RNNs reveal a new optimal stopping rule in sequential sampling for decision-making

Jialin Li
NYU Grossman School of Medicine
New York, NY 10016
jl16057@nyu.edu

Kenway Louie
NYU Grossman School of Medicine
New York, NY 10016
kl837@nyu.edu

Paul Glimcher*
NYU Grossman School of Medicine
New York, NY 10016
paulg@nyu.edu

Bo Shen*
NYU Grossman School of Medicine
New York, NY 10016
bo.shen@nyu.edu

Abstract

Sequential sampling is a commonly studied decision-making environment where evidence accumulates over time until a decision boundary is reached. Normative theories of optimal stopping largely address homogeneous streams of evidence, where each time step carries the same amount of information. However, evidence from natural environments is often heterogeneous, with the informativeness of evidence varying over time; the optimal stopping rule under such conditions remains unknown. We trained recurrent neural networks (RNNs) to make decisions when receiving heterogeneous evidence streams and considering sampling costs and a time constraint. In addition to replicating classic collapsing boundaries for stopping under a time constraint, we found a novel early commitment effect: the RNN adopts a lower decision boundary in the earliest time steps of decision-making. Normative analysis validated such a strategy as optimal. Examination of model policies showed that early commitment and collapsing boundary were driven by distinct mechanisms associated with sampling cost and time constraint, respectively. By bridging artificial networks and normative analysis, our work identifies early commitment as an optimal policy for decision-making in naturalistic environments.

1 Introduction

Decision making is often modeled as a dynamic process where evidence is accumulated over time towards a fixed boundary to inform choices (Gold and Shadlen, 2007; Ratcliff, 1978; Wald, 1947). In the standard drift diffusion model (DDM) for this process, such a decision boundary is assumed to be fixed. Recent findings, however, suggest that humans and animals do not use constant decision boundaries (Churchland et al., 2008; Cisek et al., 2009; Drugowitsch et al., 2012; Forstmann et al., 2010; Hawkins et al., 2015). Normative models support this conclusion, revealing that the decision boundary 'collapses' to a lower evidentiary threshold at long deliberation; agents trade off between the benefits of accumulating more evidence against the potential costs of running out of time (Bogacz et al., 2006; Tajima et al., 2016, 2019). In addition to this well-tested phenomenon, lower thresholds for stopping may also occur at the very early time of decision. For example, people are more willing to commit to a choice with "good enough" evidence in the early stage of decision-making when the cost (investment) they have spent is small. We term this phenomenon the *early commitment* effect; decision-makers commit to a choice at a lower evidentiary threshold at the very early time of decision.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: CogInterp: Interpreting Cognition in Deep Learning Models.

One possible reason that the *early commitment* effect may have been overlooked is a disparity between the temporal characteristics of evidence used in laboratory tasks and real-world settings. Most laboratory paradigms employ homogeneous streams of evidence, where each time step provides a consistent amount of information (with fewer exceptions: Drugowitsch et al., 2016; Kira et al., 2015; Knowlton et al., 1994; Yang and Shadlen, 2007; Yates et al., 2017). In contrast, naturalistic environments often involve heterogeneous streams of evidence, where the informativeness of sample varies over time. The optimal stopping policy under heterogeneous evidence remains poorly understood.

Here, we investigate how recurrent neural networks (RNNs) learn time-varying decision thresholds when receiving heterogeneous inputs. We find that these well-trained networks, in addition to implementing a collapsing threshold, develop a policy of early commitment—adopting lower decision thresholds in the earliest time periods of decision-making. Normative analysis using dynamic programming validates this additional early strategy as optimal. Analyzing the policies RNNs learn by examining hidden units uncovers that the network adjusts its decision policies in a time-varying manner. Distinct mechanisms drive the early commitment and the collapsing threshold strategies.

In summary, this work identifies early commitment as a novel optimal decision policy when receiving heterogeneous streams of evidence. The findings highlight the importance understanding optimal decision strategies in naturalistic environments for both artificial systems and biological brains.

2 Methods

2.1 Sequential sampling decision-making task

We trained RNNs (GRU architecture, 64 hidden units, Fig. 1A; see details in Appendix A.1) (Cho et al., 2014) to perform a sequential decision-making task. In this task (Fig. 1B), a hidden target (*A* or *B*) was randomly set as the correct response at the beginning of each trial. The stimulus that provided evidence about which target was correct was then drawn from a distribution of possible signals associated with that target (Fig. 1E), one at a time with replacement, and provided to the RNN (Fig. 1A). The RNN updated its recurrent dynamics upon receiving each new stimulus. The agent's goal was to infer the correct target to select in order to receive a reward (r). Since each stimulus contains a different amount of information about the targets (Figs. 1C-E), it results in varying amounts of evidence from each sampling, consisting of a heterogeneous evidence stream.

At each time step, the agent obtains $r = +1$ for a correct decision or $r = 0$ if wrong; otherwise, the agent continues sampling a new piece of evidence at a cost of c. To test how different levels of sampling costs and time constraints affect the decision strategies of RNNs, we trained the RNNs under different sampling cost levels (i.e., 0.01, 0.02, 0.03, 0.04, 0.05). These levels of sampling costs were tested under two different time-constraint conditions. When there is no time constraint (Environment 1), the agent is allowed to sample for an infinite number of steps. When under a time constraint (Environment 2), there is a 10-time-step limit for the RNN to sample information and make choices; out of time would result in zero reward and still paying the sampling costs in the past steps.

2.2 Training algorithm

The RNNs were trained as reinforcement-learning agents on-policy using an advantage actor–critic (A2C) algorithm. The actor produced a policy over the three possible actions (choosing A, choosing B, or continuing sampling), while the critic estimated the expected value of the current state (see details in Appendix A.2). Training objectives combined policy optimization, value prediction, and an entropy regularizer to promote exploration. Models were optimized with standard backpropagation and subsequently evaluated on held-out testing trials with frozen network weights.

3 Results

3.1 Overall performance of RNNs

Across two environments, the agent achieved comparable levels of choice accuracy and average return (Table 1 and 2, Fig. S1C and Fig. S2C in Appendix A.3) with greater difficulties, compared to monkey performance on similar task (85% and 80%) (Kira et al., 2015). The reaction times of the agents resulted in chronometric curves typically observed in humans and animals (Fig. S1A,

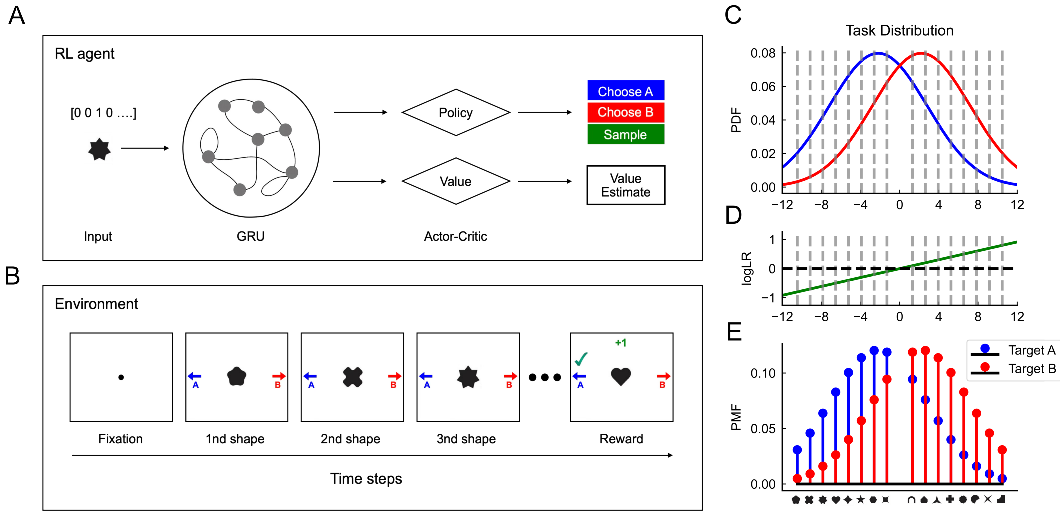

Figure 1: The RNN and the task setup. (A) The RL agent of a 64-unit GRU was trained using an A2C algorithm. At every time step, the agent received one out of sixteen stimuli using one-hot vector coding, with the actor output providing actions and a critic output providing value expectation. (B) The sequential sampling task. Each episode included a single trial. The agent received one evidentiary stimulus at a time drawn from the latent target distribution. The episode was terminated when the agent chose one of two targets. (C) The stimuli were generated from two Gaussian distributions, N(-2.2, 5) (target A, blue) and N(2.2, 5) (target B, red). (D) Sixteen discrete stimuli were sampled from the Gaussian distributions with log-likelihood ratios (logLRs) uniformly spaced between -0.8 and 0.8. (E) The frequencies and logLRs of sixteen discrete stimuli were visualized.

Fig. S2A). The choice probability curves (Fig. S1B, Fig. S2B) showed psychometric curves across cumulative evidence levels. The decision weight associated with each stimulus impact on the agents' binary choices exhibited linear curves that captured the true log-likelihood ratios (logLRs) (Fig. S1B; Appendix A.6.1), indicating a near-optimal accumulation of evidence.

3.2 Early commitment under sampling cost

We found that the agents adopted a time-varying decision threshold even at the earliest times for decision-making. By visualizing the distributions of logLRs of the evidence when the agents decided to choose A, choose B, or continue sampling (blue, red, and green outlines, respectively, in Fig. 2A, D), we observed an *early commitment* effect from both environments: the agents tend to commit to a choice using more liberal criteria at the earlier time steps of decision (Red and blue ticks as the mean values of LogLRs for the chosen trials in Figs. 2A, D). Notably, monkeys show the same pattern in the early periods of decision-making (Kira et al., 2015). By changing the sampling costs, the early commitment effect was robustly preserved in both time-constraint environments (Fig. 2B, E).

3.3 Collapsing boundary under time constraint

It is widely acknowledged that time constraints lead to collapsing of decision boundaries (Tajima et al., 2016, 2019). Our RNNs replicated this effect when trained under a time constraint (Fig. 2D, E), but not in the environment without a time constraint (Fig. 2A, B). Specifically, the agents collapse the decision threshold in the later stages of decision-making.

3.4 Normative validations

Why do RNNs adopt this decision strategy? To answer this question, we performed a normative analysis within a resource-rational framework (Lieder and Griffiths, 2020) using dynamic programming to derive the optimal policy in the current task setting with sampling costs, time constraints, and *heterogeneous evidence stream* (see Appendix A.5). In contrast to the assumption of classic theory (Wald, 1947) (Appendix A.4), we found that both the early commitment and the collapsing boundary strategies implemented by the RNNs appeared in the optimal solutions (Appendix A.5, Fig. S2).

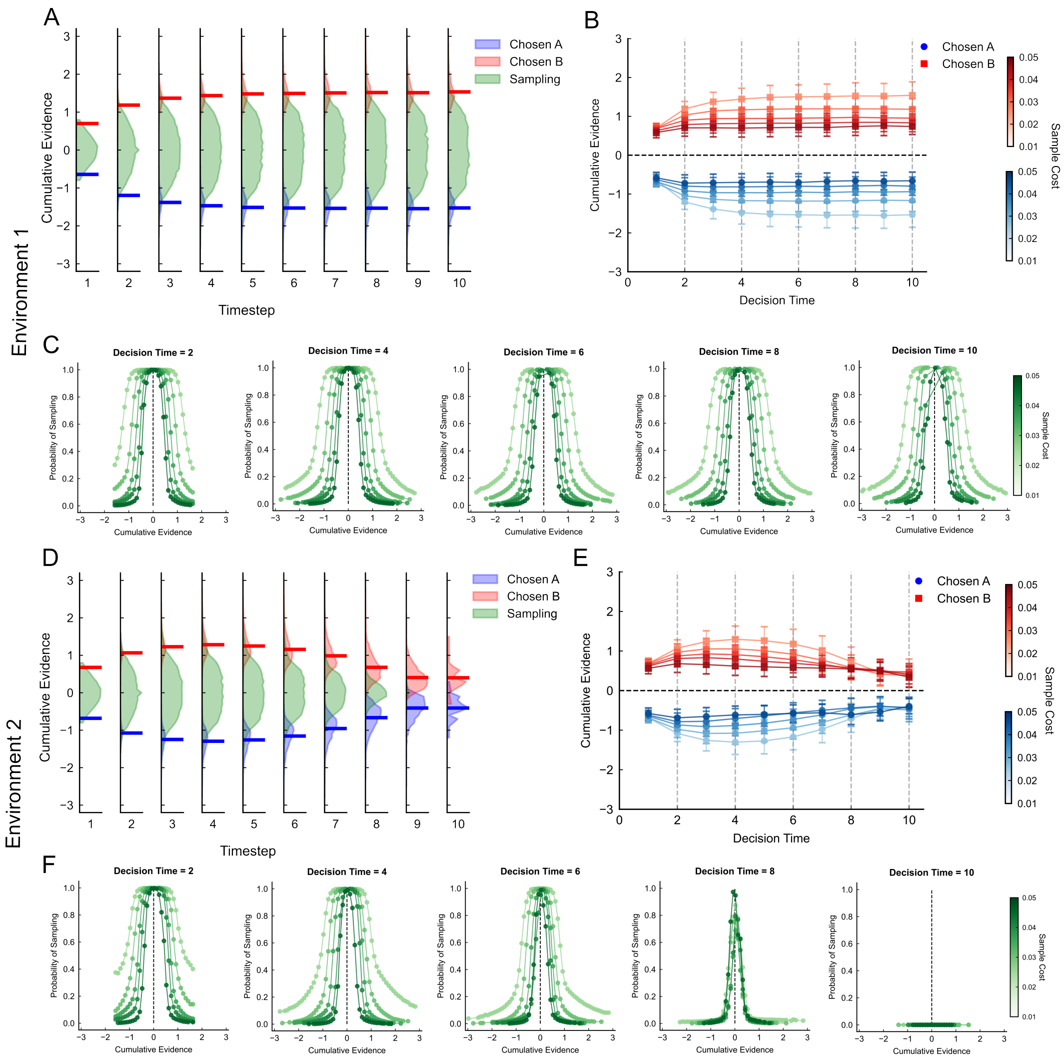

Figure 2: Networks learn time-varying decision thresholds and policies under different time constraints and sampling costs (A) The model trained under a small sampling cost ($c = -.01$) without a time constraint showed an early commitment effect in the early stage of decision, but no collapsing boundary at the late stage. Outlines showed the distributions of LogLRs in the trials of chosen A (blue), chosen B (red), and sampling (green). The blue and red ticks indicated the mean values of LogLRs of the distributions, i.e., the classical definition of decision boundary. (B) Without a time constraint, the mean values of decision boundaries for chosen A (blue) and chosen B (red) uncovered robust early commitment across levels of sampling costs (graded colors). It exhibited an decreasing early commitment effect by lowering decision thresholds when increasing sampling cost (C) The probabilities of sampling encoded in the RNN policies were aggregated according to the cumulative evidence logLRs. The early commitment effect were driven by the sharpness of the sampling curves (D) Similar to (A) but with with a time constraint at time step 10, the RNNs exhibited collapsing boundary at the later time steps in addition to early commitment. (E) Similar to (B) but with a time constraint of 10 time steps, the RNN showed robust mixtures of collapsing boundaries and early commitment across levels of sampling costs. (F) The collapsing boundary were driven by shrinking and flattening the sampling curve to ensure that the decision is made.

When the environment posed sampling cost but no time constraints, the optimal decision boundary showed an early commitment effect, where lower criteria were used in the earlier time steps (Fig. S2A). Analysis revealed that this was due to the trade-off between a time-varying marginal expected gain of sampling and the cost of sampling (Appendix A.5). This reflects a fact that a heterogeneous evidence stream allows the amount of evidence to jump above a "good enough" level with even very few samples in the earliest steps, leading to more liberal criteria early on in a trial. The early commitment effect became less prominent when the sampling cost increased, due to the lowered thresholds overall in the later time steps, consistent with the RNN's strategy (Fig. S2A; Fig. 2B).

By explicitly adding the urgency costs to mimick the time constraint setting in environment 2, we replicated the collapsing boundaries (Fig. S2B). Higher urgency costs lead to steeper collapse in the later stages - a well-documented phenomenon in the literature (Tajima et al., 2016, 2019).

3.5 Distinct mechanisms between early commitment and collapsing boundary

We further analyzed the policies π_θ learned by the RNNs over time steps, indicating the inferred probabilities over three possible actions: choosing A, choosing B, and continuing sampling. The probability of sampling p_{sampling} exhibited a bell-shaped dependence on the cumulative logLRs of the evidence. This pattern was consistently observed across two environments and persisted over temporal stages over the decision-making process. In general, the RNN agents were more likely to stop sampling and commit to a choice when the evidence was larger (Figs. 2C, F). However, the shapes of the curves were "soft", indicating a stochastic, instead of deterministic, decision strategy. In other words, even the exact same magnitude of cumulative logLRs could lead to either a choice or continued sampling in a probabilistic manner.

By comparing the policies learned across different environments, we found that early commitment and collapsing boundaries were driven by distinct mechanisms. Increasing sampling costs reduced the width of the bell-shaped curves, indicating that the agent adopted lower p_{sampling} across all levels of evidence (see narrower boundaries under higher sampling costs in Fig. 2C, colored in green gradients; such a pattern remained across different time steps). The reduction was greater in larger evidence logLRs. This indicates a trade-off between the cost of sampling and the marginal gain from sampling, which more likely leads to stopping sampling when the evidence is "good enough". Given the heterogeneous environment, "good enough" evidence would show up unevenly in the earlier stages with very few samples, which causes the early commitment effect we observed here.

In contrast, under a time constraint, p_{sampling} decreased over time, regardless of the sampling costs. This indicated a separate mechanism from the sampling cost (Fig. 2F). Moreover, the impact of the time constraint was time-varying. In the earlier time steps, the agent gradually shrinked the width of the bell shapes over time; when approaching the end of the trial, the agent exhibited a large decrease in p_{sampling} by flattening the bell shape curves to ensure making a choice within the time limit. Further PCA analysis of the RNNs found a time-coding component in their hidden states, suggesting an adaptive strategy adopted by the networks regarding the time constraints (Fig. S4).

4 Conclusion

The current study investigates how recurrent neural networks integrate heterogeneous evidence sequentially to inform a decision. Our findings reveal early commitment as a novel strategy to maximize reward when receiving heterogeneous streams of evidence, in addition to the classical collapsing boundaries under time constraints. These findings show significance in understanding decision strategies in naturalistic settings for both artificial systems and biological brains.

5 Limitations

The current implementation of normative analysis is based on an assumption that the agent has one-step myopic computation on the future reward predictions instead of a backward induction on the values. Although many studies show that myopic computation converges with the values of backward induction, we may further compare the potential differences. In addition, the time constraint during RNNs training was set as a hard boundary at the 10th time step; whereas, the time cost during dynamic programming was explicitly set as a cost increasing over time steps. Further analysis is needed to better interpret the connection between neural network behavior and normative analysis.

Acknowledgments

We thank Marcelo Mattar and the other members of the Paul Glimcher lab for feedback on this paper. We declare no competing interests.

Code availability

The codes are available through an open source license at `https://github.com/JiaLinLi0822/proj-rnn-sprt`. We trained the models with CPUs on Slurm using pytorch. You may also train the models on a standard computer with enough RAM to support the parallel training.

References

Bhui, R., Lai, L., and Gershman, S. J. (2021). Resource-rational decision making. *Current Opinion in Behavioral Sciences*, 41:15–21.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., and 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.

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation.

Churchland, A. K., Kiani, R., and Shadlen, M. N. (2008). Decision-making with multiple alternatives. *Nature Neuroscience*, 11(6):693–702.

Cisek, P., Puskas, G. A., and El-Murr, S. (2009). Decisions in changing conditions: The urgency-gating model. *Journal of Neuroscience*, 29(37):11560–11571.

Drugowitsch, J., Moreno-Bote, R., Churchland, A. K., Shadlen, M. N., and Pouget, A. (2012). The cost of accumulating evidence in perceptual decision making. *Journal of Neuroscience*, 32(11):3612–3628.

Drugowitsch, J., Wyart, V., Devauchelle, A.-D., and Koechlin, E. (2016). Computational precision of mental inference as critical source of human choice suboptimality. *Neuron*, 92(6):1398–1411.

Forstmann, B. U., Anwander, A., Schäfer, A., Neumann, J., Brown, S., Wagenmakers, E.-J., Bogacz, R., and Turner, R. (2010). Cortico-striatal connections predict control over speed and accuracy in perceptual decision making. *Proceedings of the National Academy of Sciences*, 107(36):15916–15920.

Gold, J. I. and Shadlen, M. N. (2007). The Neural Basis of Decision Making. *Annual Review of Neuroscience*, 30(1):535–574.

Hawkins, G. E., Forstmann, B. U., Wagenmakers, E.-J., Ratcliff, R., and Brown, S. D. (2015). Revisiting the Evidence for Collapsing Boundaries and Urgency Signals in Perceptual Decision-Making. *Journal of Neuroscience*, 35(6):2476–2484.

Jensen, K. T. (2024). An introduction to reinforcement learning for neuroscience.

Kingma, D. P. and Ba, J. (2017). Adam: A method for stochastic optimization.

Kira, S., Yang, T., and Shadlen, M. N. (2015). A neural implementation of wald's sequential probability ratio test. *Neuron*, 85(4):861–873.

Knowlton, B. J., Squire, L. R., and Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning and Memory*, 1(2):106–120.

Lieder, F. and Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, 43:e1.

Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., Silver, D., and Kavukcuoglu, K. (2016). Asynchronous methods for deep reinforcement learning. In *Proceedings of the 33rd International Conference on Machine Learning*, pages 1928–1937. PMLR.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85(2):59–108.

Tajima, S., Drugowitsch, J., Patel, N., and Pouget, A. (2019). Optimal policy for multi-alternative decisions. *Nature Neuroscience*, 22(9):1503–1511.

Tajima, S., Drugowitsch, J., and Pouget, A. (2016). Optimal policy for value-based decision-making. *Nature Communications*, 7(1):12400.

Wald, A. (1947). Foundations of a general theory of sequential decision functions. *Econometrica*, 15(4):279–313.

Yang, T. and Shadlen, M. N. (2007). Probabilistic reasoning by neurons. *Nature*, 447(7148):1075–1080.

Yates, J. L., Park, I. M., Katz, L. N., Pillow, J. W., and Huk, A. C. (2017). Functional dissection of signal and noise in MT and LIP during decision-making. *Nature Neuroscience*, 20(9):1285–1292.

A Appendix

A.1 Model architecture

The model consisted of an input layer, 64 gated recurrent units (GRU) (Cho et al., 2014), an actor head, and a critic head. At each time step, after receiving the input $\boldsymbol{x}$ of stimuli encoded as a one-hot vector, the GRU updates its hidden state according to:

$$\boldsymbol{r}_t = \sigma(\boldsymbol{W}_r \boldsymbol{x}_t + \boldsymbol{U}_r \boldsymbol{h}_{t-1} + \boldsymbol{b}_r) \tag{1}$$

$$\boldsymbol{z}_t = \sigma(\boldsymbol{W}_z \boldsymbol{x}_t + \boldsymbol{U}_z \boldsymbol{h}_{t-1} + \boldsymbol{b}_z) \tag{2}$$

$$\tilde{\boldsymbol{h}}_t = \tanh(\boldsymbol{W}_h \boldsymbol{x}_t + \boldsymbol{U}_h \left(\boldsymbol{r}_t \odot \boldsymbol{h}_{t-1}\right) + \boldsymbol{b}_h) \tag{3}$$

$$\boldsymbol{h}_t = (1 - \boldsymbol{z}_t) \odot \boldsymbol{h}_{t-1} + \boldsymbol{z}_t \odot \tilde{\boldsymbol{h}}_t \tag{4}$$

Then, the hidden states $\boldsymbol{h}_t$ are projected into the actor and critic using a single linear layer, respectively. The actor function $\theta(\cdot)$ produces the policy π_θ, indicating the inferred probabilities over three possible action (choosing A, choosing B, or continuing sampling). The critic function $\phi(\cdot)$ estimates the expected value of the current state.

A.2 Training algorithm

To maximize reward, we trained the reinforcement learning(RL) agent on-policy using the advantage actor-critic (A2C) algorithm (Mnih et al., 2016; Jensen, 2024). The objective function followed the typical form was used to do backpropagation:

$$J(\theta, \phi) = \mathbb{E}_t \left[\underbrace{\log \pi_\theta(\boldsymbol{a}|s_t) A_t}_{\text{actor}} + \underbrace{\beta_v \left(R_t - V_\phi(s_t)\right)^2}_{\text{critic}} + \underbrace{\beta_e \, \mathcal{H}(\pi(\cdot|s_t))}_{\text{entropy}} \right] \tag{5}$$

where π_θ is the decision policy, s_t is the current state given the observations seen so far at step t, $\boldsymbol{a}$ are the possible actions, $A_t = R_t - V_\phi(s_t)$ is the advantage function between the actual reward R_t and the reward expectation $V_\phi(s_t)$. The policy (actor) term encourages the network to take actions to maximize returns. The value term (critic) trains the network to predict the amount of return. The entropy term encourages exploration behavior to prevent the network from being trapped in local minima.

During training, the auxiliary loss coefficients were initially set to $\beta_v = 0.05$ and $\beta_e = 0.05$. To progressively reduce the contribution of the auxiliary entropy-related term relative to the policy loss, β_e was annealed from its initial value to 1×10^{-3} over the first half of training. Concurrently, the learning rate was decayed from 1×10^{-3} to 1×10^{-5} in order to facilitate a finer exploration of the gradient. All models were trained for 1.0×10^7 episodes (trials) with a batch size of 128, employing the Adam optimizer (Kingma and Ba, 2017). After training, we froze the network weights and analyzed the models' behaviors based on 5×10^4 testing trials.

A.3 Behavioral metrics of the RNNs trained under different environments

Table 1: Average return of RNNs under different environments with varied sampling costs

Sample cost	0.01	0.02	0.03	0.04	0.05
Environment 1	0.8730	0.7948	0.7383	0.6981	0.6706
Environment 2	0.8290	0.7722	0.7294	0.6926	0.6658

Table 2: Decision accuracy of RNNs under different environments with varied sampling costs

Sample cost	0.01	0.02	0.03	0.04	0.05
Environment 1	0.9627	0.9215	0.8770	0.8390	0.8085
Environment 2	0.8863	0.8656	0.8431	0.8193	0.7907

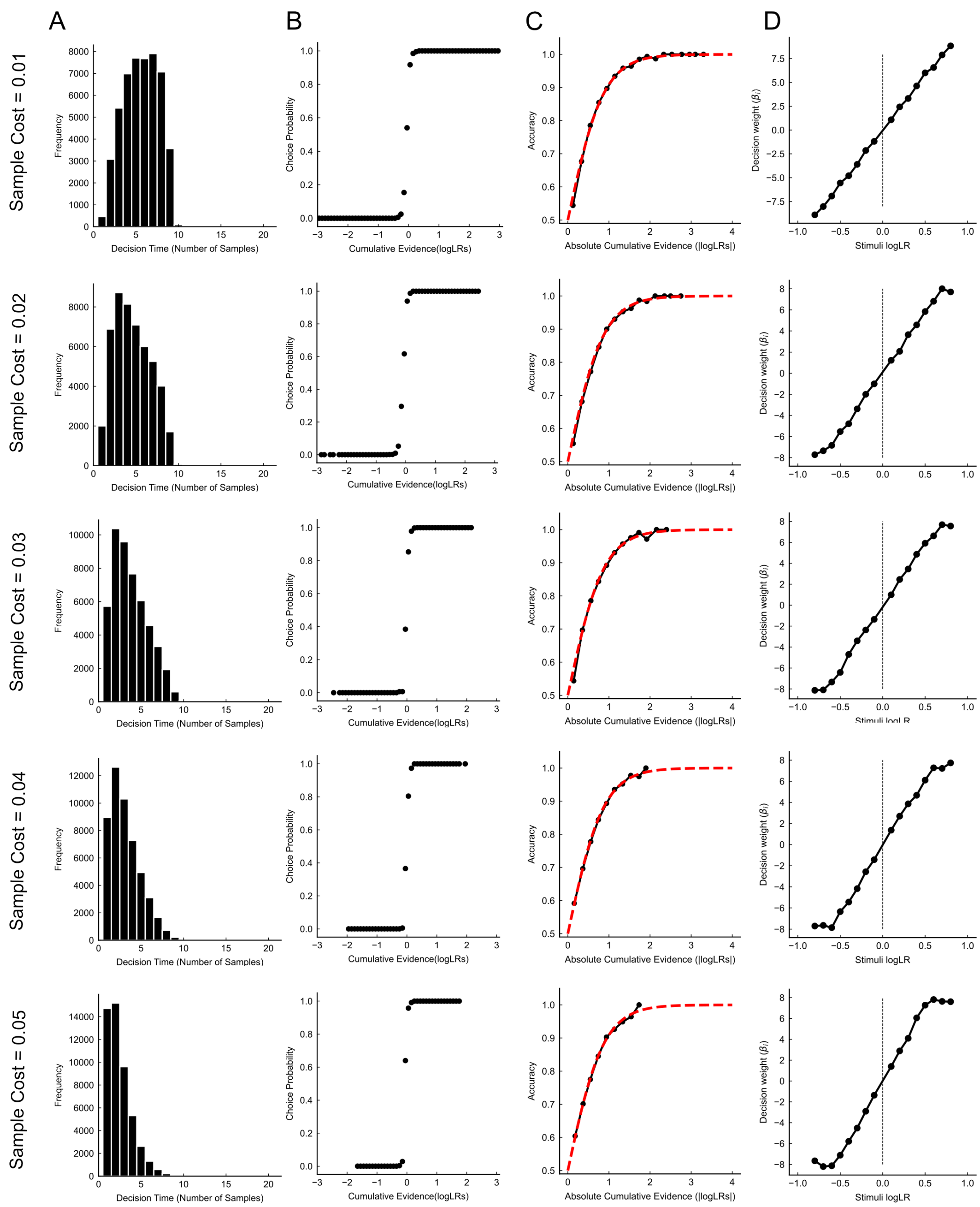

Figure S1: **Behaviors Metrics of the RNNs trained in Environment 1 with different sampling cost** (A) Models showed typical positive skewed distributions of decision times. (B)The psychometric curves of choice probability as a function of the logLRs of the cumulative evidence. (C) The psychometric curves of choice accuracy as a function of the logLRs of the cumulative evidence. Red lines indicate the theoretical optimum. (D) The models' decision weights on the sixteen stimuli varying in the amount of information showed quasi-linear curves between the ground truth logLRs and the decision weights contributed to binary choices.

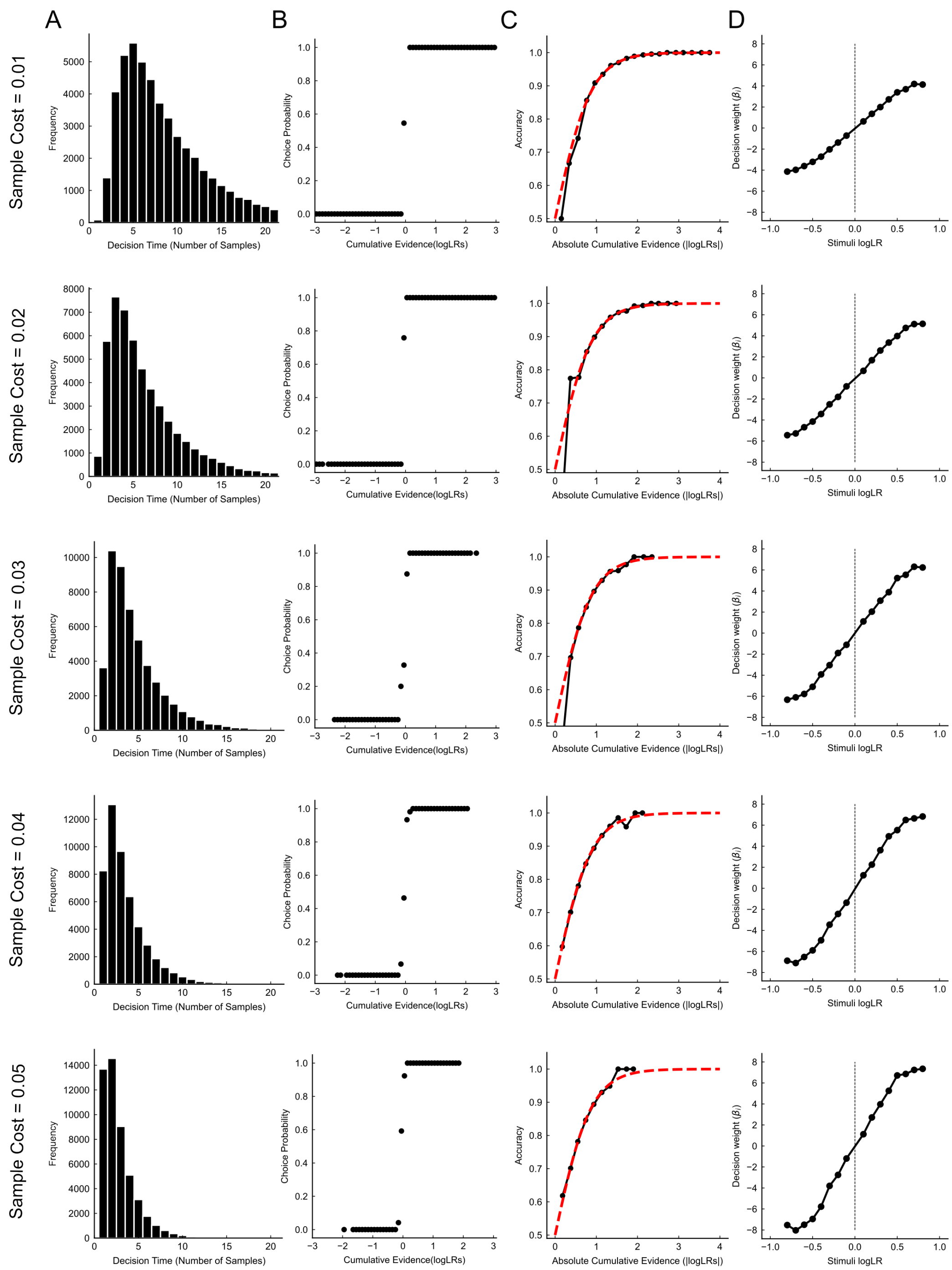

Figure S2: **Behaviors Metrics of the RNNs trained in Environment 2 with different sampling cost** (A) Models showed typical positive skewed distributions of decision times. (B)The psychometric curves of choice probability as a function of the logLRs of the cumulative evidence. (C) The psychometric curves of choice accuracy as a function of the logLRs of the cumulative evidence. Red lines indicate the theoretical optimum. (D) The models' decision weights on the sixteen stimuli varying in the amount of information showed quasi-linear curves between the ground truth logLRs and the decision weights contributed to binary choices.

A.4 The classic theory of sequential probability ratio test

When decision evidence is sampled sequentially, the classic theory employs the sequential probability ratio test (SPRT) as an optimal evidence accumulation process (Wald, 1947).

Consider that we observed a series of samples $\{x_1, x_2, ..., x_t\}$ from either $f_0(X)$ or $f_1(X)$ distribution. That leads to two alternative hypotheses H_0 and H_1, when inferring the underlying distribution given the observed samples. Let α and β be the error rates of rejecting H_0 and H_1, respectively. Wald (1947) minimized the samples taken when subject to a certain false rate:

$$\min_{S \in \mathcal{S}} \quad \pi_0 \mathbb{E}_0[n] + \pi_1 \mathbb{E}_1[n] \tag{6}$$

$$\text{s.t.} \quad \Pr_0^S(\text{refuse } H_0) \leq \alpha, \qquad \Pr_1^S(\text{accept } H_0) \leq \beta \tag{7}$$

To achieve a specific first-type error rate α and second-type error rate β, SPRT predicts a limited number of samples to reach the specified level of accuracy. derived that the decision threshold a and b can be simply written as the following forms:

$$a \approx \log \frac{\beta}{1-\alpha} \tag{8}$$

$$b \approx \log \frac{\alpha}{1-\beta} \tag{9}$$

At each time step, the likelihood ratio between these two hypotheses given the current evidence x_t is log-transformed and defined as ℓ_t:

$$\ell_t = \ln \frac{f_1(x_t)}{f_0(x_t)}. \tag{10}$$

Integrating the likelihood of evidence in a Bayesian manner over n time steps results in an accumulated log-likelihood as follows,

$$\ell_n = \sum_{t=1}^{n} \ln \frac{f_1(x_t)}{f_0(x_t)} \tag{11}$$

The decision terminates when ℓ_n reaches the decision threshold a or b. It is proven that it requires a limited number of samples to achieve the specified level of accuracy. However, this approach does not specify what threshold is optimal when under cost, which is elaborated below.

A.5 The normative model of sequential probability ratio test under costs

Distinguished from the classic theory of sequential probability ratio test above, we reconsidered this question under the resource-rational framework(Lieder and Griffiths, 2020; Bhui et al., 2021). Specifically, we developed a normative model to assess the optimal policies in our heterogeneous environment with different types of costs.

To formalize the question, this sequential sampling task can be viewed as a Meta-level Markov decision processes (Meta-MDPs), which is defined with a set of world states $\mathcal{W}$, a set of mental states $\mathcal{M}$, a set of computations $\mathcal{C}$, and a reward function $\mathcal{R}$.

The world states $\mathcal{W}$ include two hypotheses $\{H_0, H_1\}$, with each inferring the corresponding true world state that generates the observed samples $\boldsymbol{x}$. Let $f_0(x_t) = p(x_t \mid H_0)$ and $f_1(x_t) = p(x_t \mid H_1)$ be the sampling probabilities of $\boldsymbol{x}$ under the two world states. In the current setup (Fig. 1A), $f_0(x_t)$ and $f_1(x_t)$ are two Gaussian distributions with different means μ_0 and μ_1 and the same variance σ:

$$
\begin{aligned}
H_0 &: x \sim \mathcal{N}(\mu_0, \sigma^2), \\
H_1 &: x \sim \mathcal{N}(\mu_1, \sigma^2)
\end{aligned}
$$

The mental states $\mathcal{M}$ include a set of posterior beliefs b_t at each time step the agent holds that supports H_1 and against H_0 given the samples $\{x_1, x_2, ..., x_t\}$ (noted as $x_{1:t}$ in the following text for simplicity). The belief state at each time step can be defined as $b_t = p(H_1 \mid x_{1:t})$.

For the computation $\mathcal{C}$, if the optimal agent chooses to sample at time $t-1$, it can be viewed as executing an additional computation c via Bayesian posterior update:

$$
b_t = p(H_1 \mid x_{1:t}) = \frac{p(x_t \mid H_1) \cdot b_{t-1}}{p(x_t)} = \frac{f_1(x_t) \cdot b_{t-1}}{f_1(x_t) \cdot b_{t-1} + f_0(x_t) \cdot (1 - b_{t-1})} \tag{12}
$$

Finally, the expected reward function $\mathcal{R}$ describes the expected value of executing each possible action. In this task, it can be defined as follows,

$$
R(b, a) = \begin{cases} R_0(b_t, \perp) & \text{if } a = 0 \text{ (choose } H_0\text{)} \\ R_1(b_t, \perp) & \text{if } a = 1 \text{ (choose } H_1\text{)} \\ -c_{\text{tot}} + \mathbb{E}_{x|b}[V_{t+\Delta t}(b'_{t+\Delta t}(x))] & \text{if } a = 2 \text{ (sample)} \end{cases} \tag{13}
$$

Here, $\perp$ indicates the terminal action that is executed (i.e., commit to A or B) to end the Meta-MDP. For the expected value of these two terminal actions:

$$
R_0(b_t, \perp) = 1 - b_t \tag{14}
$$

$$
R_0(b_t, \perp) = b_t \tag{15}
$$

Here we assume that it depends on the current posterior belief π_t on H_0 and H_1. In particular, the more confident the agent is in one of the two hypotheses, the higher the expected value of committing to a target.

For the expectation of sampling, it depends on the sum of the total cost c_{tot} and the expected value of obtaining a new sample in the next time step:

$$
\mathbb{E}_{x|b}\left[V_{t+\Delta t}\left(b'_{t+\Delta t}(x)\right)\right] = \int_{-\infty}^{\infty} V_{t+\Delta t}\left(b'_{t+\Delta t}(x)\right) f(x \mid b)\, dx \tag{16}
$$

where $b'_{t+\Delta t}(x)$ denotes the possible updated belief in the next time step after obtaining a new sample, $\mathbb{E}_x[V(b'(x))]$ denotes the expected future value over all possible beliefs following a new sample x, $f(x \mid b_{t+\Delta t}) = b_t \cdot f_1(x) + (1 - b_t) \cdot f_0(x)$ indicates the marginal observed probability distribution. Here we obtained the posterior value $V_{t+\Delta t}\left(b'_{t+\Delta t}(x)\right)$ by looking up the value function $V_t\left(b'_{t+\Delta t}(x)\right)$ in the current time step.

To demonstrate the effect of sampling cost, we used different sampling cost c_{total} under two environment. When no time constraints, there is only sampling cost for taking one more step further. When there is a time constraint, we introduced an explicit urgency cost that increases over time to model time pressure. Thus, the total cost can be written as:

$$c_{\text{total}} = c_{\text{sample}} + \alpha t \tag{17}$$

where α controls the urgency cost. When $\alpha = 0$, there is no time constraint. When α increases, time pressure becomes stronger.

To find the optimal policy for this problem and delineate the optimal thresholds over time, we used dynamic programming approach, computing the value function by value iteration following Bellman's optimality:

$$V^*(b,t) = \max \left\{ \underbrace{R_0(b_t, \perp)}_{\text{Choose } H_0}, \underbrace{R_1(b_t, \perp)}_{\text{Choose } H_1}, \underbrace{-c + \mathbb{E}_{x|b}\left[V(b'(x))\right]}_{\text{Sample}} \right\} \tag{18}$$

After obtaining the whole value function by iterating over every possible belief over time steps, we then derived the optimal policy $\pi^*(b_t)$ by choosing the action that has the largest value. We defined the threshold as the transition belief states where the optimal policy changes from sampling to committing to target A or B.

We tested the optimal model under different combinations of sampling cost and urgency cost. When there is only sampling cost (Fig. S2A), the optimal policy exhibited early commitment behavior at the earlier time steps. The early commitment effect becomes less prominent when the sample cost increases, which is consistent with our findings in the neural network. Then, we changed the levels of urgency cost and fixed the sample cost at -.01 (Fig. S2B). The optimal policies show collapsing boundaries at the late stage. Increasing the urgency cost leads to steeper declines of the boundaries.

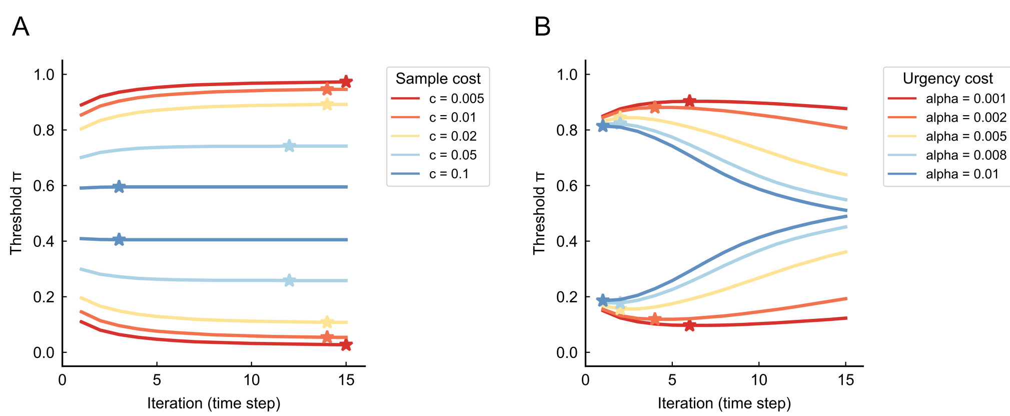

Figure S3: **Optimal policy under different cost structure** (A) We set urgency cost to 0 in this test, and changed the sampling cost as follows: 0.005, 0.01, 0.02, 0.05, 0.1. The asteroids indicated the maximum/minimum of each curve. (B) We fixed the sampling cost at 0.01 and changed the urgency cost α as follows: 0.001, 0.002, 0.005, 0.008, 0.01. Here we used the same Gaussian distributions as used in the training of the neural networks.

A.6 Revealing the model's decision weight on each stimulus

A.6.1 Estimating the decision weight of each stimulus on the model's choices

To estimate the subjective weight of each stimulus on the RNN's binary choices, we performed the following logistic regression:

$$\text{logit}(P(\text{chosen B})) = \beta_0 + \sum_{i=1}^{8} \beta_i n_i \tag{19}$$

Here n_i is the count of each stimulus that appeared in each trial. We then used the fitted logistic coefficients β_i to quantify the contribution of each stimulus to the model's choices and compared them to the ground-truth logLR of each stimulus.

A.6.2 Estimating the model's decision weights on evidence over time

The sequential order of a stimulus might impact the model's decision weight on the corresponding evidence. To test this hypothesis, we performed a logistic regression over time steps,

$$\text{logit}(P(\text{chosen B})) = \beta_0 + \sum_{t=1}^{T} \beta_t \cdot \text{logLR}_t \tag{20}$$

Here, T is the decision time, logLR_t is the ground truth evidence provided by the stimulus at every time step, calculated as the log-likelihood ratio of the sampling probability between the two target distributions. We used the fitted logistic coefficient β_t to indicate the model's decision weights on evidence at different time points of a sequence.

Here, we found that RNNs trained from both environments exhibited recency bias (Fig. S4); the pattern was stronger in the environment 1. This informed an intriguing strategy of unequal temporal weights on evidence when under the adaptive decision strategies.

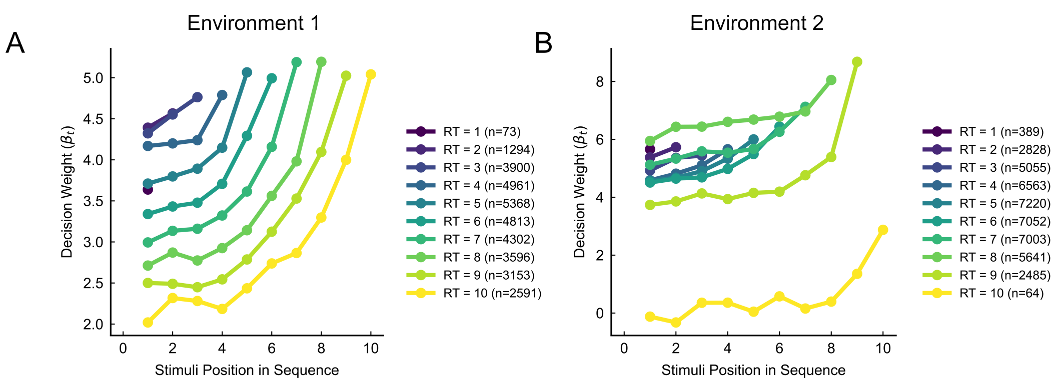

Figure S4: Temporal weights of the RNNs on evidence over time. The top and bottom panels showed the logistic regression weights and its matrix (β) of the RNNs' choice on the evidence over the time steps before decisions under two environments. Colored lines indicated the trials aggregated by reaction times. The figures were made with a same sampling cost ($c = 0.01$ in both environments). (A) Agent exhibited strong recency bias in environment 1. The temporal weight at earlier evidence declined when a decision was made later. (B) Agent in environment 2 also exhibited recency bias overall, but the temporal weights in the earlier time steps approached a similar levelthen showed a large reduction at the last time step.

A.7 Investigating the internal representation of the hidden states

To examine how environmental structure shapes the internal representations of the neural network, we performed principal component analysis (PCA) on the hidden states. We plotted the first ten principal components for environment 1 and environment 2 (Fig. S5A). The results revealed that in the absence of a time constraint, the first principal component alone explained more than 80% of the variance. In contrast, when a time constraint was imposed, the second principal component increased largely in capturing the variance: the first component accounted for approximately 45% of the variance, while the second component explained nearly 40%. This shift suggested that time pressure altered the dimensional structure of the network's internal representation.

We further found that the first principal component (PC1) in both environments was selectively associated with the cumulative logLRs, rather than with other task variables (e.g., stimulus identity, sampling probability, and etc.). The activation of PC1 exhibited an approximately linear relationship with cumulative logLRs at each time step, indicating a near-optimal encoding of evidence (Fig. S5B). To validate this representation, we computed the cross-temporal correlation between PC1 activity at time step i and cumulative logLR at time step j (Fig. S5C). The strongest correlations appeared along the anti-diagonal of the heatmap, indicating that the agent continuously tracked the evolving evidence over time. These findings suggested that even though the logLRs of evidence were not explicitly provided to the agent, it learned to internally represent this key decision variable and performed evidence accumulation in an information-theoretically optimal manner.

Interestingly, the agent in the environment 2 showed a feature that represented the time-coding information when we plot the hidden states of some example trial trajectories in the 2D space of PC1 and PC2 (Fig. S5D). Just like the time-varying decision boundaries shown in Fig. 2, the trajectories of hidden states also exhibited a boundary-hitting like feature, with early commitment in the earlier decision trials. However, when we checked the environment 1, we did not find a large impact from the PC2 direction.

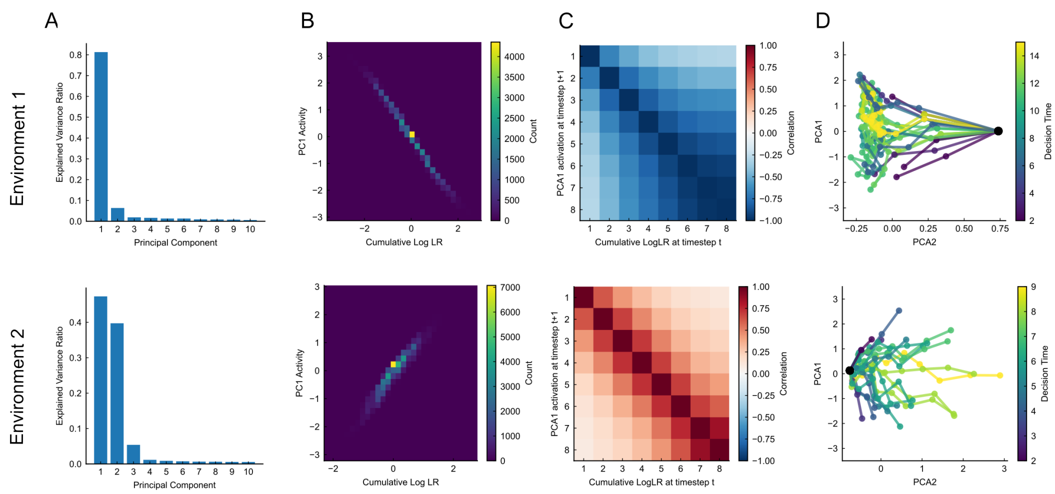

Figure S5: Internal representation of the hidden states under the two environments. The top and bottom rows showed the results in environment 1 and environment 2, respectively. (A) The scree plot for the first 10 components. PC1 accounted for 80% of the variance in the environment 1, and the second component increased largely in the environment 2. (B) 2D-histogram on PC1 activity and cumulative evidence (logLR) at time step 8. PC1 exhibited a strong correlation with the cumulative evidence in both environments. (C) The correlation between PC1 activity at time step i and cumulative evidence at time step j. The strongest correlation appeared at the anti-diagonal lines. (D) The example trial trajectories when projecting the hidden states into 2D space of PC1 and PC2. PC2 revealed the representation of time in the environment 2 but notin the environment 1.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: Please refer to the abstract, introduction and conclusion sections.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Please refer to the limitation section.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: Please refer to the appendix A.4 and A.5 for full proof.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Please refer to the method section.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: Please refer to Appendix A.1 and the github repository.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: We specify all these details in the method section.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: We provides the definition of error bars in the results section.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.

- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes] ,

   Justification: Please refer to the Code availability.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The paper conforms with the NeurlIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: Please refer to the discussion section of the paper

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to

generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: The paper poses no such risks

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [NA]

Justification: The paper does not use existing assets.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: The paper does not release new assets.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [NA]

Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.