Humans Learn to Weight Evidence Unevenly Over Time

Hua-Dong Xiong[1]
School of Psychology
Georgia Institute of Technology
hdx@gatech.edu

Li Ji-An[1]
Neurosciences Graduate Program
University of California San Diego
jil095@ucsd.edu

Marcelo G. Mattar[2]
Department of Psychology
New York University
marcelo.mattar@nyu.edu

Robert C. Wilson[2]
School of Psychology
Georgia Institute of Technology
rwilson337@gatech.edu

Abstract

In perceptual decision-making tasks, humans integrate noisy sensory evidence over time to guide their choices. The optimal integration process assumes that all evidence is weighted equally within a trial and that different trials are independent. However, humans exhibit systematic deviations from optimality, including uneven weighting of evidence within trials and influences from previous trials. Prior studies have demonstrated that biological constraints can account for this suboptimality. In this study, we present evidence that humans adapt their evidence integration strategies over time in response to task demands, and that the suboptimal uneven weighting is gradually learned over the course of the task. By explicitly modeling this adaptation through online gradient-based learning, our model outperforms existing approaches in capturing human behavior and unifies both observed forms of suboptimality in the Click task: dependence across trials emerges from an error-driven learning process that also gives rise to uneven integration weights within trials. We further propose a bounded-rational adaptation account to explain why humans progressively learn to weight evidence unevenly within a trial.

Our modeling framework provides a general approach of resource-rational adaptation. It captures how initially uninformed agents can gradually update their strategies through error-driven learning and is applicable to a broad range of learning and decision-making scenarios.

Keywords: perceptual decision-making; computational modeling; meta-learning; online gradient-based learning; bounded-rational adaptation

Introduction

To navigate an uncertain world, organisms must make accurate decisions by integrating noisy sensory evidence over time. In perceptual decision-making, evidence integration models have successfully explained many aspects of human and animal behavior. These models typically assume that decision-makers employ statistically optimal strategies, assigning equal weight to all evidence within a trial and treating each trial independently. They also assume that these strategies remain fixed throughout the experiment. However, biological decision-makers must learn their strategies through experience and adapt them continuously based on task feedback—a dynamic process often overlooked by traditional computational models.

Consider the widely studied Click task (Brunton, Botvinick, & Brody, 2013; Keung, Hagen, & Wilson, 2019), where participants hear a sequence of clicks presented to either the left or right ear and report the side with more clicks. Classical models of behavior in this task assume a strategy that integrates clicks with equal weights on each side, enabling decisions based on cumulative evidence. These optimal strategies have explained key qualitative aspects of human and animal behavior and have revealed neural correlates of evidence integration (Scott et al., 2017; Brunton et al., 2013). However, both humans and animals also exhibit systematic deviations from optimality in evidence integration (Lau & Glimcher, 2005; Scott, Constantinople, Erlich, Tank, & Brody, 2015; Keung et al., 2019). For example, studies using the Click task have identified two key forms of suboptimal behavior: individuals tend to weight evidence unevenly within single trials (intra-trial suboptimality) and allow prior trials to influence current decisions (inter-trial suboptimality). Although separate mechanisms have been proposed to explain these effects—such as memory drift or divisive normalization for uneven weighting (Brunton et al., 2013; Keung, Hagen, & Wilson, 2020), and win-stay-lose-shift strategies for trial history effects (Keung et al., 2019)—a unified framework explaining how these suboptimalities arise and interact remains lacking. Moreover, existing explanations often attribute these behaviors to "non-optimal" heuristics, rather than grounding them in a normative analysis.

To address this gap, we introduce a novel framework to model the online task adaptation process and the resulting temporal suboptimalities in the Click task. Our framework models across-trial dependencies as an error-driven learning process, in which evidence integration weights are updated via gradient-based learning informed by task feedback. This model explicitly captures the learning dynamics and shows that uneven evidence weighting emerges naturally from this process, thereby accounting for both within- and between-trial suboptimalities in behavior. Inspired by these findings, we offer a bounded-rational adaptation account of uneven integration in the Click task as an interaction between trial-difficulty modulation, diminishing information gain, and a gating mechanism. Our results provide a computational explanation for temporal dependencies in decision-making and suggest that behavioral patterns previously viewed as suboptimal may instead reflect adaptive learning processes.

[1]Co-first author
[2]Co-senior author

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

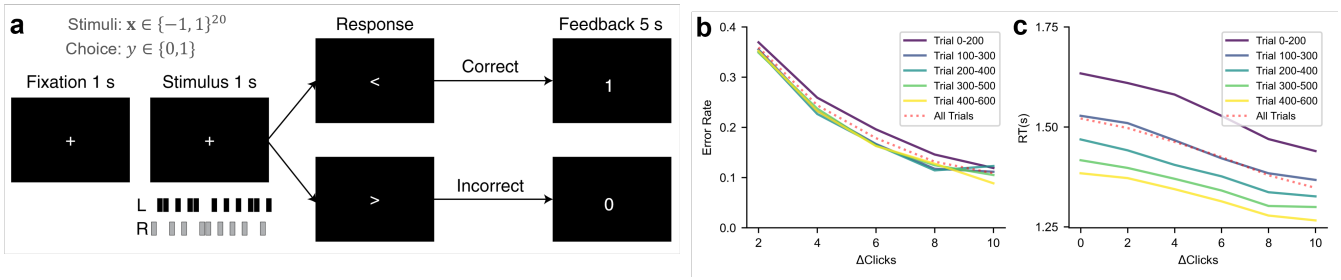

Figure 1: **a,** Click task. Participants hear a one-second sequence of 20 clicks presented to either the left (L, black bars) or right ear (R, gray bars) and decide which side had more clicks. After responding, an arrow indicated their choice, followed by feedback (1 for correct, 0 for incorrect). **b,** The error rate (proportion of incorrect choices) decreases as the experiment progresses and increases with trial difficulty (measured by $|\Delta\text{Click}|$, the difference in the number of clicks between left and right sides; easier trials have larger $|\Delta\text{Click}|$). **c,** Reaction time decreases as the experiment progresses and increases with trial difficulty.

Methods

Click task

To investigate suboptimalities in human evidence integration across timescales, we analyzed the behavior of 186 participants performing the Bernoulli Click Task (Fig. 1a). Portions of these data have been partially reported in previous studies (Keung et al., 2019, 2020). Each trial began with a fixation period, followed by a sequence of $S = 20$ rapid clicks delivered every 50 ms over one second to either the left or right ear. Participants wore headphones and indicated which ear received more clicks by pressing a key for "left" or "right"; failure to respond resulted in a timeout. After the response, an arrow on the screen displayed their choice for 500 ms, followed by performance feedback.

Click stimuli were generated via a Bernoulli process. On each trial, one side (left or right) was pseudo-randomly designated as the high-probability side. Each click was then generated as $x \sim \text{Bern}(0.55)$, where $x = +1$ indicates a click on the left and $x = -1$ indicates a click on the right. The high-probability side was counterbalanced across trials. The correct response was determined by which side received more clicks; if both sides received the same number, either response was accepted as correct. Participants were not informed about the stimulus generation process and only received feedback about the correct side after responding.

Participants continued the task until they reached either 500 correct responses or a 50-minute time limit. For consistency, we analyzed only the first 600 trials per participant. Trials with reaction times exceeding three standard deviations above the participant's mean were excluded.

Behavioral Models

Below, we describe all models using consistent notation. These models share two key components: evidence integration and action selection.

Evidence integration. On trial t, participant n ($1 \le n \le N = 186$) observes a sequence of clicks $\mathbf{x}_t^{(n)} = \left(x_{t,1}^{(n)}, x_{t,2}^{(n)}, \ldots, x_{t,S}^{(n)}\right)^\top \in \{-1,+1\}^S$, where -1 denotes a click on the left and $+1$ a click on the right. Each trial contains $S = 20$ clicks. We assume the participant integrates these clicks using a weight vector $\mathbf{w}_t^{(n)} = \left(w_{t,1}^{(n)}, w_{t,2}^{(n)}, \ldots, w_{t,S}^{(n)}\right)^\top \in \mathbb{R}^S$. The integrated evidence is computed as a weighted sum: $z_t^{(n)} = \sum_{s=1}^{S} w_{t,s}^{(n)} x_{t,s}^{(n)} = \mathbf{w}_t^{(n)\top} \mathbf{x}_t^{(n)}$.

Action selection. All models use a logistic function to determine the probability of choosing the right side. Specifically, the probability that participant n chooses the right side on trial t is given by:

$$p_t^{(n)} = \frac{1}{1 + \exp\left(-z_t^{(n)}\right)}. \tag{1}$$

We fit all models by minimizing the negative log-likelihood between observed human behavioral choices and predicted choice probabilities. Let $c_t^{(n)} \in \{0,1\}$ denote the actual choice of participant n on trial t. The fitting loss is defined as

$$\mathcal{L}_{\text{fit}} = -\sum_{t=1}^{T}\sum_{n=1}^{N}\left[c_t^{(n)} \log\left(p_t^{(n)}\right) + \left(1 - c_t^{(n)}\right) \log\left(1 - p_t^{(n)}\right)\right]. \tag{2}$$

This loss is minimized with respect to the relevant model parameters, which vary depending on the specific models, as described below.

Logistic Regression Logistic regression is a baseline model without an online learning component. Each participant applies a fixed integration weight across all trials, i.e., $\mathbf{w}_t^{(n)} \equiv \mathbf{w}^{(n)}$ for all t. The weight vectors $\{\mathbf{w}^{(n)}\}_{n=1}^N$ are directly optimized to minimize the loss $\mathcal{L}_{\text{fit}}$ (see Eq. 2).

Gradient-Based Meta-Learning To model how humans adapt their evidence integration strategies over time, we adopt a model-agnostic meta-learning framework (Finn, Abbeel, & Levine, 2017). We implement an error-driven (gradient-based) online learning process that explicitly captures participants' adaptation during the task: they continuously adjust their integration strategies via gradient descent. This framework comprises inner and outer learning loops. The inner loop

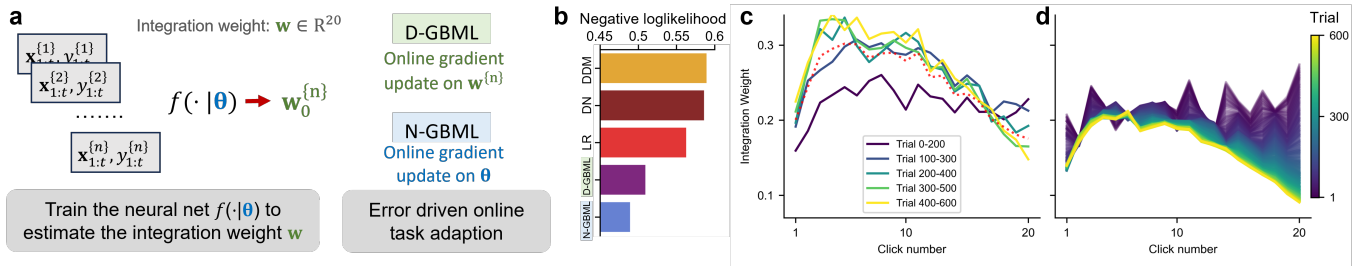

Figure 2: **Gradient-Based Meta-Learning model (GBML). a**, Diagram of GBML (n: participant. t: trial index). **b**, Model Comparison. Our GBML model achieved better performance (lower negative log-likelihood indicates a better fit) using nested cross-validation compared to the drift diffusion model (DDM), divisive normalization (DN), and logistic regression (LR). **c,** Click integration weights estimated by logistic regression, averaged within overlapping windows of consecutive trials and across subjects. **d,** Click integration weights estimated by GBML for each trial, averaged across subjects.

updates a subset of parameters on each trial, corresponding to fast weights that capture the participant's online learning process. The outer loop updates the remaining parameters to fit overall behavioral patterns.

On each trial t, the model computes an online prediction error using the cross-entropy loss between the predicted probability $p_t^{(n)}$ (based on the participant's choice $c_t^{(n)}$) and the correct response $y_t^{(n)} \in \{0,1\}$ (i.e., the side with more clicks). For participant n, the online loss is defined as:

$$\mathcal{L}_{\text{online}}^{(n)} = -y_t^{(n)} \log\left(p_t^{(n)}\right) - \left(1-y_t^{(n)}\right) \log\left(1-p_t^{(n)}\right). \quad (3)$$

Our two model variants differ in the parameters they update (Fig. 2a).

D-GBML Our first model, Direct Gradient-Based Meta-Learning (D-GBML), updates the integration weight vector for participant n on each trial based on the gradient of the online loss:

$$\mathbf{w}_{t+1}^{(n)} \leftarrow \left(1-\lambda^{(n)}\right)\mathbf{w}_t^{(n)} - \alpha^{(n)} \nabla_{\mathbf{w}_t^{(n)}} \mathcal{L}_{\text{online}}^{(n)}, \quad (4)$$

where $\alpha^{(n)} \in \mathbb{R}$ is the learning rate and $\lambda^{(n)} \in \mathbb{R}$ is the weight decay rate. In the outer loop, we optimize $\{\alpha^{(n)}\}_{n=1}^N$, $\{\lambda^{(n)}\}_{n=1}^N$, and the initial weight vectors $\{\mathbf{w}_0^{(n)}\}_{n=1}^N$ to minimize the overall loss $\mathcal{L}_{\text{fit}}$.

N-GBML While D-GBML directly updates the integration weights $\mathbf{w}_t^{(n)}$ via gradient descent, this approach may be too restrictive to capture more complex learning dynamics and relationships among click weights. To address this limitation, we introduce Neural Network Gradient-Based Meta-Learning (N-GBML), which uses a neural network to parameterize the integration weights.

For each participant n on trial t, the weight vector is computed as:

$$\mathbf{w}_t^{(n)} = \text{ReLU}\left(\mathbf{W}_t^{\text{in}} \mathbf{e}^{(n)} + \mathbf{b}_t^{\text{in}}\right) \mathbf{W}^{\text{out}} + \mathbf{b}^{\text{out}}, \quad (5)$$

where $\mathbf{W}_t^{\text{in}} \in \mathbb{R}^{H \times D}$ and $\mathbf{b}_t^{\text{in}} \in \mathbb{R}^H$ are the input weights and biases, updated on each trial; $\mathbf{W}^{\text{out}} \in \mathbb{R}^{H \times S}$ and $\mathbf{b}^{\text{out}} \in \mathbb{R}^S$ are the output weights and biases, fixed across trials; $\mathbf{e}^{(n)} \in \mathbb{R}^D$ is the participant-specific embedding; H is the hidden dimensionality; and D is the embedding dimensionality.

Only the input parameters $\theta_t^{\text{in}} = \{\mathbf{W}_t^{\text{in}}, \mathbf{b}_t^{\text{in}}\}$ are updated on each trial using the gradient of the online loss:

$$\theta_{t+1}^{\text{in}}[d] \leftarrow \left(1-\lambda^{(d)}\right)\theta_t^{\text{in}}[d] - \alpha^{(d)} \nabla_{\theta_t^{\text{in}}[d]} \mathcal{L}_{\text{online}}^{(n)}, \quad (6)$$

where $\alpha^{(d)} \in \mathbb{R}$ is the learning rate, $\lambda^{(d)} \in \mathbb{R}$ is the weight decay rate, and d indexes the D dimensions. In the outer loop, we optimize α, λ, the initial input parameters θ_0^{in}, and the output parameters $\mathbf{W}^{\text{out}}$ and $\mathbf{b}^{\text{out}}$ to minimize the overall choice loss.

Results

Humans adapt their behaviors to the task over time

We first examined whether humans learn evidence integration strategies online during the Click task. Trials were grouped by difficulty level, defined as the absolute difference in the number of clicks between the left and right ($|\Delta\text{Click}|$). A sliding window of length 200 was applied, and a windowed average was computed over time. Analysis of the performance of 186 participants revealed that both error rate (Fig. 1b) and reaction time (Fig. 1c) decreased over time, indicating improvement in both speed and accuracy throughout the experiment. These behavioral trends suggest that participants adapt their strategies over the course of the task.

Gradient-based online learning models outperform competing models

To test whether across-trial dependencies in human behavior can be explained by online gradient-based learning, we compared our models (D-GBML and N-GBML) with several established baselines: the drift-diffusion model (DDM), the divisive normalization (DN) model (Keung et al., 2019, 2020), and the logistic regression (LR) model. Our two GBML models significantly outperformed all baselines in terms of negative log-likelihood, as evaluated using nested cross-validation (Fig. 2b). Moreover, the N-GBML model outperformed the D-GBML model, reflecting its greater computational flexibility.

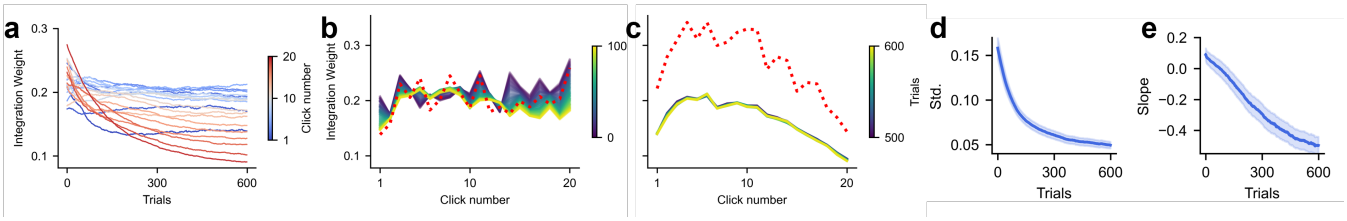

Figure 3: **Emergence of uneven integration weights**. **a,** Trial-by-trial changes in integration weights estimated by N-GBML. Early clicks (i.e., those with smaller click numbers) show slightly increasing weights across trials, while later clicks exhibit decreasing weights. **b, c**, Integration weights estimated by logistic regression (red dashed lines) and N-GBML for the first 100 trials (**b**) and the last 100 trials (**c**). **d,** Standard deviation of integration weights across trials, computed per participant and averaged to quantify individual-level variability. **e,** Normalized slope of the *average* integration weights across trials, indicating trends in group-average integration weights. Shaded areas indicate the 95% confidence interval.

These results suggest that incorporating trial-by-trial online learning via gradient-based weight updates more accurately captures the dynamics of human decision-making. Given its superior predictive performance, we used the N-GBML model for all subsequent analyses.

Emergence of uneven integration weights over time

Next, we examined how integration weights evolved over trials by comparing estimates from logistic regression (Fig.2c) and N-GBML (Fig.2d, 3a). Early in the experiment, integration weights were relatively uniform; as learning progressed, they converged toward increasingly uneven - a pattern robustly captured by both models.

To better understand these temporal changes, we visualized the average integration weights for the first and last 100 trials using both logistic regression and N-GBML (Fig. 3b,c). Most weight changes occurred during early trials, after which the integration profile gradually stabilized. Initially, weights were relatively flat (Fig. 3b), but over time developed a pronounced bump-like pattern (Fig. 3c). Although this increasing unevenness may seem counterintuitive given the observed behavioral improvement (Fig. 1b,c), it reflects a transition from noisy, inconsistent early strategies to more stable, individually biased ones. This interpretation is supported by a decline in within-subject variability over time (Fig. 3d). To quantify the emergence of this uneven pattern, we measured the slope of the group-average integration weights and found that it became increasingly negative over time (Fig. 3e).

Overall, we found that while individuals improved their performance on the task, a consistent bias shared across participants gradually emerged, leading to increasingly uneven integration weights as a result of online task adaptation. Why and how do humans learn this uneven, bump-like integration pattern (Fig. 2d)? We address these questions in the following sections.

Bump-shaped updates in integration weights

Given that our N-GBML model captures the gradient-based learning process, we analyzed the averaged trial-by-trial changes of the integration weight vector ($\Delta \mathbf{w}_t = \mathbf{w}_t - \mathbf{w}_{t-1}$) for both correct and incorrect trials (Fig. 4), grouped by trials with the same difficulty.

We consider three analyses by averaging the integration weights corresponding to (i) all click positions(Fig. 4a,b); (ii) click positions on the chosen side (Fig. 4c,d); and (iii) click positions on the unchosen side (Fig. 4e,f). Note that the analyses (ii) and (iii) reflect the change in integration weights for stimuli presented on either the chosen or unchosen side of the trial (e.g., if a participant correctly chooses the left, weight updates are measured for left-side stimuli). While not all click positions contributed to weight updates in every trial, we plotted the weight changes as a function of all click positions.

We found that integration weights averaged over all click positions are approximately zero (Fig. 4a,b), suggesting that the click positions on the chosen side and on the unchosen side produce opposite effects, effectively cancelling each other. In correct trials, the averaged weights for click positions on the chosen side increased (Fig. 4c), while those for the unchosen side decreased (Fig. 4e). This pattern is consistent with a "win-stay, lose-shift" strategy in updating integration weights: when decisions are correct, participants amplify the evidential weights supporting their choice and suppress those contradicting it. The pattern in incorrect trials is similar: weights for clicks on the chosen side decrease (Fig. 4d), while those for the unchosen side increase (Fig. 4f).

Further, we observed difficulty-modulated updates to the integration weights: weight updates were larger for harder trials (i.e., smaller $|\Delta\text{Click}|$) than for easier ones. In addition, the characteristic bump-shaped pattern of weight updates is most pronounced in the most difficult trials.

A bounded-rational adaptive account of uneven bump-shaped integration weights

Our previous results show that the uneven, bump-shaped integration weights observed in human perceptual decision-making may arise from an online learning process driven by errors (i.e., online loss gradients). But why do participants manifest uneven integration weights instead of an optimal uniform weighting strategy?

We interpret the emergence of uneven integration weights through the lens of bounded-rational adaptation, following the principles of resource-rational analysis (Lieder & Griffiths,

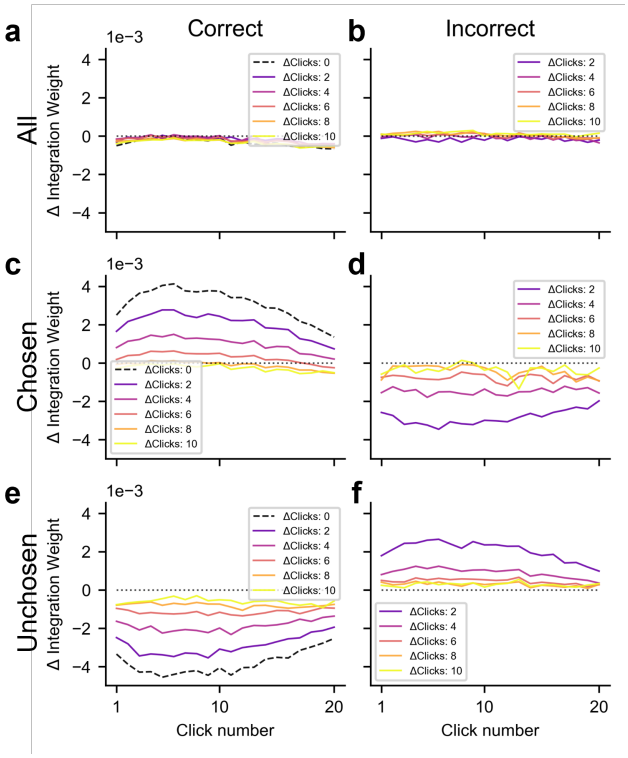

Figure 4: **Updates in integration weights.** Trial-by-trial changes in integration weights for correct (**a,c,e**) and incorrect (**b,d,f**) trials, grouped by trial difficulty. Panels show integration weight changes averaged across all click indices (**a,b**), click indices on the chosen side (**c,d**), and click indices on the unchosen side (**e,f**).

2020). Rather than assuming that agents compute the globally optimal solution during the task, this framework posits that behavior reflects the best strategy achievable under limited resources—such as time, noise, and computational capacity constrained by low-level neurobiology. These constraints impose costs that shape how learning unfolds, often leading to heuristic or partially optimized solutions acquired through prior learning and reused across tasks.

To explain the observed bump kernel (Fig. 4), we suggest that there might be three sources of constraints: a functional constraint from the error-driven updates in the inner loop, a time constraint from speed-accuracy tradeoff, and a gating mechanism subject to biological constraints.

Functional constraint. Trials have different levels of difficulty d, directly modulating the magnitude of weight updates (in Eq. 4 and 6). In easier trials (larger $|\Delta\text{Click}|$), the predicted probability for the correct action is closer to 1, leading to smaller (cross-entropy) error signals and weight updates. In contrast, in more difficult trials, the probability for the correct action is closer to 0.5, leading to larger error signals and weight updates.

Time constraint. Intuitively, due to the constraint of limited time and the diminishing marginal information offered by each click, subjects may feel the urgency to decide quickly based on limited evidence collected. Such a decision may lead to fewer rewards, but effectively save time and mental efforts for making future decisions. We note that while the Click task does not have explicit time pressure (Fig. 1), such speed-accuracy tradeoff (Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006) is universal in perceptual decision-making mechanisms, plausibly shaped by evolution and development.

Formally, we model this using a Beta–Bernoulli model. Since the clicks in each trial are sampled from a Bernoulli distribution, the subject's goal for each trial is to estimate the Bernoulli parameter p. Assume that the subject's prior for p before click x_i is described by $\text{Beta}(\alpha_i, \beta_i)$, summarizing past clicks in the current trial. The information gain IG_i due to the observation of click x_i, is defined as the expected reduction in uncertainty (i.e., the mutual information between p and x_i, in which x_i is marginalized) (Houlsby, Huszár, Ghahramani, & Lengyel, 2011):

$$IG_i = H\left(\mathbb{E}[z]\right) - \mathbb{E}_{z\sim\text{Beta}(\alpha_i,\beta_i)}[H(z)], \tag{7}$$

where $p = \mathbb{E}[z] = \frac{\alpha_i}{\alpha_i+\beta_i}$ is the predictive mean before observing x_i and $H(p) = -p\log p - (1-p)\log(1-p)$ is the entropy of the Bernoulli distribution. This quantity measures the epistemic value of click x_i by quantifying how much it is expected to reduce uncertainty in the posterior. Notably, IG_i decreases as i increases, reflecting diminishing information provided by later clicks.

Biological constraint. Evidence integration in the brain is implemented by neuronal circuits that are subject to biophysical constraints. Rather than acting as perfect integrators, these circuits are shaped by excitatory-inhibitory dynamics. We propose that, as in other perceptual decision processes, click integration relies on a disinhibitory circuit motif common in the cortex (Yang, Murray, & Wang, 2016; Wang & Yang, 2018), where dendritic gating is controlled by interneuron cascades. Specifically, peridendrite-targeting interneurons are tonically active before trial onset, suppressing noise by blocking excitatory inputs. A trial-onset control signal activates interneuron-targeting interneurons, which in turn inhibit the gating cells, thereby disinhibiting dendritic integration. This mechanism gradually opens the integration gate during stimulus presentation, effectively increasing the weight of later clicks. We capture this effect using a sigmoidal gating function m_i:

$$m_i = \sigma\left(\tau \cdot i - c\right), \tag{8}$$

where $\sigma(\cdot)$ is the sigmoid function, τ and c control the steepness and timing of gate opening, respectively.

Integration weight updates. Taken together, these constraints produce an adaptive update strategy that prioritizes earlier signals for rapid decisions while avoiding improper responses to task-irrelevant noise — a bounded-rational adaptation to the specific characteristics and demands of the perceptual decision-making task. Formally, the integration weight update for click i at trial t can be modeled as:

$$\Delta \mathbf{w}_{t,i} \propto IG_i \cdot m_i \cdot \frac{\partial \mathcal{L}_{\text{online}}^{\text{ideal}}}{\partial \mathbf{w}_{t,i}}, \tag{9}$$

where IG_i is the information gain for i-th click, m_i is the gating dynamics, and $\frac{\partial \mathcal{L}_{\text{online}}^{\text{ideal}}}{\partial \mathbf{w}_{t,i}}$ is the ideal, uniform weight updates modulated by trial difficulty d_t.

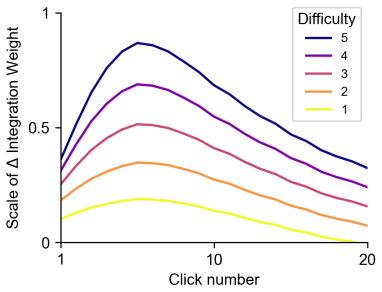

Figure 5: **Interaction between trial difficulty, information gain, and a gating mechanism produces bump-shaped weight updates.** Colors indicate trial difficulty, corresponding to the magnitude of the error-driven weight update.

With suitable parameters (sharpness τ and onset timing c), we can qualitatively match our resource rational account, modulated by trial difficulty, early gating dynamics, and late diminishing information gain, to the online learning dynamics discovered by N-GBML (comparing Fig. 5 to Fig. 4). Thus, the observed bump-shaped weight updates, and the resulting uneven integration weights, emerge as a bounded-rational strategy adapted to the resource constraints and computational demands faced by biological organisms.

Discussion

In this study, we introduced a computational framework to explain how humans adapt evidence integration strategies over time in perceptual decision-making. Rather than assuming fixed strategies or attributing "suboptimalities" to behavioral biases, we modeled decision-making as an error-driven, online learning process. Our GBML model accounts for both intra-trial and inter-trial suboptimalities within a unified and principled framework. Moreover, we show that uneven, bump-shaped integration weights—often labeled suboptimal—emerge naturally from a rational adaptation process shaped by cognitive and computational constraints. These findings suggest that the temporal structure of evidence weighting is not a violation of optimality, but rather a signature of rational learning under bounded cognitive resources.

Our work makes a technical contribution by introducing a general method for inferring explicit cognitive constraints from behavioral data. Specifically, we parameterize these constraints directly within the model's integration weight update rules. This flexible approach allows researchers to explore diverse forms of decision-making and learning under well-defined hypotheses. In our study, different model variants instantiate distinct cognitive assumptions. D-GBML posits that integration weights are updated via error gradients, potentially corresponding to plasticity within neural circuits responsible for evidence integration. In contrast, N-GBML uses a neural network to generate integration weights, with the network itself trained via gradient descent. This architecture could reflect plasticity in upstream neural circuits that modulate evidence-integration mechanisms.

Inspired by insights from N-GBML, we propose that trial-by-trial learning is modulated by three interacting constraints: information gain, which quantifies how much each click reduces uncertainty; trial-difficulty modulation, which scales the magnitude of error signals; and gating dynamics, which suppress irrelevant noise. Together, these constraints give rise to a bump-shaped profile of weight updates, with the largest updates occurring mid-trial and during more difficult trials. As learning progresses, participants' integration weights converge toward this uneven pattern, reflecting rational adaptation under cognitive constraints. Thus, what might appear as a behavioral bias can instead be understood as an efficient, adaptive strategy for allocating limited computational resources.

In addition, our framework can be interpreted through the lens of control-as-inference (Todorov, 2008; Levine, 2018), which recasts decision-making as probabilistic inference over latent optimal actions or policies. From this perspective, participants adaptively update their policy—in this case, the integration weights—to increase the posterior likelihood of selecting actions aligned with inferred task goals. In our model, online gradient-based updates to the integration weights correspond to posterior inference steps, modulated by three resource-rational signals. This formulation captures how agents under control-as-inference allocate learning effort to moments of high uncertainty or informativeness. Thus, the observed non-uniform, bump-shaped integration profiles do not reflect deviations from optimality, but rather approximate solutions to the problem of inferring latent control policies under temporally structured uncertainty.

More broadly, our findings offer a general perspective on decision-making as an online, gradient-based learning process. This framework enables more precise hypotheses about the cognitive and neural mechanisms that shape adaptive behavior, and it may align with neural signals—such as phasic arousal or neuromodulatory gain—that modulate task-dependent learning. Beyond perceptual decision-making, the model can be extended to other domains, including sequence learning and memory, to investigate how internal representations evolve over time. Ultimately, this approach may support the development of integrated models of cognition that unify learning, inference, and adaptation in dynamic environments.

Acknowledgments

This work was supported by start-up funding from the Georgia Institute of Technology awarded to RCW. We also acknowledge the use of the Partnership for an Advanced Computing Environment (PACE) at the Georgia Institute of Technology, which provided essential computational resources for this research.

References

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, *113*(4), 700–765. Retrieved 2024-05-09, from https://doi.apa.org/doi/10.1037/0033-295X.113.4.700 doi: 10.1037/0033-295X.113.4.700

Brunton, B. W., Botvinick, M. M., & Brody, C. D. (2013, April). Rats and Humans Can Optimally Accumulate Evidence for Decision-Making. *Science*, *340*(6128), 95–98. Retrieved 2025-01-13, from https://www.science.org/doi/10.1126/science.1233912 doi: 10.1126/science.1233912

Finn, C., Abbeel, P., & Levine, S. (2017, July). Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. In *Proceedings of the 34th International Conference on Machine Learning* (pp. 1126–1135). PMLR. Retrieved 2024-11-13, from https://proceedings.mlr.press/v70/finn17a.html (ISSN: 2640-3498)

Houlsby, N., Huszár, F., Ghahramani, Z., & Lengyel, M. (2011, December). *Bayesian Active Learning for Classification and Preference Learning*. arXiv. Retrieved 2025-05-11, from http://arxiv.org/abs/1112.5745 (arXiv:1112.5745 [stat]) doi: 10.48550/arXiv.1112.5745

Keung, W., Hagen, T. A., & Wilson, R. C. (2019, June). Regulation of evidence accumulation by pupil-linked arousal processes. *Nature Human Behaviour*, *3*(6), 636–645. Retrieved 2022-12-08, from http://www.nature.com/articles/s41562-019-0551-4 doi: 10.1038/s41562-019-0551-4

Keung, W., Hagen, T. A., & Wilson, R. C. (2020, May). A divisive model of evidence accumulation explains uneven weighting of evidence over time. *Nature Communications*, *11*(1), 2160. Retrieved 2023-09-15, from https://www.nature.com/articles/s41467-020-15630-0 (Number: 1 Publisher: Nature Publishing Group) doi: 10.1038/s41467-020-15630-0

Lau, B., & Glimcher, P. W. (2005). Dynamic Response-by-Response Models of Matching Behavior in Rhesus Monkeys. *Journal of the Experimental Analysis of Behavior*, *84*(3), 555–579. Retrieved 2025-01-25, from https://onlinelibrary.wiley.com/doi/abs/10.1901/jeab.2005.110-04 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1901/jeab.2005.110-04) doi: 10.1901/jeab.2005.110-04

Levine, S. (2018, May). *Reinforcement Learning and Control as Probabilistic Inference: Tutorial and Review*. arXiv. Retrieved 2023-02-18, from http://arxiv.org/abs/1805.00909 (arXiv:1805.00909 [cs, stat]) doi: 10.48550/arXiv.1805.00909

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*, e1. Retrieved 2022-11-07, from https://www.cambridge.org/core/product/identifier/S0140525X1900061X/type/journal_article doi: 10.1017/S0140525X1900061X

Scott, B. B., Constantinople, C. M., Akrami, A., Hanks, T. D., Brody, C. D., & Tank, D. W. (2017, July). Fronto-parietal Cortical Circuits Encode Accumulated Evidence with a Diversity of Timescales. *Neuron*, *95*(2), 385–398.e5. Retrieved 2025-01-14, from https://www.cell.com/neuron/abstract/S0896-6273(17)30511-1 (Publisher: Elsevier) doi: 10.1016/j.neuron.2017.06.013

Scott, B. B., Constantinople, C. M., Erlich, J. C., Tank, D. W., & Brody, C. D. (2015, December). Sources of noise during accumulation of evidence in unrestrained and voluntarily head-restrained rats. *eLife*, *4*, e11308. Retrieved 2025-01-25, from https://doi.org/10.7554/eLife.11308 (Publisher: eLife Sciences Publications, Ltd) doi: 10.7554/eLife.11308

Todorov, E. (2008, December). General duality between optimal control and estimation. In *2008 47th IEEE Conference on Decision and Control* (pp. 4286–4292). Retrieved 2025-04-30, from https://ieeexplore.ieee.org/abstract/document/4739438 (ISSN: 0191-2216) doi: 10.1109/CDC.2008.4739438

Wang, X.-J., & Yang, G. R. (2018, April). A disinhibitory circuit motif and flexible information routing in the brain. *Current Opinion in Neurobiology*, *49*, 75–83. Retrieved 2025-05-12, from https://www.sciencedirect.com/science/article/pii/S0959438817302490 doi: 10.1016/j.conb.2018.01.002

Yang, G. R., Murray, J. D., & Wang, X.-J. (2016, September). A dendritic disinhibitory circuit mechanism for pathway-specific gating. *Nature Communications*, *7*(1), 12815. Retrieved 2025-05-12, from https://www.nature.com/articles/ncomms12815 (Publisher: Nature Publishing Group) doi: 10.1038/ncomms12815