



People display consistent recency and primacy effects in behavior and neural activity across perceptual and value-based judgments

Minhee Yoo¹ · Giwon Bahg² · Brandon Turner¹ · Ian Krajbich^{1,3,4} 

Accepted: 18 February 2025 / Published online: 26 March 2025
© The Author(s) 2025

Abstract

Retrospective judgments require decision-makers to gather information over time and integrate that information into a summary statistic like the average. Many retrospective judgments require putting equal weight on early and late information, in contrast to prospective judgments that involve predicting the future and so rely more on late information. We investigate how people weight information over time when continuously reporting the average stimulus strength in a sequence of displays. We investigate the consistency of these temporal profiles across perceptual and value-based tasks using both behavior and functional magnetic resonance imaging (fMRI) data. We found that people display remarkably consistent temporal weighting functions across choice domains, with a generally strong recency bias and modest primacy bias. The fMRI data revealed evidence-tracking activity in the cuneus in both tasks and in the left dorsolateral prefrontal cortex in the value-based task. Finally, a network of cognitive control regions is more active for people who exhibit a stronger primacy vs. recency bias. Together, our behavioral findings indicate that people consistently overweight recency when evaluating past information, and the neural data suggest that overcoming this tendency may require cognitive control.

Keywords Decision-making · Primacy · Recency · Decision neuroscience · Evidence averaging · fMRI

Introduction

Many of the decisions that we make require us to gather information over time and integrate that information into a summary evaluation (e.g., the average). For example, we might judge the average quality of a TV show, competence of an employee, or trustworthiness of a politician. Such decisions rely on temporal integration of evidence, from early to recent timepoints.

There are two kinds of judgments that one can make based on past evidence. There are prospective judgments, which involve predicting the future, and retrospective

judgments, which involve explaining the past. For example, suppose we are sports analysts tasked with evaluating a team's performance. We may want to predict whether that team will succeed in the future, e.g., later in the season, in the playoffs, or the following season. Conversely, we may want to determine how well that team has played so far, e.g., to give out awards or positions in the playoffs.

Prospective judgments are well understood; there is a large literature on reinforcement learning and how people should, and do, put more weight on more recent evidence (Rescorla & Wagner, 1972; Sutton & Barto, 1998). In our sports example, playoff outcomes are better predicted by end-of-season performance than early-season performance (Durand et al., 2021; Metz & Jog, 2023).

In comparison, retrospective judgments are less well understood. Do people mistake retrospective judgments for prospective judgments and put too much weight on recent evidence? Or do they jump to conclusions early, putting more weight on early evidence? Anecdotally, retrospective judgments are also overly influenced by recent performance. For example, selections for the college American football playoffs depend heavily on the last games of the season, and selections for the Major League Baseball All Star game,

✉ Ian Krajbich
krajbich@ucla.edu

¹ Department of Psychology, The Ohio State University, Columbus, OH, USA

² Department of Psychology, Vanderbilt University, Nashville, TN, USA

³ Department of Economics, The Ohio State University, Columbus, OH, USA

⁴ Department of Psychology, University of California Los Angeles, Los Angeles, CA, USA

which happens mid-season, depend disproportionately on that season's performance. Here, we seek a more rigorous investigation of these issues.

Retrospective judgments rely on memory, and memory is known to exhibit primacy and recency effects. When participants are presented with a list of words, they tend to better remember the words from both the beginning (primacy effect) and the end of the list (recency effect) than those from the middle of the list (Murdock Jr., 1962; Oberauer et al., 2018). Primacy and recency effects are not confined to verbal memory. They have been reported across different domains, such as visual and spatial memory (Hurlstone et al., 2014).

In the studies of decision-making and perceptual averaging, there is evidence for primacy and recency effects. In some cases, decision-makers appear to weight the evidence equally over time (Brunton et al., 2013; Wyart et al., 2012), as in Bayesian updating. By contrast, other work has shown that decision-makers put more weight on recent evidence (Cakici & Zaremba, 2023; Cheadle et al., 2014; Do et al., 2022; Ge et al., 2012; Hertwig et al., 2004; Johar et al., 1997; Li & Epley, 2009; Mantonakis et al., 2009; Mohrschladt, 2021; Tong et al., 2019; Tsetsos et al., 2011, 2012; Usher & McClelland, 2001; Wulff et al., 2015, 2018). Conversely, a bias towards early evidence (i.e., primacy) has also been observed (Carney & Banaji, 2012; Hubert-Wallander & Boynton, 2015; Kiani et al., 2008; Mantonakis et al., 2009; Rey et al., 2020; Wilming et al., 2020; Yates et al., 2017; Zylberberg et al., 2012). There are also individual differences; some people choose the option favored early or late in the sequence, whereas others do not show a temporal bias (Pietsch & Vickers, 1997; Tong & Dubé, 2022; Tsetsos et al., 2011; Usher & McClelland, 2001). Although these individual differences have been rigorously studied in the context of perceptual decisions, it is less clear to what degree they exist in value-based decisions.

What also remains unclear is whether these findings on memory and decision-making extend to continuous judgments, such as averaging. With rare exceptions (Do et al., 2022; Hubert-Wallander & Boynton, 2015; Johar et al., 1997; Tong & Dubé, 2022; Tong et al., 2019), primacy and recency biases are inferred from discrete choices. They are typically studied in settings where information must be stored and retrieved later, whereas in continuous judgments the decision-maker need only maintain a running average. It is unknown how the competition between primacy, recency, and optimality combines to form an overall temporal weighting function in estimating average evidence.

It is also unknown whether biases in temporal integration, if any, are consistent across perceptual and value-based domains. There has been growing interest in the commonalities and differences between perceptual and value-based decisions (Frydman & Nave, 2017; Polanía et al., 2014;

Shadlen & Shohamy, 2016; Smith & Krajbich, 2021). However, none of this work has looked at continuous judgments.

Finally, the neural mechanisms of evidence averaging are also unknown. There has been influential work studying the accumulation of evidence in discrete choice (Gluth et al., 2012; Hare et al., 2011; Heekeren et al., 2006; Juechems et al., 2017; O'Connell et al., 2018; PISAURO et al., 2017; Rodriguez et al., 2015; Shadlen & Shohamy, 2016). This work has identified input regions—perceptual areas, such as fusiform face area, parahippocampus (Heekeren et al., 2004, 2006) or medial temporal areas (Ho et al., 2009; Kayser et al., 2010), and value-based areas, including vmPFC and striatum (Gluth et al., 2012; Hare et al., 2011; PISAURO et al., 2017; Rodriguez et al., 2015), as well as accumulator regions including (pre-) motor cortex, dlPFC, and parts of parietal cortex (Liu & Pleskac, 2011; Turner et al., 2019; Wilming et al., 2020). The input regions reflect the nature of the stimuli being evaluated, whereas the accumulator regions largely reflect the response modality. Despite this influential work, there has not been any investigation of how evidence is averaged over time in the brain.

We study the ways in which people temporally weight evidence when forming perceptual and value-based judgments. We investigate individual differences in temporal weighting functions and how these individual differences arise in the brain. We use a modified interrogation paradigm (Bahg et al., 2020) in which subjects are asked to continuously report their estimate of the average evidence as new evidence becomes available, analogous to how college football rankings are updated each week after the latest set of games. In the original interrogation paradigm (Ratcliff, 2006; Turner et al., 2017; Usher & McClelland, 2001), a response cue is presented after a delay following the presentation of a stimulus. Subjects are instructed to make a single choice immediately after the response cue. In contrast, in the modified interrogation paradigm, subjects continuously report their estimates of average evidence throughout a trial. These ongoing measures allow us to investigate how temporal biases impact the estimates of average evidence. Moreover, this paradigm allows us to identify brain regions that track the average evidence by correlating the reports with fMRI signals.

To preview the results, we find that people show considerable recency bias and some primacy bias. These temporal weighting functions are highly consistent within an individual, showing strong correlations between perceptual and value-based tasks. As expected, parietal cortex and the fronto-striatal reward network encode the inputs for perceptual and value-based tasks, respectively, while only the dorsolateral prefrontal cortex (dlPFC) represents the averaged evidence, and only in the value-based task. Finally, we find that people who exhibit more primacy bias compared to recency bias display more activity in the cognitive control network including the intraparietal sulcus and dlPFC.

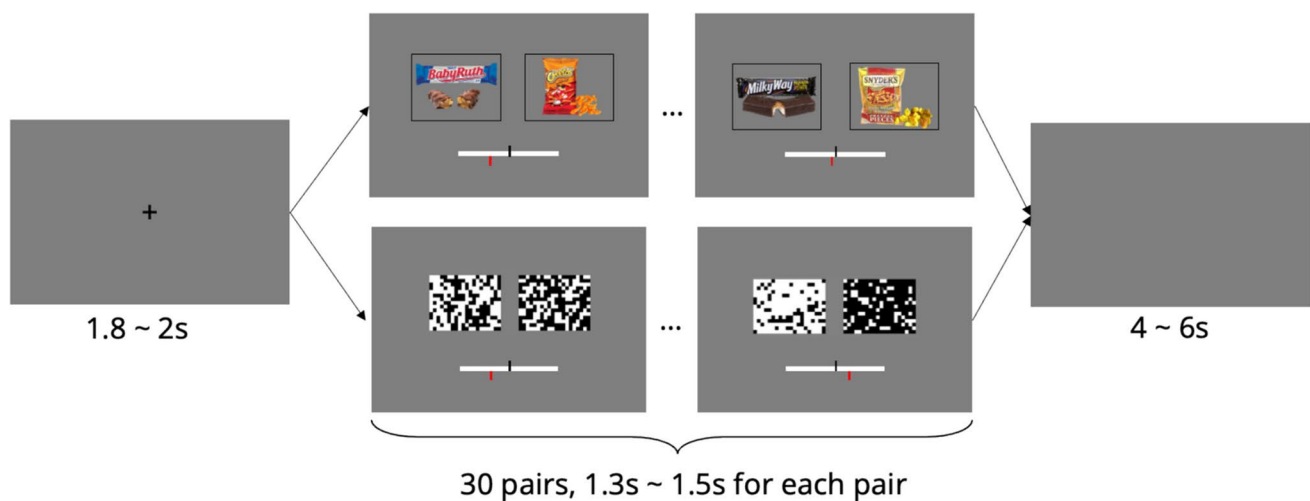


Fig. 1 Timeline of the evidence-averaging task. After a fixation cross, subjects saw 30 pairs of stimuli in a trial. They saw each pair for 1.3 to 1.5 s. Square grids and snack foods were presented in the perceptual and value-based tasks, respectively. Subjects continuously

reported the average difference between the right and left sides by moving the red slider bar along the scale with a joystick. The center of the scale was marked with a black vertical bar. At the end of each trial, there was a 4- to 6-s delay before the next trial

Methods

Subjects

Forty-six subjects participated in this study (23 males, mean age 21.4 years). Subjects were recruited from The Ohio State University Experimental Economics Pool. They received \$35 as a base payoff and could earn an additional payoff up to \$10. All subjects were right-handed and had normal or corrected-to-normal vision.

We excluded eight subjects from the analyses. One subject had a structural abnormality in their brain, and six subjects moved their head more than 3 mm. We additionally excluded one subject owing to failure to understand the task, based on a postexperiment questionnaire where they indicated that they tracked the difference between single pairs of stimuli rather than the average.

Experimental procedure

Rating task

Subjects completed two tasks. In the first task, which took place outside the MRI scanner, they rated 144 snack foods on a continuous scale from 0 (least liked) to 10 (most liked). Before rating the foods, subjects saw all of them in a slideshow. After the slideshow, subjects saw one food at a time and rated how much they would like to eat it at the end of the experiment. They used the mouse to move a slider along a rating scale and clicked the left mouse button to mark their ratings.

Evidence-averaging task

After the rating task, subjects completed an evidence-averaging task in the MRI scanner, which took approximately 50 min to complete. The evidence-averaging task consisted of two blocks of a perceptual task and two blocks of a value-based task. The two tasks alternated. The order of the blocks was counterbalanced across subjects. Each block had 15 trials.

In each trial, subjects saw 30 pairs of stimuli in series (Fig. 1). Each pair was presented for 1.3 to 1.5 s. The stimuli were grids of black and white squares in the perceptual task and snack foods in the value-based task. While watching each series of stimuli, subjects continuously reported the average difference between the right and left stimulus by using a joystick (Current Designs Tethyx) to move a slider bar along a scale. Specifically, they reported which side of the screen had on average more white squares or better foods, and by how much.

There were no response cues during the task. The slider began each trial at the center of the scale. Subjects moved the slider whenever they thought the average difference had changed. The slider remained in the same position until subjects moved the joystick. The tilt of the joystick determined how fast the slider moved in the tilted direction. We recorded the slider position at 30 Hz.

Before the MRI task, subjects learned how to map stimulus differences to positions on the scale. Subjects first learned the meaning of the endpoints on the scale. In the perceptual task, they were informed that the endpoints indicated that one side had 60% more white dots. In the value-based task, they were informed that the endpoints indicated

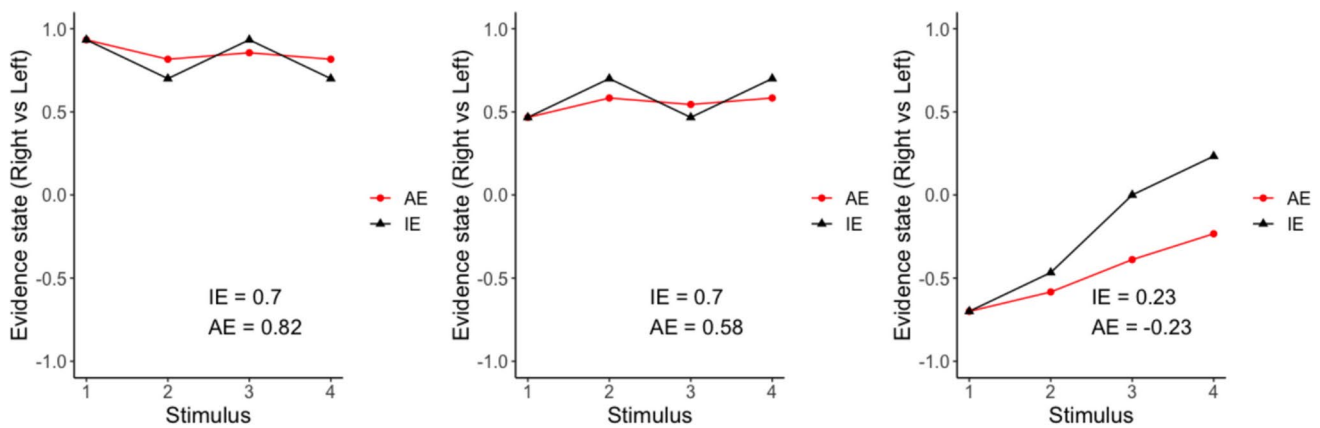


Fig. 2 The difference between instantaneous evidence (IE) and average evidence (AE). Time course examples of the first four pairs of stimuli in three trials. Each panel shows how IE and AE change over time in a trial. IE (black) represents the difference between a pair of

stimuli at the current point in time. AE (red) represents the average evidence up to that point in time. In other words, the red AE curve is the cumulative mean of the black IE curve. The IE and AE written in each panel represent their values after the fourth stimulus pair

that foods on that side were better by x (the rounded maximum rating difference between two foods) out of 10. Once subjects learned the meaning of the scales' endpoints, we displayed several examples where we moved the slider from the left end to the right end of the scale and showed corresponding stimulus pairs. We showed three examples at nine points on the scale for 1.5 s each. After seeing the examples, subjects practiced a trial without feedback.

Subjects earned money based on their performance in the evidence-averaging task. We determined their earnings based on the deviation of the slider from the correct answer at the end of each stimulus pair, averaged across all stimulus pairs. The correct answer was the average of all the stimuli presented up to that time in a trial. Subjects earned the maximum and minimum bonus when the slider deviated from the correct answer by less than 4% and 20% of the scale length, respectively. The bonus decreased gradually from \$10 to \$1 between these extremes. Subjects earned \$2.63 on average (standard deviation = 2.28).

Sequence

We generated two block-length sequences to present stimuli in each task. Each sequence was designed to minimize the correlation between the instantaneous evidence (IE) and average evidence (AE), i.e., the mean of IE. The IE is the difference in the number of white squares (perceptual) or subjective values (value-based) between the two stimuli on the screen, right minus left. The IE had nine levels, in arbitrary units from -0.4 to 0.4 , increasing by 0.1 . To generate a trial-level sequence, we simulated a random walk process with 30 steps corresponding to 30 stimulus pairs in a trial. We used a random walk process to generate stimulus sequences to

ensure that the input would gradually change over time. This gradual change allowed the AE to evolve smoothly, which made it easier for subjects to update their AE and report it using a joystick. The initial value was drawn from the nine possible IE levels. We determined the direction of the random walk by sampling a value from *Uniform*(0,1) and calculating whether it was greater or smaller than 0.5. Given the direction, we determined the step size by sampling a value from $[Poisson(0.3) + 1]/10$. Examples of the sequences for the first four stimulus pairs in a trial are illustrated in Fig. 2.

We created 15 trial-level sequences per block. We generated 100 block-level sequences by repeating this process and then selected the one with the lowest correlation between IE and AE. We generated two such sequences and used them in both perceptual and value-based tasks, in the same order. The order of the block-level sequences was fixed within each subject but was counterbalanced across subjects.

Stimulus

In the perceptual task, we manipulated the difference in the number of white squares between the two grids. Each grid had 300 squares, with 15 rows and 20 columns. The size of each square was 20×20 pixels. The total number of white squares on the screen was constant; when we increased the number of white squares in one grid, we also decreased it by the same amount in the other grid. When IE was 0, half of each grid was white and the other half was black. As the magnitude of IE increased, we increased the percentage of white in one grid by 7% and decreased the percentage of white in the other grid by 7%. This resulted in difference levels of 0%, 14%, 28%, 42%, and 56% of the total number of white squares between the two grids. The endpoint of the

slider scale roughly corresponded to the maximum difference between the two grids, namely 60%.

In the value-based task, we manipulated the rating difference between the two foods. We chose stimuli based on each subject's ratings. We excluded the lowest and highest rated items (5% each), including all items rated 0 or 10. We then computed the rating differences between all pairs of the remaining snack foods. Next, we evenly divided the range of value differences into nine bins, which were then matched with the nine levels of IE. Stimulus pairs from each bin were drawn according to the specified sequence. The endpoint of the scale was customized for each subject and corresponded to the rounded maximum rating difference between two foods.

For behavioral data analysis, we rescaled the position of the slider, the rating difference between snack foods, and the difference in the number of white squares between grids to be in the range of -1 to $+1$. This rescaling process made it easier to compare results between the two decision domains and fit computational models to the data.

Behavioral data analysis

Accuracy

We first assessed how well subjects tracked the AE. If subjects were able to perfectly average all evidence without any noise in the evidence-averaging process, their responses would align with the running average of the IE (simple AE). To quantify the accuracy of subjects in the evidence-averaging task, we computed the mean error (ME). Specifically, ME was calculated as the mean absolute difference between the last slider position $y_{b,l,s}$ of stimulus pair s in trial l of block b and the simple $AE_{b,l,s}$: $\frac{1}{N} \sum_b \sum_l \sum_s |y_{b,l,s} - simpleAE_{b,l,s}|$ where $simpleAE_{b,l,s} = \sum_s IE_{b,l,s} / N_s$. Here, N refers to the total number of stimulus pairs (i.e., 2 blocks per decision domain \times 15 trials per block \times 30 pairs per trial = 900 pairs), and N_s refers to the number of stimulus pairs presented up to that time in a trial. We computed ME for each subject on each task.

Next, we examined how well subjects tracked AE changed over time. If subjects showed a temporal bias, their responses would deviate more from the simple AE as time passed within a trial. To test this possibility, we ran a linear regression on the error in the evidence-averaging task: $|y_{b,l,s} - simpleAE_{b,l,s}|$. Time within a trial, task type (value-based vs. perceptual), and their interaction were included as regressors. Standard errors were clustered at the subject level.

Lastly, we examined how sensitive subjects were to IE and whether the sensitivity to IE changed over time. When computing the average, the influence of IE on AE decreases over time. Thus, we expect to see a decreasing effect of IE

on AE as time passes within a trial. To test this possibility, we ran a linear regression on AE updates in response to each IE within a trial. AE updates were defined as the difference between the last slider positions of two consecutive stimulus pairs: $y_{b,l,s} - y_{b,l,s-1} \cdot IE(IE_{b,l,s})$, time within a trial, task type (value-based vs. perceptual), and all their interactions were included as regressors. Standard errors were clustered at the subject level.

Deviance index

Next, we examined temporal bias in the evidence-averaging process. Specifically, we assessed the degree to which subjects exhibited recency bias. At one extreme a subject could perfectly report the AE, while at the other extreme a subject could discard past information and simply report the IE (Fig. 2). IE and AE are the same at the beginning of the trial but diverge over time. To quantify how closely subjects' responses aligned with AE vs. IE, we computed the ratio of how much the last slider position $y_{b,l,s}$ of stimulus pair s in trial l of block b deviated from IE vs. AE:

$$d_a = \frac{\sum_b \sum_l \sum_s |y_{b,l,s} - IE_{b,l,s}|}{\sum_b \sum_l \sum_s |y_{b,l,s} - simpleAE_{b,l,s}|}.$$

We first computed the deviation of the last slider position from IE and AE for each stimulus pair, then summed up those deviations, and then computed the ratio between the two sums. This deviance index (d_a) quantifies the overall response deviation from IE relative to AE. The deviance index has a range from 0 to infinity, with larger values corresponding to better performance, i.e., better tracking of AE. We computed d_a for each subject on each task.

Averaging Diffusion Model

We used the Averaging Diffusion Model (Turner et al., 2017) to investigate how subjects computed AE over time. The ADM is a modification of the Drift Diffusion Model (DDM). The ADM assumes that a decision maker integrates noisy samples of evidence to estimate the mean of that evidence. The ADM explains how the estimate of the average evidence evolves over time and links it to a decision maker's response at any given time.

To calculate the decision variable of the ADM at time t (denoted $\mu(t)$), we used a weighted average of IE. Specifically, when a total of s_t stimulus pairs have been presented up to time t and the weight on IE of the i -th stimulus pair is $w_{s_t,i}$, the ADM's decision variable is a weighted average of IE presented up to that time: $\mu(t) = \sum_{i=1}^{s_t} w_{s_t,i} IE_i / \sum_{i=1}^{s_t} w_{s_t,i}$. Note that i indexes stimulus pairs while t indexes measurements of the slider position. The weight on the i -th stimulus pair is determined by a temporal weighting

function (Galdo et al., 2022; Pooley et al., 2011): $w_{s_t,i} = \left[1 - \left(1 - \epsilon_{primacy}^i\right)\left(1 - \epsilon_{recency}^{s_t-i+1}\right)\right](1 - \eta) + \eta$. Since s_t is updated with each new stimulus pair, the weight on each stimulus pair and the ADM's decision variable are updated every time a new IE is presented. A primacy parameter $\epsilon_{primacy}$ and a recency parameter $\epsilon_{recency}$ determine the weights on early and recent IE. The lowest possible weight on a stimulus pair is determined by η . We set η as 0.01 and estimated $\epsilon_{primacy}$ and $\epsilon_{recency}$. The temporal weighting function determines the weights on early and recent decision evidence relative to the baseline η . As long as the baseline is low, $\epsilon_{primacy}$ and $\epsilon_{recency}$ can adjust to match an individual's temporal weighting function. Thus, we fixed η and estimated only $\epsilon_{primacy}$ and $\epsilon_{recency}$.

The ADM can represent two mechanisms for updating the AE. First, it can represent a process in which subjects recompute the AE every time a new IE is presented. When a new IE is presented, the weights on each IE are updated using a temporal weighting function and the AE is recomputed with the updated weights and the IEs stored in memory. Second, under a certain constraint, the ADM can represent a process in which subjects maintain their current estimate of AE and update it with each new IE. The AE in the ADM at time t can be re-expressed as: $AE(t) = \frac{\sum_{i=1}^{s_t-1} w_{s_t,i} IE_i + w_{s_t,s_t} IE_{s_t}}{\sum_{i=1}^{s_t} w_{s_t,i}}$. When the weight on each IE is fixed and does not change over time ($w_{1,i} = w_{2,i} \dots w_{30,i} = w_i$), the AE in the ADM can be reformulated as: $AE(t) = \frac{\sum_{i=1}^{s_t-1} w_i}{\sum_{i=1}^{s_t} w_i} AE_{s_t-1} + \frac{w_{s_t}}{\sum_{i=1}^{s_t} w_i} IE_{s_t}$. This formulation represents the process of maintaining the most recent estimate of AE and updating it with each new IE.

The measured average evidence at time t (i.e., the position of the slider), $E(t)$, follows a normal distribution $E(t) \sim Normal(\mu(t), \sigma^2(t))$. $\sigma(t)$ represents the standard deviation of the average evidence. As the subject collects more evidence, their estimate of the AE becomes increasingly precise. Thus, the standard deviation of AE decreases over time: $\sigma(t) = \sigma_w / \sqrt{t}$. Here, σ_w is the standard deviation of within-trial variability in the samples of evidence.

We compared the ADM fits to subjects' AE reports in the tasks. To do so, we computed the likelihood of the slider position $y_{b,l,t}$ at time t in trial l of block b , given the distribution of the measured average evidence at that time:

$$L(\theta|D) = \prod_b \prod_l \prod_t Normal(y_{b,l,t} | \mu_{b,l}(t), \sigma_{b,l}^2(t))$$

where θ is a generic notation for the ADM parameters. We evaluated this likelihood for all recorded slider positions, except for those in the first stimulus pair of each trial. These initial measurements were excluded to account for the large movements that are initially required in the task, coupled with the sluggish nature of moving the slider with the joystick.

We constructed four variants of the ADM based on hypotheses about temporal bias and noise in the evidence-averaging process. First, we tested whether the temporal bias is consistent across the two tasks or not. To test this hypothesis, we constructed two models: a separate-temporal-bias model and a common-temporal-bias model. In the separate-temporal-bias model, we separately estimated $\epsilon_{primacy}$ and $\epsilon_{recency}$ for the perceptual task and the value-based task. In contrast, we estimated a single $\epsilon_{primacy}$ and $\epsilon_{recency}$ across the two tasks in the common-temporal-bias model.

Next, we considered whether the noise in the evidence-averaging process differs between the perceptual task and the value-based task. We constructed two models, a separate-noise model and a common-noise model, to test this hypothesis. In the separate-noise model, σ_w was allowed to differ between the perceptual task and the value-based task. However, the common-noise model assumes a single noise parameter for both tasks, and thus only one σ_w was estimated across the two tasks.

We fitted four models to each subject's data by considering all possible combinations of these hypotheses (Table S1). To determine the better model for each subject, we compared the performance of the four models using the widely applicable information criterion (WAIC; Vehtari et al., 2017; Watanabe, 2010). The model with the lower WAIC was chosen as the preferred model for that subject.

We fitted these models to each subject's data using STAN (Stan Development Team n.d.). We set $Beta(2,4)$ as the prior of $\epsilon_{primacy}$ and $\epsilon_{recency}$, and $Normal(-1, 1)$ as the prior of $\log(\sigma_w)$. In the first fitting attempt, we collected 4000 samples from each of the four chains after discarding the first 1000 samples. Then, we computed \hat{R} values (Bürkner et al., 2023; Vehtari et al., 2021) to assess the convergence of the chains. \hat{R} values serve as indicators of convergence in posterior chains. Ideally, \hat{R} values should be close to 1, indicating that the chains have converged. \hat{R} values were higher than 1.01 for some subjects, suggesting a divergence of chains. Divergence was observed in 14 subjects in the separate-temporal-bias and separate-noise model, 9 subjects in the common-temporal-bias and separate-noise model, 12 subjects in the separate-temporal-bias and common-noise model, and 10 subjects in the common-temporal-bias and common-noise model. To address this issue, we re-fitted the models for subjects where divergence occurred. We initialized chains with informed initial values in this case. We selected the chain that had the highest log posterior probability. Then, we drew random values from the 95% highest-density interval (HDI) of parameters of the selected chain to initialize chains in the second fitting. In the second fitting, we used four chains and collected 4000 samples from each chain after discarding 1000 samples. After the second fitting, \hat{R} values of all parameters were smaller than 1.01, indicating a successful convergence.

MRI data acquisition and analysis

We collected the functional and structural MRI data at the Center for Cognitive and Brain Imaging at The Ohio State University. We used a 32-channel head coil and 3 T MRI scanner (MAGNETOM Tim Trio; Siemens Medical Solutions). We first collected structural MRI data [repetition time (TR) = 2400 ms; echo time (TE) = 2.22 ms; flip angle = 8 degree; 208 slices; voxel size $0.8 \times 0.8 \times 0.8$ mm; 300×320 matrix size]. Then, we collected four runs of functional MRI data (740 volumes per run, around 12 min). Functional MRI data was acquired with a multiband echo-planar imaging sequence [TR = 1000 ms; TE = 28 ms; flip angle = 60 degree; 45 interleaved slices, voxel size $3 \times 3 \times 3$ mm, 72×72 matrix size; multiband acceleration factor = 3]. We tilted the acquisition plane 15 degrees upwards from the line connecting the anterior commissure and posterior commissure to reduce signal dropout in the orbitofrontal cortex (Deichmann et al., 2003). A field map was collected to correct the spatial distortion caused by inhomogeneity of the magnetic field [TR = 670 ms; TE1 = 5.19 ms; TE2 = 7.65 ms; flip angle = 60 degree; 60 slices; voxel size $3 \times 3 \times 3$ mm; 72×72 matrix size].

Stimuli were presented using MATLAB Psychtoolbox (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) in MATLAB and displayed with a DLP projector onto a screen.

Preprocessing

We preprocessed and analyzed the MRI data with Statistical Parametric Mapping 12 (SPM12; the Wellcome Trust Centre for Neuroimaging, University College London, UK). We corrected the different slice acquisition time of the functional MRI data. Then, we realigned and unwarped functional MRI data to correct head movements and spatial distribution using a voxel displacement map. The voxel displacement map was created by processing the field map with the field map toolbox. The structural MRI data was normalized to the MNI space by a unified segmentation. Structural MRI data was coregistered to the functional MRI data. Functional MRI data was normalized to the MNI space by using the deformation field used to normalize structural MRI data into the MNI space. We smoothed the normalized functional MRI data with an 8-mm full width at half maximum Gaussian kernel.

Statistical analysis

We constructed two generalized linear models (GLMs) to identify brain regions tracking AE and IE. Each block was modeled with four regressors and six realignment parameters that were generated during preprocessing.

GLM 1

The first regressor was the onset of each stimulus pair with a duration of zero. This regressor had three parametric modulators: unsigned IE, unsigned AE, and the joystick position. Unsigned IE is the absolute difference between stimuli in a pair (normalized to lie between 0 and 1). Unsigned AE is the absolute AE predicted by the ADM with separate temporal weighting functions and common noise for both tasks. The joystick position captures how much the joystick was tilted from the center (normalized to lie between 0 and 1). We used the first sampled joystick position after stimulus presentation. We turned off orthogonalization of parametric modulators. All regressors were modeled with a canonical hemodynamic response function without temporal or dispersion derivatives.

GLM 2

Regressors in GLM2 were similar to those in GLM1, except that instead of unsigned AE, it included subjects' responses, specifically their absolute final slider position for each stimulus pair. If the slider was at the center or at either end of the scale, this variable was zero or one, respectively. Note that the tilt of the joystick determined how fast the slider moved in the tilted direction and not the position of the slider. Thus, subjects' responses (slider position) and the joystick position were not usually the same.

Both GLMs

We created four contrast images for each GLM. For each task, we made a positive contrast image for the unsigned IE regressor and a positive contrast image for the unsigned AE regressor. We ran a one-sample *t*-test with each contrast image at the group-level. To correct for multiple comparisons, we applied a threshold of corrected $p < 0.05$. We defined clusters as $p = 0.001$ and further corrected the *p*-values based on the size of clusters (Woo et al., 2014) or the peak activation of a cluster.

ROI analysis

We explored the neural mechanisms of temporal bias using a ROI analysis. In our task, the influence of each new sample on the average decreases as a trial progresses. As a result, a decision maker may adopt a strategy that prioritizes early samples and downweights new samples. This strategy leads to a primacy bias. Thus, primacy bias may arise from the activity of brain regions involved in inhibiting the updates of the average, such as brain regions related to cognitive control.

We defined ROIs using a term-based meta-analysis at Neurosynth (Yarkoni et al., 2011). We used the association

test to create a brain map that is preferentially related to “cognitive control.” Based on the meta-analysis map, clusters in bilateral intraparietal cortex (Left: $x = -32$, $y = -56$, $z = 44$; Right: $x = 52$, $y = -45$, $z = 46$) and bilateral dorsolateral prefrontal cortex (Left: $x = -44$, $y = 20$, $z = 30$; Right: $x = 44$, $y = 17$, $z = 34$) were selected as our ROIs. Then, we extracted the beta estimates of the stimulus onset regressor in the GLM1 using the MarsBaR package (Brett et al., 2002) in SPM. Beta estimates of the stimulus onset regressor were correlated with the mean of the difference between primacy and recency parameters for each decision domain. We also examined the correlation between beta estimates of the stimulus onset regressor the mean of primacy and recency parameter, respectively (Table S7).

Results

Accuracy

We used the ME to evaluate subjects’ performance in the evidence-averaging task. If a subject had no temporal bias and successfully averaged all the evidence without any noise in the evidence-averaging process, their responses would follow the simple AE, resulting in an ME of zero. Therefore, lower ME values indicate better tracking of the simple AE due to a smaller temporal bias and less noise in the evidence-averaging process.

In both the perceptual and value-based tasks, the ME was larger than zero (one sample *t*-test, Perceptual task: $t(37) = 19.19$, $p = 10^{-21}$; Value-based task: $t(37) = 20.88$, $p = 10^{-22}$). The average ME across subjects was 0.27 for both tasks. This means that, on average, subjects deviated from the simple AE by approximately 13% of the length of the scale. Notably, there was no significant difference in performance between the two decision domains (paired *t*-test, $t(37) = 0.58$, $p = 0.57$).

The deviation of subjects’ responses from the simple AE grew over time in a trial (Table S8). Time had a positive effect on the error in the evidence-averaging task (time: $\beta = 0.003$, $p = 0.0004$). The error was initially larger in the value-based task (task type value-based: $\beta = 0.02$, $p = 0.001$) but increased more slowly than in the perceptual task (interaction of time and task type value-based: $\beta = -0.002$, $p = 0.0001$).

Subjects became less sensitive to IE over time within a trial when computing AE (Table S9). IE had a positive effect on AE updates (IE: $\beta = 0.18$, $p < 10^{-16}$), and the positive effect of IE on AE decreased over time within a trial (interaction of IE and time: $\beta = -0.005$, $p < 10^{-16}$). The effect of IE on AE was smaller in the value-based task (interaction of IE and task type value-based: $\beta = -0.04$, $p = 10^{-11}$), but decreased less over time than in the perceptual task

(interaction of IE, time, and task type value-based: $\beta = 0.001$, $p = 10^{-11}$).

Deviance index

We used a deviance index to quantify temporal bias in the evidence-averaging task. The deviance index ranges from 0 to infinity. The higher deviance index means that subjects were tracking the simple AE rather than tracking the IE. A deviance index of 1 corresponds to being equally distant from the simple AE and the IE, whereas a deviance index < 1 corresponds to being closer to the IE than the simple AE. Thus, a deviance index < 1 indicates behavior more in line with reporting the evidence from only the most recent sample.

Individual differences in temporal bias were evident in our study. The deviance index was significantly higher than 1 in both tasks ($M_{\text{Perceptual}} = 1.50$, $M_{\text{Value-based}} = 1.50$; one-sample *t*-test, Perceptual task: $t(37) = 3.87$, $p = 10^{-4}$; Value-based task: $t(37) = 5.91$, $p = 10^{-7}$; Fig. 3A). This indicates that, on average, subjects tended to track the AE more than the IE.

However, the range of the deviance index was considerable, varying from 0.34 to 3.09 in the perceptual task and from 0.69 to 2.72 in the value-based task. Notably, 13 subjects in the perceptual task and five subjects in the value-based task had deviance indexes less than one, showing a strong recency bias.

Despite significant individual differences in temporal bias, there was remarkable consistency within an individual across tasks. The deviance index in the perceptual task was not significantly different from the deviance index in the value-based task (paired *t*-test, $t(37) = 0.04$, $p = 0.97$; Fig. 3A). Additionally, there was a very strong positive correlation between the deviance indexes for the two tasks ($r(36) = 0.96$, $p < 10^{-16}$; Fig. 3B). This shows that subjects with a strong recency bias in the perceptual task also had a strong recency bias in the value-based task.

Averaging Diffusion Model

Among the four models considered, the model assuming separate temporal bias and separate noise for perceptual task and value-based task provided the best fit for all subjects based on the WAIC goodness-of-fit metric (see Fig. 4 for model fits and Table S1 for WAIC). This indicates that the exact shape of the temporal weighting function differed between the two tasks, and the noise in the evidence-averaging process also varied across tasks.

To further investigate temporal bias using the ADM, we examined the posterior distributions of parameters from the separate-temporal-bias and separate-noise model

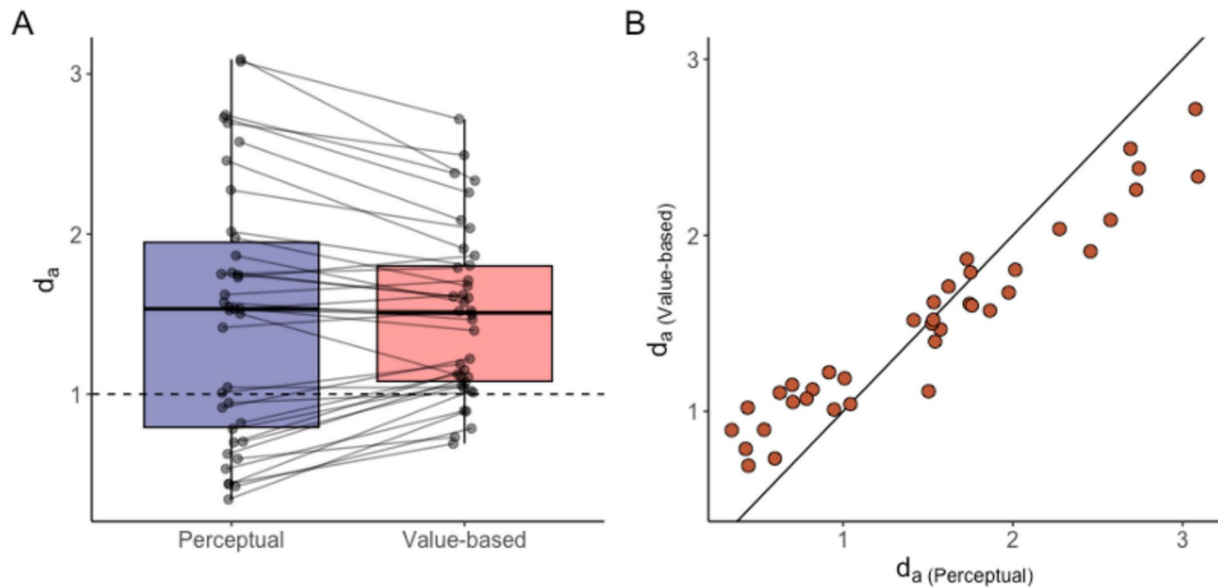


Fig. 3 Deviance index. **(A)** Deviance index for the perceptual and the value-based tasks. Each dot represents the deviance index of an individual subject, and each line connects an individual subject's devi-

ance index across the two tasks. **(B)** Scatter plot of the deviance index between the two tasks. Each dot represents the deviance index of an individual subject. The line represents the identity line ($y = x$)

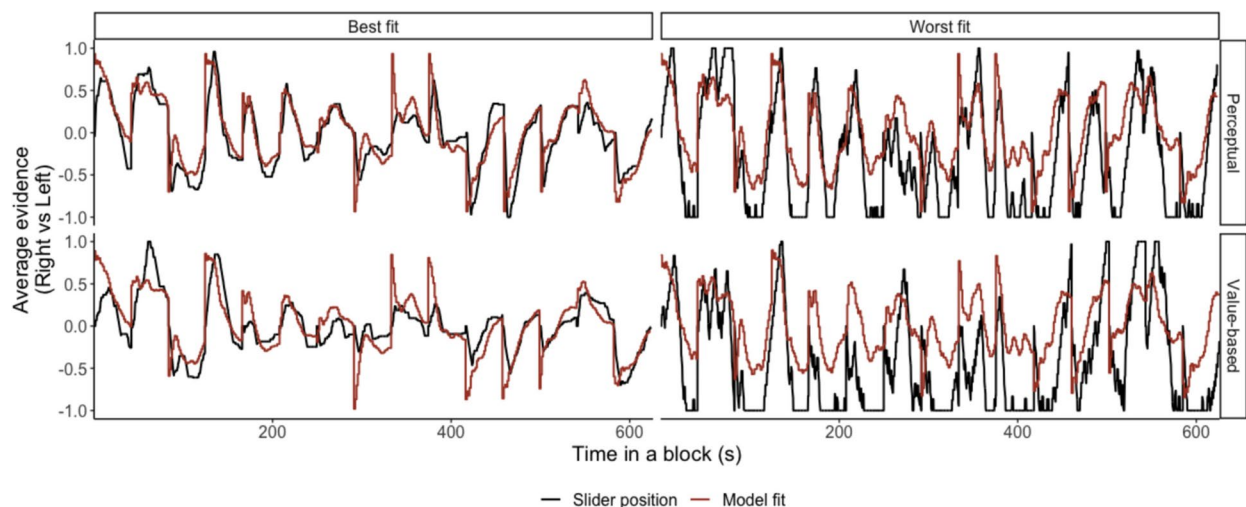


Fig. 4 Model fits. The figure shows the responses of two subjects, i.e., slider positions on the scale (black line), and fits of the separate-temporal-bias model (red line). The two subjects' data from a block of perceptual and value-based tasks are displayed. The two blocks share the same stimulus sequence. These subjects had the lowest (left

column, best fit) and highest (right column, worst fit) WAIC to illustrate how well the Averaging Diffusion Model (ADM) explains the data. The mean of posterior distributions was used to generate the model fits of the ADM

(Table S2). We quantified temporal bias by computing the difference between $\epsilon_{primacy}$ and $\epsilon_{recency}$ in each task (i.e., $\epsilon_{primacy} - \epsilon_{recency}$). Then, we assessed whether the 95% HDI of $\epsilon_{primacy} - \epsilon_{recency}$ was strictly positive or negative, and whether $\epsilon_{primacy} - \epsilon_{recency}$ was correlated between tasks.

We observed a substantial recency bias in both the perceptual and value-based tasks. In the perceptual task, 34

of 38 subjects had higher values for $\epsilon_{recency}$ than $\epsilon_{primacy}$ (Figs. 5A-B). The 95% HDI of $\epsilon_{primacy} - \epsilon_{recency}$ was strictly negative and the posterior probability that $\epsilon_{recency}$ was larger than $\epsilon_{primacy}$ was 1 for these 34 subjects. By contrast, four subjects showed the opposite pattern, indicating a primacy bias. The 95% HDI of $\epsilon_{primacy} - \epsilon_{recency}$ was strictly positive

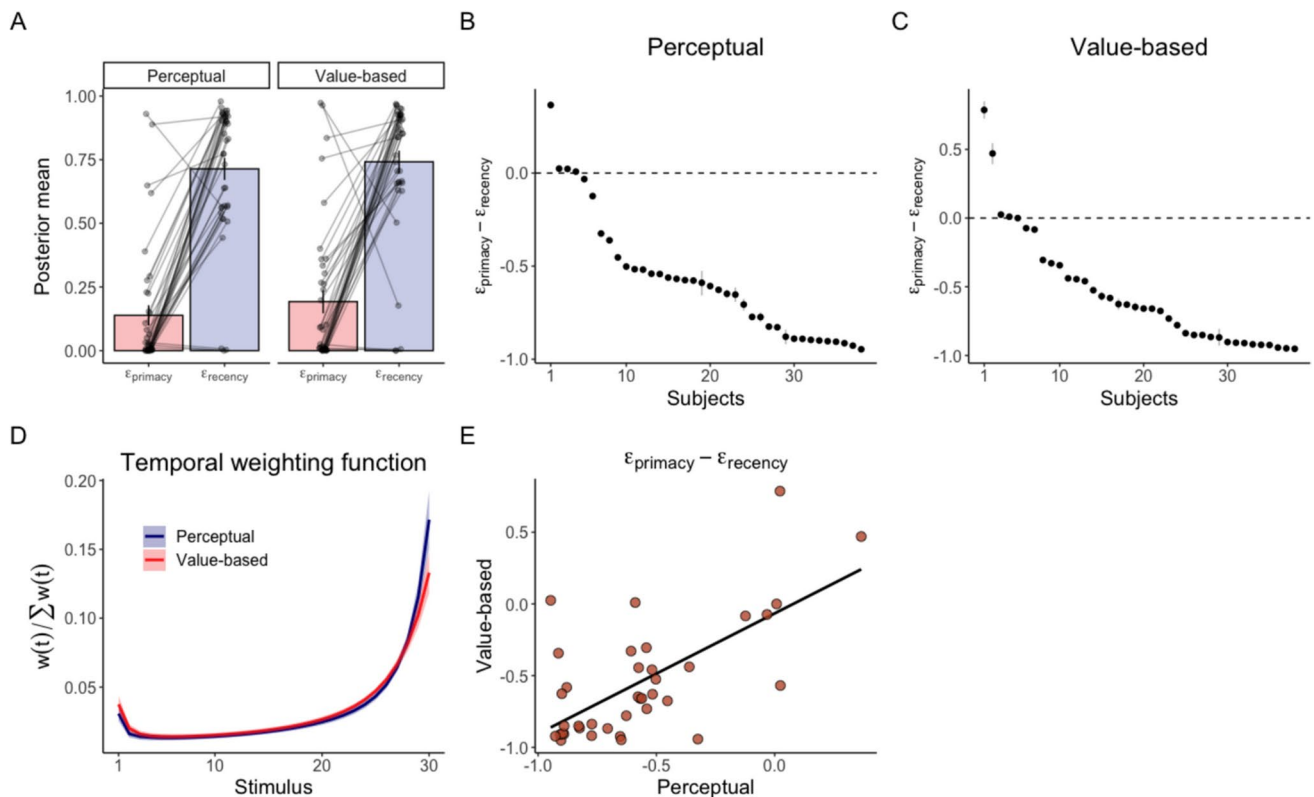


Fig. 5 Posterior distribution of $\epsilon_{primacy}$ and $\epsilon_{recency}$. (A) The mean of posterior distribution of $\epsilon_{primacy}$ and $\epsilon_{recency}$. Bars show the mean of subjects' posterior means. Error bars represent standard errors across subjects. Each dot and line represent the mean of the posterior distribution of an individual subject. The difference between $\epsilon_{primacy}$ and $\epsilon_{recency}$ in the perceptual task (B) and the value-based task (C). Dots represent the mean and gray bars represent the 95% highest density interval. Subjects are sorted by the mean. (D) Temporal

weighting function. The temporal weighting function $w(t)$ was computed using the posterior mean of $\epsilon_{primacy}$ and $\epsilon_{recency}$ of an individual subject. Then, $w(t)$ was averaged across subjects. The line shows the mean temporal weighting function. The shaded area represents standard errors across subjects. (E) Scatter plot of the difference between $\epsilon_{primacy}$ and $\epsilon_{recency}$ in the two tasks. Each dot represents the mean of an individual subject's posterior distribution. The line represents the mean trend from linear regression

and the posterior probability that $\epsilon_{recency}$ was larger than $\epsilon_{primacy}$ was 1 for these four subjects.

We observed similar results in the value-based task (Figs. 5A and C). As in the perceptual task, a majority of subjects—34 out of 38—had higher values for $\epsilon_{recency}$ than $\epsilon_{primacy}$. The 95% HDI of $\epsilon_{primacy} - \epsilon_{recency}$ was strictly negative and the posterior probability that $\epsilon_{recency}$ was larger than $\epsilon_{primacy}$ was 1 for these 34 subjects. By contrast, three subjects had a strictly positive HDI of $\epsilon_{primacy} - \epsilon_{recency}$ and a posterior probability that $\epsilon_{primacy}$ was larger than $\epsilon_{recency}$ equal to 1. For one subject, the 95% HDI included zero and the posterior probability that $\epsilon_{recency}$ was larger than $\epsilon_{primacy}$ was 0.66. While temporal bias for this subject cannot be definitively determined based on the 95% HDI, the posterior distribution supports a recency bias.

We also computed the temporal weighting function at the 30th stimulus pair (i.e., the last stimulus pair on each trial) using the posterior means of $\epsilon_{primacy}$ and $\epsilon_{recency}$. These temporal weighting functions clearly show that the weights on late stimulus pairs were higher than those on early stimulus

pairs in both the perceptual and value-based tasks (Fig. 5D; see Fig. S1 for the subject-level functions).

The temporal weighting function was highly consistent across tasks. The means of $\epsilon_{primacy} - \epsilon_{recency}$ in the two tasks were positively correlated (Fig. 5E; $r(36) = 0.66$, $p = 10^{-5}$). Consistent with the deviance results, this analysis indicates that subjects with a strong recency bias in one task also showed a strong recency bias in the other task.

In addition to the temporal bias, we examined the diffusion noise σ_w in both decision domains. We assessed whether the 95% HDI of the difference in σ_w between the perceptual and value-based tasks was positive or negative. Thirty-three subjects had strictly negative HDIs, indicating that they had larger σ_w in the value-based task compared to the perceptual task (Figs. 6A–B). The posterior probability that σ_w was larger in the value-based task than in the perceptual task was 1 for these 33 subjects. By contrast, five subjects had strictly positive HDIs, and a probability that σ_w in the perceptual task was larger than in the value-based

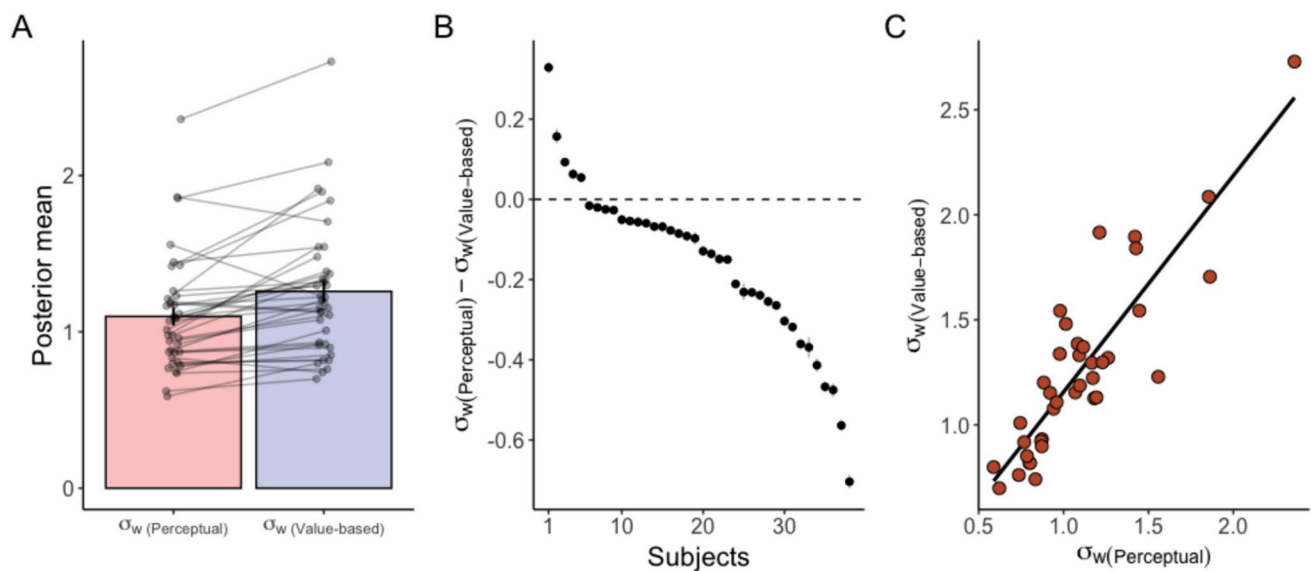


Fig. 6 Posterior distribution of σ_w . (A) The mean of the posterior distribution of σ_w . Bars show the mean of subjects' posterior means. Error bars represent standard errors across subjects. Each dot and line represent the mean of the posterior distribution of an individual subject. (B) The difference in σ_w between the perceptual and the value-based tasks. The dots represents the means, and the gray

bars represent the 95% highest density intervals. Subjects are sorted by the mean. (C) Scatter plot of σ_w in the perceptual and the value-based tasks. Each dot represents the mean of an individual subject's posterior distribution. The line represents the mean trend from linear regression

task equal to 1 (Figs. 6A–B). Also, the mean of the posterior distribution of σ_w in the two tasks was positively correlated (Fig. 6C; $r(36) = 0.88$, $p = 10^{-13}$). Thus, subjects with a smaller σ_w in one task also tended to have a smaller σ_w in the other task.

Lastly, we explored the relationship between subjects' performance in the evidence-averaging task and the parameters of the ADM. Specifically, we examined the correlation between ME and the posterior means of ADM parameters. We found that ME in the evidence-averaging task was negatively correlated with the posterior mean of $\epsilon_{\text{primacy}} - \epsilon_{\text{recency}}$ in both tasks (perceptual task: $r(36) = -0.34$, $p = 0.04$; value-based task: $r(36) = -0.37$, $p = 0.02$). In other words, more error was correlated with more recency vs. primacy bias. Also, the ME and σ_w were positively correlated in both tasks (perceptual task: $r(36) = 0.66$, $p = 10^{-6}$; value-based task: $r(36) = 0.88$, $p = 10^{-13}$).

Neuroimaging results

GLM1. The model with ADM-predicted AE

In the GLM with ADM-predicted AE, we found distinct brain regions that tracked the unsigned IE (Table S3) in the two tasks. In the perceptual task, the inferior parietal gyrus ($x = -42$, $y = -58$, $z = 50$, $Z = 4.62$; Fig. 7A) and angular gyrus ($x = 42$, $y = -74$, $z = 40$, $Z = 6.30$) tracked

the unsigned IE. In the value-based task, both the ventral striatum ($x = -12$, $y = 8$, $z = -8$, $Z = 5.86$; Fig. 7B) and the medial prefrontal cortex ($x = -2$, $y = 40$, $z = 24$, $Z = 4.96$; Fig. 7B) tracked the unsigned IE.

Surprisingly, we only observed significant neural correlates with unsigned ADM-predicted AE in the perceptual task (Table S4). The occipital cortex ($x = -16$, $y = -90$, $z = -6$, $Z = 5.12$; $x = 16$, $y = -86$, $z = -4$, $Z = 4.67$; Fig. 7C) showed significant activation corresponding to the unsigned AE in the perceptual task. There was no significant cluster tracking the unsigned AE in the value-based task.

GLM2. The model with behavioral AE

As in GLM1, distinct brain regions were found to track the unsigned IE in the two tasks in GLM with behavioral AE (Table S5). In the perceptual task, the middle temporal gyrus ($x = 52$, $y = -70$, $z = 6$, $Z = 4.86$; Fig. 8A) and middle occipital cortex including angular gyrus ($x = 40$, $y = -78$, $z = 36$, $Z = 4.52$; Fig. 8A) tracked the unsigned IE. Like in GLM1, the ventral striatum ($x = 12$, $y = 12$, $z = -6$, $Z = 5.61$; Fig. 8B) and the ventromedial prefrontal cortex ($x = 2$, $y = 40$, $z = 4$, $Z = 4.51$; Fig. 8B) tracked the unsigned IE in the value-based task.

Common and distinct brain regions tracked the unsigned behavioral AE in the two tasks (Table S6). The cuneus tracked the unsigned behavioral AE in both the perceptual task ($x = -4$, $y = -84$, $z = 34$, $Z = 4.82$; Fig. 9A) and

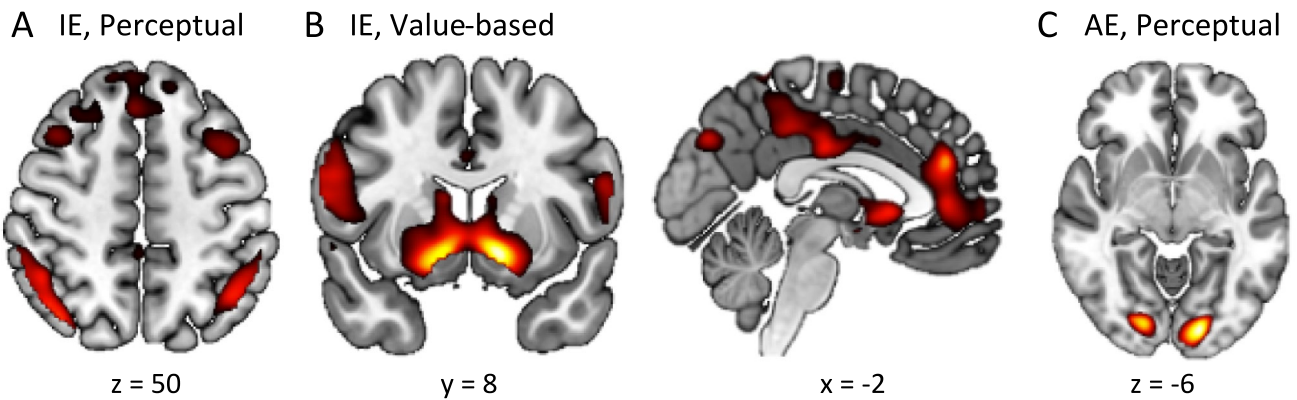


Fig. 7 Brain regions tracking unsigned instantaneous evidence (IE) and average evidence (AE) (GLM1). Brain regions tracking the unsigned IE in the perceptual task (A) and the value-based task (B).

(C) Brain regions tracking the unsigned ADM-predicted AE in the perceptual task. P values were set to .001, uncorrected for visualization. ADM: Averaging Diffusion Model

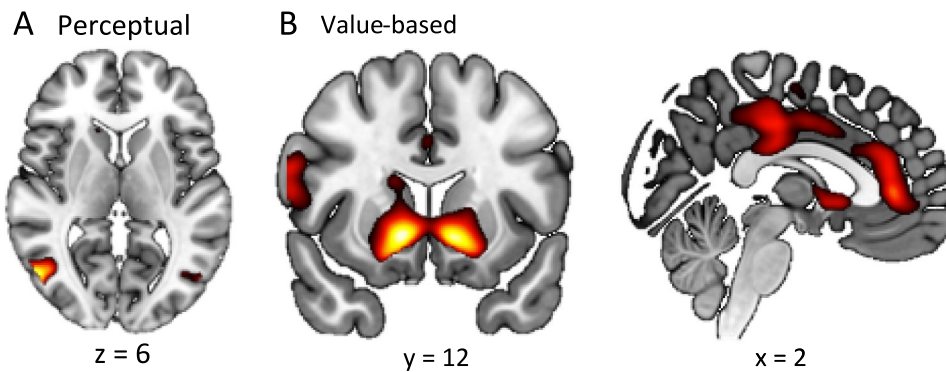


Fig. 8 Brain regions tracking the unsigned instantaneous evidence (IE) (GLM2). Brain regions tracking the unsigned IE in the perceptual task (A) and the value-based task (B). P values were set to .001, uncorrected for visualization

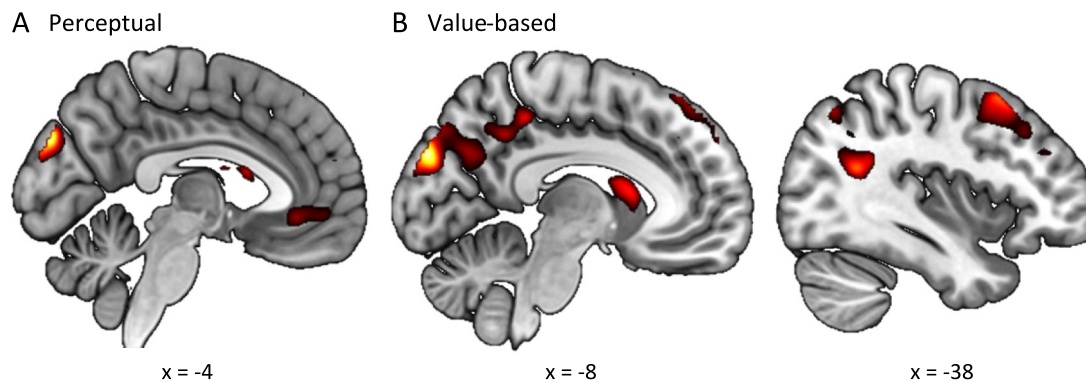


Fig. 9 Brain regions tracking the unsigned behavioral average evidence (AE) (GLM2). Brain regions tracking the unsigned behavioral AE in the perceptual task (A) and the value-based task (B). P values were set to .001, uncorrected for visualization

the value-based task ($x = -8$, $y = -90$, $z = 32$, $Z = 5.23$; Fig. 9B). Additionally, the left dorsolateral prefrontal cortex ($x = -38$, $y = 16$, $z = 56$, $Z = 4.43$; Fig. 9B) tracked the unsigned behavioral AE in only the value-based task.

ROI analysis

We found that the brain regions associated with cognitive control reflected individual differences in temporal bias.

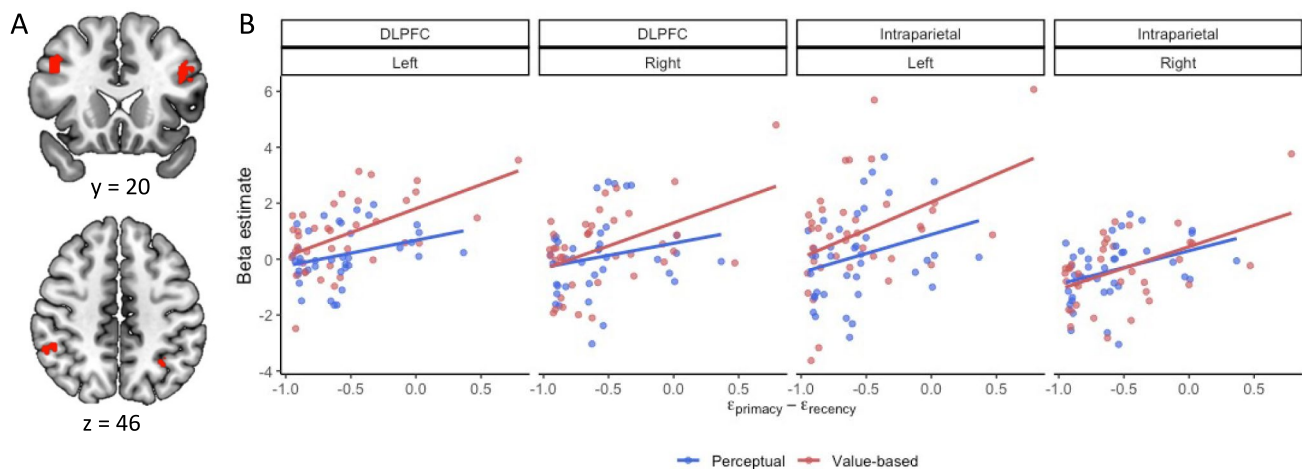


Fig. 10 Region of interest (ROI) analysis results. **(A)** ROIs created from a term-based meta-analysis at Neurosynth. Bilateral DLPFC (top) and intraparietal sulcus (bottom) were chosen as our ROIs. **(B)** The correlation between beta estimates for the stimulus onset regres-

sor in GLM1 and the difference between primacy and recency parameters. Dots represent individual subjects' data and lines represent mean trends from linear regression

Bilateral DLPFC and left intraparietal sulcus were more active at stimulus onset for individuals with a stronger primacy bias in the value-based task (left DLPFC: $r(36)=0.55$, $p<0.001$; right DLPFC: $r(36)=0.47$, $p=0.003$; left intraparietal sulcus: $r(36)=0.43$, $p=0.008$; Fig. 10). In the right intraparietal sulcus, activity at stimulus onset was higher for individuals with a stronger primacy bias in both perceptual and value-based tasks (perceptual $r(36)=0.35$, $p=0.03$; value-based $r(36)=0.05$, $p=0.001$; Fig. 10). This result suggests that individuals with stronger primacy biases recruit more cognitive control when averaging evidence, perhaps to inhibit evidence from later in the stimulus sequence.

In the supporting information, we also report the results of a psychophysiological interaction (PPI) analysis to identify brain regions whose activity depends on the activity of brain regions tracking the unsigned IE.

Discussion

We investigated temporal bias in the evidence-averaging process. In particular, we aimed to understand how temporal bias affects the continuous updating of average evidence over time. We found that people generally exhibit substantial recency bias but only a little primacy bias in their estimates of average evidence. We established these biases in two domains: perceptual and value-based choice. These biases are highly consistent within an individual; people who are more recency-biased in one domain are also more recency-biased in the other. In the value-based, but not perceptual task, we found that the dIPFC tracks the average evidence. Finally, we found that a brain network associated with

cognitive control (dIPFC and intraparietal sulcus) is more active for people who exhibit relatively more primacy bias than recency bias.

Our experiment directly examines temporal biases in information processing, rather than indirectly inferring it from choices (Cheadle et al., 2014; Tsetsos et al., 2011, 2012; Wyart et al., 2012). By having subjects continuously update their estimates on a continuous scale, we could infer their temporal weighting functions from their continuous reports. Our results qualitatively replicate past work, showing the prevalence of recency bias and a small proportion of primacy bias. Our approach allows for a more granular evaluation of the relative impacts of primacy and recency biases in the evidence-averaging process.

With this approach, we identified strong consistency of the temporal weighting function within individuals, across different domains. While the best-fitting model supported different parameters in the two tasks, subjects who showed a strong recency bias in the perceptual task also showed a strong recency bias in the value-based task. The stability of the temporal weighting function indicates that temporal bias is not confined to a specific decision domain. This lends further credence to the idea that perceptual and value-based decisions are governed by overlapping principles (Frydman & Nave, 2017; Glimcher, 2022; Polanía et al., 2014; Shadlen & Shohamy, 2016; Summerfield & Tsetsos, 2012; Smith & Krajbich, 2021).

Although temporal bias is relatively stable within an individual, it does vary substantially across individuals. The majority of subjects showed a strong recency bias, putting more weight on recent evidence in their evaluations. Conversely, a smaller number of subjects showed a strong

primacy bias, putting more weight on early evidence in their evaluations. The prevalence of recency bias in our task may reflect a misapplied adaptive decision-making strategy that only works for prospective judgments. The stimulus sequence in our study was not generated by sampling from a single distribution; it was generated from a random walk. Thus, our stimulus sequence mimics a volatile environment. In such an environment, assigning greater weights to the most recent decision evidence is an adaptive decision strategy when making prospective judgments (Behrens et al., 2007; Hubert-Wallander & Boynton, 2015). It is likely that people are more familiar with making prospective judgments (i.e., predicting the future) and incorrectly apply strategies from that setting to retrospective judgments.

Another reason for recency bias in retrospective judgment may be limited memory capacity. To compute average evidence (AE), a decision maker may try to store individual evidence samples and compute their average. When memory capacity is limited, early evidence might be discarded to store and process the latest evidence. On the one hand, limited memory could be a major issue in our study as we presented long sequences of stimuli. Conversely, subjects in our study could also have avoided the memory burden by using the on-screen slider bar to store prior evidence. The fact that we observed neural correlates of AE, at least in the value-based task, suggests that subjects did mentally track AE and thus could have faced memory constraints, which could have contributed to the pronounced recency bias in our study.

Although primacy opposes recency, they are not mutually exclusive, and a primacy bias may also have its advantages. As a trial progresses, each new sample has a smaller impact on the average. Thus, a decision maker might choose to save cognitive resources by focusing on the stimuli early in a trial and then later inhibiting further updates. Our fMRI results support this explanation. We found that activity in brain regions associated with cognitive control correlate with individual differences in temporal bias. Specifically, bilateral DLPFC and intraparietal sulcus were more active for individuals with a stronger primacy vs. recency bias. This suggests that the primacy biases that we observed were a deliberate strategy on the part of our subjects. This behavior might be analogous to people who fail to update their beliefs once they have formed a strong initial impression (Fourakis & Cone, 2020; Sullivan, 2019).

Primacy bias is also a feature of some sequential sampling models, such as the Ornstein–Uhlenbeck model and the leaky competing accumulator model (Polanía et al., 2014; Usher & McClelland, 2001). There is considerable support for these models (Busemeyer & Townsend, 1993; Roe et al., 2001; Tsetsos et al., 2011; Turner et al., 2018), consistent with the results from our study.

Earlier studies have accounted for temporally weighted evidence-averaging processes with population-coding and

evidence-accumulation frameworks (Brezis et al., 2015; Bronfman et al., 2016; Keung et al., 2020). Brezis & colleagues (2015) propose a dual-component model of numerical averaging in which the number of items to be averaged influences the computational mechanism. They assumed that a greater number of items would lead cognitive agents to use more “intuitive” neurophysiological evidence provided by the population code than an analytic solution. A dynamic leaky competing accumulator model (Bronfman et al., 2016) accounts for the primacy and recency effects as emergent properties of its evidence accumulation mechanism, assuming increasing leakage and decreasing inhibition over time. Leaky inhibitory mechanisms can also implement divisive normalization (Keung et al., 2020).

Both approaches account for the temporal bias as a product of decision processes. For example, population-coding approaches (Brezis et al., 2015, 2016, 2018) can use the decay of neural activations induced by earlier stimuli to realize the recency effect. The Bronfman model does not calculate the arithmetic mean of evidence; instead, the leak and inhibition in the evidence accumulation dynamics determine the relative contribution of evidence from individual stimuli. The Keung model further provides a method to quantify the temporal weighting function directly from the model.

Compared with the Bronfman model, we use a mathematical function that directly calculates the item-wise temporal weight (Pooley et al., 2011). This function does not provide a mechanistic account of temporal weighting (unlike the Bronfman model) but allows us to evaluate and incorporate the item-wise weights seamlessly within the ADM. Some of the aforementioned studies (Brezis et al., 2015; Bronfman et al., 2016) report temporal weighting profiles, but they are calculated from behavioral data in a post-hoc manner.

The temporal weighting scheme in our model (Turner et al., 2017) relies on a process of value averaging that is similar to the mechanism of divisive normalization. Divisive normalization has been proposed as a way to reweight neural activations over time and has been observed in several brain areas (Carandini & Heeger, 2012; Pouget et al., 2013). Many authors have argued that divisive normalization could be performed through the neural computations of inhibition and leakage (Keung et al., 2020; Louie et al., 2014).

Our study also revealed distinct neural mechanisms involved in evaluating the current evidence and tracking the average evidence over time. We found minimal representation of the average evidence signals; aside from the dIPFC in the value-based task, the main result was activity in visual cortex, presumably having to do with the position of the slider bar. However, we found sensible, domain-specific representations of the current evidence in posterior parietal cortex for the perceptual task and in vmPFC and ventral striatum for the value-based task. Parietal lobe is involved in

processing numbers (Leibovich et al., 2017); it shows higher activation when comparing Arabic numbers or the quantities of visual stimuli (Cantlon et al., 2009; Harvey et al., 2013; Kanayet et al., 2014). Thus, the involvement of the parietal lobe is reasonable given that subjects were judging the number of white squares in the display. The vmPFC and ventral striatum are part of the reward network (Bartra et al., 2013; Clithero & Rangel, 2014) and represent subjective value in many decisions, including those based on sequential sampling (Gluth et al., 2012; Hare et al., 2011; PISAURO et al., 2017; Rodriguez et al., 2015). Consistent with these findings, subjects in our study recruited the reward system to evaluate the relative value of the snack foods.

One reason that we may have struggled to find neural representations of AE is the high correlation between IE and AE in the observed responses. We designed the experiment to minimize the correlation between IE and AE, assuming no temporal bias. However, the prevalence of recency bias resulted in a strong correlation between IE and AE in the majority of subjects. The high collinearity between these two regressors may have made it difficult to disentangle the neural signals associated with each process. To address this limitation in future studies, a potential approach would be to adaptively generate sequences for each subject based on their estimated temporal bias, thereby minimizing the correlation between IE and AE (Cavagnaro et al., 2010; Myung & Pitt, 2009).

It is also worth noting that the fMRI results from our model-based GLM1 did not reveal as much as our behavior-based GLM2. As noted above, one reason for this might be the high correlation between IE and AE for many subjects. The noise in behavior might have helped to decorrelate these two measures. An alternative explanation is that our model does not adequately capture subjects' behavior. Comparisons between behavior and our simulated model indicate a generally good fit, but there was of course heterogeneity in how well the model captured each subject's behavior. A poor fit for even a few subjects could be pivotal in an fMRI with a limited sample size.

In our task, participants were asked to explicitly track and report the average evidence over the course of each trial. It is possible that the cognitive and neural mechanisms would have been different if we had only asked subjects to report the average evidence at the end of each trial. Thus, the scope of our findings is limited to cases where people are continuously tracking the average over time, as in our motivating example of the college football rankings.

Retrospective evaluation is an important aspect of our lives. We often look back in time to pinpoint the relationships among events in the past and events in the present. Often, this requires a careful and equal consideration of each event in the past. Our results indicate that people are typically bad at doing this. Most people tend to put too much

weight on recent information, and another smaller group tend to put too much weight on the earliest information (i.e., jumping to conclusions). Both strategies lead to substantial distortions of the truth, distortions that appear to be stable across both perceptual and value-based domains.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.3758/s13415-025-01285-1>.

Author contributions BT and IK conceived the study, all authors designed the experiment, MY programmed the experiment and collected the data, GB contributed analysis tools, MY analyzed the data and wrote the first draft of the paper, and all authors edited the paper.

Funding This study was funded by National Science Foundation CAREER grant 1847603 awarded to Brandon Turner and National Science Foundation grant 233979 to Ian Krajbich.

National Science Foundation, 1847603, 233979

Data availability All behavioral data is available on OSF (<https://osf.io/38ugk/>).

Code availability All code for analysis is available on OSF (<https://osf.io/38ugk/>).

Declarations

Ethics approval The Biomedical Sciences Institutional Review Board at the Ohio State University approved this study (2014H0338).

Consent to participate Written informed consent was obtained from all individual participants included in the study.

Consent for publication All participants provided informed consent for the publication of data.

Conflicts of interests The authors declare no competing interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bahg, G., Evans, D. G., Galdo, M., & Turner, B. M. (2020). Gaussian process linking functions for mind, brain, and behavior. *Proceedings of the National Academy of Sciences*, 117(47), 29398–29406. <https://doi.org/10.1073/pnas.1912342117>
- Bartra, O., McGuire, J. T., & Kable, J. W. (2013). The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments

- examining neural correlates of subjective value. *NeuroImage*, 76, 412–427. <https://doi.org/10.1016/j.neuroimage.2013.02.063>
- Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, 10(9), Article 9. <https://doi.org/10.1038/nn1954>
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433–436. <https://doi.org/10.1163/156856897X00357>
- Brett, M., Anton, J.-L., Valabregue, R., & Poline, J.-B. (2002). Region of interest analysis using an SPM toolbox. *8th International Conference on Functional Mapping of the Human Brain*.
- Brezis, N., Bronfman, Z. Z., Jacoby, N., Lavidor, M., & Usher, M. (2016). Transcranial direct current stimulation over the parietal cortex improves approximate numerical averaging. *Journal of Cognitive Neuroscience*, 28(11), 1700–1713. https://doi.org/10.1162/jocn_a_00991
- Brezis, N., Bronfman, Z. Z., & Usher, M. (2015). Adaptive spontaneous transitions between two mechanisms of numerical averaging. *Scientific Reports*, 5(1), 10415. <https://doi.org/10.1038/srep10415>
- Brezis, N., Bronfman, Z. Z., & Usher, M. (2018). A perceptual-like population-coding mechanism of approximate numerical averaging. *Neural Computation*, 30(2), 428–446. https://doi.org/10.1162/neco_a_01037
- Bronfman, Z. Z., Brezis, N., & Usher, M. (2016). Non-monotonic temporal-weighting indicates a dynamically modulated evidence-integration mechanism. *PLOS Computational Biology*, 12(2), e1004667. <https://doi.org/10.1371/journal.pcbi.1004667>
- Brunton, B. W., Botvinick, M. M., & Brody, C. D. (2013). Rats and humans can optimally accumulate evidence for decision-making. *Science*, 340(6128), 95–98. <https://doi.org/10.1126/science.1233912>
- Bürkner, P.-C., Gabry, J., Kay, M., & Vehtari, A. (2023). *posterior: Tools for Working with Posterior Distributions* (Version R package version 1.5.0) [Computer software]. <https://mc-stan.org/posterior/>.
- Bussemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, 100(3), 432–459. <https://doi.org/10.1037/0033-295X.100.3.432>
- Cakici, N., & Zaremba, A. (2023). Recency bias and the cross-section of international stock returns. *Journal of International Financial Markets, Institutions and Money*, 84, 101738. <https://doi.org/10.1016/j.intfin.2023.101738>
- Cantlon, J. F., Libertus, M. E., Pinel, P., Dehaene, S., Brannon, E. M., & Pelphrey, K. A. (2009). The neural development of an abstract concept of number. *Journal of Cognitive Neuroscience*, 21(11), 2217–2229. <https://doi.org/10.1162/jocn.2008.21159>
- Carandini, M., & Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, 13(1), Article 1. <https://doi.org/10.1038/nrn3136>
- Carney, D. R., & Banaji, M. R. (2012). First Is Best. *PLoS ONE*, 7(6), e35088. <https://doi.org/10.1371/journal.pone.0035088>
- Cavagnaro, D. R., Myung, J. I., Pitt, M. A., & Kujala, J. V. (2010). Adaptive design optimization: A mutual information-based approach to model discrimination in cognitive science. *Neural Computation*, 22(4), 887–905. <https://doi.org/10.1162/neco.2009.02-09-959>
- Cheadle, S., Wyart, V., Tsetsos, K., Myers, N., de Gardelle, V., Hecce Castañón, S., & Summerfield, C. (2014). Adaptive gain control during human perceptual choice. *Neuron*, 81(6), 1429–1441. <https://doi.org/10.1016/j.neuron.2014.01.020>
- Clithero, J. A., & Rangel, A. (2014). Informatic parcellation of the network involved in the computation of subjective value. *Social Cognitive and Affective Neuroscience*, 9(9), 1289–1302. <https://doi.org/10.1093/scan/nst106>
- Deichmann, R., Gottfried, J. A., Hutton, C., & Turner, R. (2003). Optimized EPI for fMRI studies of the orbitofrontal cortex. *NeuroImage*, 19(2), 430–441. [https://doi.org/10.1016/S1053-8119\(03\)00073-9](https://doi.org/10.1016/S1053-8119(03)00073-9)
- Do, J., Eo, K. Y., James, O., Lee, J., & Kim, Y.-J. (2022). The representational dynamics of sequential perceptual averaging. *Journal of Neuroscience*, 42(6), 1141–1153. <https://doi.org/10.1523/JNEUROSCI.0628-21.2021>
- Durand, R. B., Patterson, F. M., & Shank, C. A. (2021). Behavioral biases in the NFL gambling market: Overreaction to news and the recency bias. *Journal of Behavioral and Experimental Finance*, 31, 100522. <https://doi.org/10.1016/j.jbef.2021.100522>
- Fourakis, E., & Cone, J. (2020). Matters order: The role of information order on implicit impression formation. *Social Psychological and Personality Science*, 11(1), 56–63. <https://doi.org/10.1177/1948550619843930>
- Frydman, C., & Nave, G. (2017). Extrapolative beliefs in perceptual and economic decisions: Evidence of a common mechanism. *Management Science*, 63(7), 2340–2352. <https://doi.org/10.1287/mnsc.2016.2453>
- Galdo, M., Weichart, E. R., Sloutsky, V. M., & Turner, B. M. (2022). The quest for simplicity in human learning: Identifying the constraints on attention. *Cognitive Psychology*, 138, 101508. <https://doi.org/10.1016/j.cogpsych.2022.101508>
- Ge, X., Häubl, G., & Elrod, T. (2012). What to say when: Influencing consumer choice by delaying the presentation of favorable information. *Journal of Consumer Research*, 38(6), 1004–1021. <https://doi.org/10.1086/661937>
- Glimcher, P. W. (2022). Efficiently irrational: Deciphering the riddle of human choice. *Trends in Cognitive Sciences*, 26(8), 669–687. <https://doi.org/10.1016/j.tics.2022.04.007>
- Gluth, S., Rieskamp, J., & Buchel, C. (2012). Deciding when to decide: Time-variant sequential sampling models explain the emergence of value-based decisions in the human brain. *Journal of Neuroscience*, 32(31), 10686–10698. <https://doi.org/10.1523/JNEUROSCI.0727-12.2012>
- Hare, T. A., Schultz, W., Camerer, C. F., O'Doherty, J. P., & Rangel, A. (2011). Transformation of stimulus value signals into motor commands during simple choice. *Proceedings of the National Academy of Sciences*, 108(44), 18120–18125. <https://doi.org/10.1073/pnas.1109322108>
- Harvey, B. M., Klein, B. P., Petridou, N., & Dumoulin, S. O. (2013). Topographic representation of numerosity in the human parietal cortex. *Science*, 341(6150), 1123–1126. <https://doi.org/10.1126/science.1239052>
- Heekeren, H. R., Marrett, S., Bandettini, P. A., & Ungerleider, L. G. (2004). A general mechanism for perceptual decision-making in the human brain. *Nature*, 431(7010), 859–862. <https://doi.org/10.1038/nature02966>
- Heekeren, H. R., Marrett, S., Ruff, D. A., Bandettini, P. A., & Ungerleider, L. G. (2006). Involvement of human left dorsolateral prefrontal cortex in perceptual decision making is independent of response modality. *Proceedings of the National Academy of Sciences*, 103(26), 10023–10028. <https://doi.org/10.1073/pnas.0603949103>
- Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, 15(8), 534–539. <https://doi.org/10.1111/j.0956-7976.2004.00715.x>
- Ho, T. C., Brown, S., & Serences, J. T. (2009). Domain general mechanisms of perceptual decision making in human cortex. *Journal of Neuroscience*, 29(27), 8675–8687. <https://doi.org/10.1523/JNEUROSCI.5984-08.2009>
- Hubert-Wallander, B., & Boynton, G. M. (2015). Not all summary statistics are made equal: Evidence from extracting summaries across time. *Journal of Vision*, 15(4), 5. <https://doi.org/10.1167/15.4.5>
- Hurlstone, M. J., Hitch, G. J., & Baddeley, A. D. (2014). Memory for serial order across domains: An overview of the literature and

- directions for future research. *Psychological Bulletin*, 140(2), 339–373. <https://doi.org/10.1037/a0034221>
- Johar, G. V., Jedidi, K., & Jacoby, J. (1997). A varying-parameter averaging model of on-line brand evaluations. *Journal of Consumer Research*, 24(2), 232–247. <https://doi.org/10.1086/209507>
- Juechems, K., Balaguer, J., Ruz, M., & Summerfield, C. (2017). Ventromedial prefrontal cortex encodes a latent estimate of cumulative reward. *Neuron*, 93(3), 705–714.e4. <https://doi.org/10.1016/j.neuron.2016.12.038>
- Kanayet, F. J., Opfer, J. E., & Cunningham, W. A. (2014). The value of numbers in economic rewards. *Psychological Science*, 25(8), 1534–1545. <https://doi.org/10.1177/0956797614533969>
- Kayser, A. S., Buchsbaum, B. R., Erickson, D. T., & D'Esposito, M. (2010). The functional anatomy of a perceptual decision in the human brain. *Journal of Neurophysiology*, 103(3), 1179–1194. <https://doi.org/10.1152/jn.00364.2009>
- Keung, W., Hagen, T. A., & Wilson, R. C. (2020). A divisive model of evidence accumulation explains uneven weighting of evidence over time. *Nature Communications*, 11(1), 2160. <https://doi.org/10.1038/s41467-020-15630-0>
- Kiani, R., Hanks, T. D., & Shadlen, M. N. (2008). Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. *Journal of Neuroscience*, 28(12), 3017–3029. <https://doi.org/10.1523/JNEUROSCI.4761-07.2008>
- Kleiner, M., Brainard, David, & Pelli, Denis. (2007). *What's new in Psychtoolbox-3?* Perception 36 ECV Abstract Supplement.
- Leibovich, T., Katzin, N., Harel, M., & Henik, A. (2017). From “sense of number” to “sense of magnitude”: The role of continuous magnitudes in numerical cognition. *Behavioral and Brain Sciences*, 40. <https://doi.org/10.1017/S0140525X16000960>
- Li, Y., & Epley, N. (2009). When the best appears to be saved for last: Serial position effects on choice. *Journal of Behavioral Decision Making*, 22(4), 378–389. <https://doi.org/10.1002/bdm.638>
- Liu, T., & Pleskac, T. J. (2011). Neural correlates of evidence accumulation in a perceptual decision task. *Journal of Neurophysiology*, 106(5), 2383–2398. <https://doi.org/10.1152/jn.00413.2011>
- Louie, K., LoFaro, T., Webb, R., & Glimcher, P. W. (2014). Dynamic divisive normalization predicts time-varying value coding in decision-related circuits. *Journal of Neuroscience*, 34(48), 16046–16057. <https://doi.org/10.1523/JNEUROSCI.2851-14.2014>
- Mantonakis, A., Rodero, P., Lesschaeve, I., & Hastie, R. (2009). Order in choice: Effects of serial position on preferences. *Psychological Science*, 20(11), 1309–1312. <https://doi.org/10.1111/j.1467-9280.2009.02453.x>
- Metz, N., & Jog, C. (2023). High stakes, experts, and recency bias: Evidence from a sports gambling contest. *Applied Economics Letters*, 30(18), 2525–2529. <https://doi.org/10.1080/13504851.2022.2099517>
- Mohrschladt, H. (2021). The ordering of historical returns and the cross-section of subsequent returns. *Journal of Banking & Finance*, 125, 106064. <https://doi.org/10.1016/j.jbankfin.2021.106064>
- Murdock, B. B., Jr. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, 64(5), 482–488. <https://doi.org/10.1037/h0045106>
- Myung, J. I., & Pitt, M. A. (2009). Optimal experimental design for model discrimination. *Psychological Review*, 116(3), 499–518. <https://doi.org/10.1037/a0016104>
- Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D. A., Conway, A., Cowan, N., Donkin, C., Farrell, S., Hitch, G. J., Hurlstone, M. J., Ma, W. J., Morey, C. C., Nee, D. E., Schweppe, J., Vergauwe, E., & Ward, G. (2018). Benchmarks for models of short-term and working memory. *Psychological Bulletin*, 144(9), 885–958. <https://doi.org/10.1037/bul0000153>
- O'Connell, R. G., Shadlen, M. N., Wong-Lin, K., & Kelly, S. P. (2018). Bridging neural and computational viewpoints on perceptual decision-making. *Trends in Neurosciences*, 41(11), 838–852. <https://doi.org/10.1016/j.tins.2018.06.005>
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4), 437–442.
- Pietsch, A., & Vickers, D. (1997). Memory capacity and intelligence: Novel techniques for evaluating rival models of a fundamental information-processing mechanism. *Journal of General Psychology*, 124(3), 229–339. <https://doi.org/10.1080/00221309709595520>
- Pisauro, M. A., Fouragnan, E., Retzler, C., & Philiastides, M. G. (2017). Neural correlates of evidence accumulation during value-based decisions revealed via simultaneous EEG-fMRI. *Nature Communications*, 8(1), 15808. <https://doi.org/10.1038/ncomms15808>
- Polanía, R., Krajbich, I., Grueschow, M., & Ruff, C. C. (2014). Neural oscillations and synchronization differentially support evidence accumulation in perceptual and value-based decision making. *Neuron*, 82(3), 709–720. <https://doi.org/10.1016/j.neuron.2014.03.014>
- Pooley, J. P., Lee, M. D., & Shankle, W. R. (2011). Understanding memory impairment with memory models and hierarchical Bayesian analysis. *Journal of Mathematical Psychology*, 55(1), 47–56. <https://doi.org/10.1016/j.jmp.2010.08.003>
- Pouget, A., Beck, J. M., Ma, W. J., & Latham, P. E. (2013). Probabilistic brains: Knowns and unknowns. *Nature Neuroscience*, 16(9), 1170–1178. <https://doi.org/10.1038/nn.3495>
- Ratcliff, R. (2006). Modeling response signal and response time data. *Cognitive Psychology*, 53(3), 195–237. <https://doi.org/10.1016/j.cogpsych.2005.10.002>
- Rescorla, R., & Wagner, A. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In *Classical Conditioning II: Current Research and Theory: Vol. 2*.
- Rey, A., Le Goff, K., Abadie, M., & Courrieu, P. (2020). The primacy order effect in complex decision making. *Psychological Research Psychologische Forschung*, 84(6), 1739–1748. <https://doi.org/10.1007/s00426-019-01178-2>
- Rodriguez, C. A., Turner, B. M., Van Zandt, T., & McClure, S. M. (2015). The neural basis of value accumulation in intertemporal choice. *European Journal of Neuroscience*, 42(5), 2179–2189. <https://doi.org/10.1111/ejn.12997>
- Roe, R. M., Busmeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionist model of decision making. *Psychological Review*, 108(2), 370–392. <https://doi.org/10.1037/0033-295x.108.2.370>
- Shadlen, M. N., & Shohamy, D. (2016). Decision making and sequential sampling from memory. *Neuron*, 90(5), 927–939. <https://doi.org/10.1016/j.neuron.2016.04.036>
- Smith, S. M., & Krajbich, I. (2021). Mental representations distinguish value-based decisions from perceptual decisions. *Psychonomic Bulletin & Review*, 28(4), 1413–1422. <https://doi.org/10.3758/s13423-021-01911-2>
- Stan Development Team. (n.d.). *RStan: The R interface to Stan* (Version 2.21.5) [Computer software]. <https://mc-stan.org/>.
- Sullivan, J. (2019). The primacy effect in impression formation: Some replications and extensions. *Social Psychological and Personality Science*, 10(4), 432–439. <https://doi.org/10.1177/1948550618771003>
- Summerfield, C., & Tsetsos, K. (2012). Building bridges between perceptual and economic decision-making: Neural and computational mechanisms. *Frontiers in Neuroscience*, 6, 70. <https://doi.org/10.3389/fnins.2012.00070>
- Sutton, R. S., & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. MIT Press. <http://www.cs.ualberta.ca/~sutton/book/the-book.html>.

- Tong, K., & Dubé, C. (2022). Modeling mean estimation tasks in within-trial and across-trial contexts. *Attention, Perception, & Psychophysics*, 84(7), 2384–2407. <https://doi.org/10.3758/s13414-021-02410-1>
- Tong, K., Dubé, C., & Sekuler, R. (2019). What makes a prototype a prototype? Averaging visual features in a sequence. *Attention, Perception, & Psychophysics*, 81(6), 1962–1978. <https://doi.org/10.3758/s13414-019-01697-5>
- Tsetsos, K., Chater, N., & Usher, M. (2012). Salience driven value integration explains decision biases and preference reversal. *Proceedings of the National Academy of Sciences*, 109(24), 9659–9664. <https://doi.org/10.1073/pnas.1119569109>
- Tsetsos, K., Usher, M., & McClelland, J. L. (2011). Testing multi-alternative decision models with non-stationary evidence. *Frontiers in Neuroscience*, 5. <https://doi.org/10.3389/fnins.2011.00063>
- Turner, B. M., Gao, J., Koenig, S., Palfy, D., & McClelland, L. J. (2017). The dynamics of multimodal integration: The averaging diffusion model. *Psychonomic Bulletin & Review*, 24(6), 1819–1843. <https://doi.org/10.3758/s13423-017-1255-2>
- Turner, B. M., Rodriguez, C. A., Liu, Q., Molloy, M. F., Hoogendijk, M., & McClure, S. M. (2019). On the neural and mechanistic bases of self-control. *Cerebral Cortex*, 29(2), 732–750. <https://doi.org/10.1093/cercor/bhx355>
- Turner, B. M., Schley, D. R., Muller, C., & Tsetsos, K. (2018). Competing theories of multialternative, multiattribute preferential choice. *Psychological Review*, 125, 329–362. <https://doi.org/10.1037/rev0000089>
- Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, 108(3), 550–592. <https://doi.org/10.1037/0033-295x.108.3.550>
- Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27(5), 1413–1432. <https://doi.org/10.1007/s11222-016-9696-4>
- Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2021). Rank-normalization, folding, and localization: An improved \widehat{R} for assessing convergence of MCMC. *Bayesian Analysis*, 16(2). <https://doi.org/10.1214/20-BA1221>
- Watanabe, S. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11(116), 3571–3594.
- Wilmings, N., Murphy, P. R., Meyniel, F., & Donner, T. H. (2020). Large-scale dynamics of perceptual decision information across human cortex. *Nature Communications*, 11(1), Article 1. <https://doi.org/10.1038/s41467-020-18826-6>
- Woo, C.-W., Krishnan, A., & Wager, T. D. (2014). Cluster-extent based thresholding in fMRI analyses: Pitfalls and recommendations. *NeuroImage*, 91, 412–419. <https://doi.org/10.1016/j.neuroimage.2013.12.058>
- Wulff, D. U., Hills, T. T., & Hertwig, R. (2015). Online product reviews and the description–experience gap. *Journal of Behavioral Decision Making*, 28(3), 214–223. <https://doi.org/10.1002/bdm.1841>
- Wulff, D. U., Mergenthaler-Canseco, M., & Hertwig, R. (2018). A meta-analytic review of two modes of learning and the description–experience gap. *Psychological Bulletin*, 144(2), 140–176. <https://doi.org/10.1037/bul0000115>
- Wyart, V., de Gardelle, V., Scholl, J., & Summerfield, C. (2012). Rhythmic fluctuations in evidence accumulation during decision making in the human brain. *Neuron*, 76(4), 847–858. <https://doi.org/10.1016/j.neuron.2012.09.015>
- Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, 8(8), Article 8. <https://doi.org/10.1038/nmeth.1635>
- Yates, J. L., Park, I. M., Katz, L. N., Pillow, J. W., & Huk, A. C. (2017). Functional dissection of signal and noise in MT and LIP during decision-making. *Nature Neuroscience*, 20(9), Article 9. <https://doi.org/10.1038/nn.4611>
- Zylberberg, A., Barttfeld, P., & Sigman, M. (2012). The construction of confidence in a perceptual decision. *Frontiers in Integrative Neuroscience*, 6. <https://doi.org/10.3389/fnint.2012.00079>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.