Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

How Humans Restructure Predictive Models: Context-Tree Dynamics in Sequential Learning

Italo Ivo Lima Dias Pinto

italoivo@gmail.com

Universidade de São Paulo

Paulo Roberto Cabral-Passos

Universidade de São Paulo

Priscila S. Azevedo

Federal University of Rio de Janeiro

Fernando Araujo Najman

Universidade Federal do ABC

Paul Bogdan

University of Southern California

Claudia Domingues Vargas

Federal University of Rio de Janeiro

Florencia Leonardi

Universidade de São Paulo

Article

Keywords:

Posted Date: March 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8970014/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

How Humans Restructure Predictive Models: Context-Tree Dynamics in Sequential Learning

Italo Ivo Lima Dias Pinto[1,*], Paulo Roberto Cabral-Passos[1], Priscila S. Azevedo[2], Fernando Araujo Najman[3], Paul Bogdan[4], Claudia D. Vargas[2], and Florencia Leonardi[1]

[1]Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, Brazil
[2]Instituto de Biofísica Carlos Chagas Filho, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil
[3]Centro de Matemática, Computação e Cognição, Universidade Federal do ABC, São Bernardo, Brazil
[4]University of Southern California, Los Angeles, CA, USA
[*]italoivo@gmail.com

ABSTRACT

Human learning in sequential environments often involves uncovering latent statistical structures that govern event regularities. In this study, we investigate how individuals adapt their internal predictive models while playing the Goalkeeper Game, a stochastic sequence prediction task driven by a probabilistic context tree. We introduce a real-time context-tree inference framework that reconstructs the evolving internal models underlying participants' trial-by-trial choices. By tracking the entropy of the inferred context trees, we reveal that learning unfolds through two intertwined processes: frequent refinements, corresponding to gradual adjustments of transition probabilities within a stable structure, and rare transitions, corresponding to structural reorganizations of the predictive model. Entropy reductions parallel improvements in success rate, demonstrating that participants progressively internalize the underlying generative process. The waiting-time distribution of transitions follows a sub-exponential Weibull law, indicating history-dependent reorganization dynamics consistent with bursty, non-memoryless adaptation. These findings suggest that human statistical learning proceeds through a balance between exploitation and exploration. Our framework provides a quantitative and interpretable tool for modeling the continuous–discrete dynamics of adaptive learning in probabilistic environments.

1 Introduction

Humans have a remarkable capacity to internalize sequential patterns through statistical learning, a domain-general mechanism for predicting future events[1]. This ability underpins a wide range of behaviors, from language processing to decision-making in sports, allowing a goalkeeper, for instance, to unconsciously learn an opponent's shot distribution over time. Even when an outcome is initially deemed unlikely given long-term statistics, humans can rapidly adjust their expectations with repeated exposure[2]. Such adaptability suggests a Bayesian-like updating process, in which recent observations dynamically reshape internal predictive models.

Context-tree models, originally introduced in the field of data compression by Rissanen[3], provide a principled framework to represent sequential dependencies with variable memory depth. By encoding conditional probabilities over contexts of varying length, these models can flexibly capture the statistical structure of human expectations. A context-tree model with adaptive memory[4] can further emulate how humans down-weight older information, enhancing prediction accuracy in dynamically changing environments. Following the idea that the brain's predictive memory can be modeled as a variable length memory process, we use a context-tree model (also known as a variable-length Markov chain) to represent how goalkeepers learn transition probabilities of opponents' actions. This context-tree proxy aligns with cognitive neuroscience evidence that sequential regularities guide human expectations[5,6].

Previous studies have demonstrated that the human brain encodes statistical regularities in structured stimuli using context-tree models, providing electrophysiological evidence of such structures from EEG recordings during auditory prediction tasks[7]. Recent work has extended this line of evidence by showing that context trees can be directly retrieved from EEG signals, revealing the brain's internal encoding of sequence structure[8]. Complementarily, behavioral analyses have shown that response times reflect context-based learning and error modulation in sequential prediction tasks[9], and that the complexity of the underlying context tree influences the learnability of probabilistic sequences in the Goalkeeper game[10]. Specifically, more intricate tree structures or disruptions to temporal regularities reduce learning efficiency and delay strategy convergence.

However, humans often rely on simple sequential cues and show strong limitations with respect to more sophisticated patterns in adversarial tasks[11]. In the context of sports, goalkeepers tend to anticipate straightforward

tendencies—such as a shooter's preferred side—while overlooking higher-order trends. This behavioral asymmetry highlights the need for computational models that can reveal how humans transition from simple heuristic responses to structured, context-based learning. Here, we address this challenge by modeling the evolution of internal predictive structures through the dynamics of inferred context trees.

The remainder of this paper is organized as follows. Section 2 presents the experimental design of the Goalkeeper Game and details the real-time context-tree estimation algorithm used to infer participants' internal predictive structures. Section 3 reports the main behavioral and computational findings, including the convergence of participants' entropy toward the theoretical ground truth, the decomposition of learning into refinement and transition processes, and the statistical characterization of their temporal dynamics. Section 4 interprets these results in the context of adaptive learning, linking the observed refinement–transition duality to exploration–exploitation trade-offs and self-organized critical dynamics. Finally, Section 5 summarizes the main contributions and discusses broader implications for modeling human predictive behavior in dynamically changing environments.

2 Methods

The procedures described here were conducted in agreement and with the approval of the Ethics Committee of the University of São Paulo (CAAE 69431623.2.0000.5407) and of the Deolindo Couto Institute of Neurology at the Federal University of Rio de Janeiro (CAAE 72756723.8.-0000.5261). All methods were performed in accordance with the relevant guidelines and regulations. All participants provided informed consent before proceeding to the experimental task.

2.1 Experimental Design

Two identical experimental set-ups were prepared in the following facilities: the Research, Innovation and Dissemination Center for NeuroMathematics (NeuroMat/FAPESP) at the University of São Paulo and the Neurosciences and Rehabilitation Laboratory at the Federal University of Rio de Janeiro. The study included 26 right-handed participants (12 female, mean age 24 $\pm$ 6.61 years), whose handedness was confirmed using the Portuguese version of the Edinburgh Handedness Inventory.

The task to which the subjects were submitted consisted in playing the Goalkeeper Game (game.numec.prp.usp.br), a video-game developed by NeuroMat, in which the player controls a goalkeeper through a sequence of penalty trials. Each participant played the game while seated in comfortable chair with arm and foot rests. For standardizing the subjects' visual angle relative to the screen, the subjects nasion landmark was aligned with the vertical midpoint of the 32-inch Display++ monitor (Cambridge Research Systems Ltda.) positioned at a distance 114 cm.

The game configurations consisted in 1500 penalty trials with no pauses between trials, except for those after trial 500 and 1000. In each trial, the goalkeeper were asked to predict the penalty taker's next choice and inform his/her prediction by pressing a button of the keyboard (CEDRUS RB-840). The penalty taker was configured to shoot to one of three options: to the left, to the center, or to the right of the goal, henceforth referred to as 0, 1, and 2, respectively. If the penalty taker choice coincided with the goalkeeper informed prediction, an animation at the end of the trial showed a successful defense; otherwise, a goal animation was presented.

Figure 1 shows a schematic representation of the trial. On the left, we show the keyboard design and the fingers' correspondence with the game option: right index finger (0), right middle finger (1), and right ring finger (2). On the right, the duration of each trial section is presented along with the game screen samples. The first section is the readiness period, during which the goalkeeper and penalty taker on the screen remain at rest ($200ms$). The second section is the response time section, marked by the appearance of three arrows at the bottom of the screen, informing that the goalkeeper is allowed to inform his/her prediction. No maximum duration for this period was set. The response time section ended with a button press, immediately initiating the feedback animation ($2300ms$). After each 500 trials, subjects were allowed to take a rest, no time limit was set to this interval. During this period, the participant was allowed to move, drink water, and communicate any concern to the researcher, notifying the researcher when he/she felt ready to return to the game.

At the beginning of the task, the goalkeepers were presented to a special screen with information about which fingers he/she should use, the finger/button/choice mapping, and the game mechanics explanation. The instruction given to the subjects was: "Your task is to avoid as many goals as you can. Attention: the penalty taker is not influenced by your choices". The penalty taker's choices followed the Variable Length Context Tree model, also termed Probabilistic Context Tree model, presented bellow.

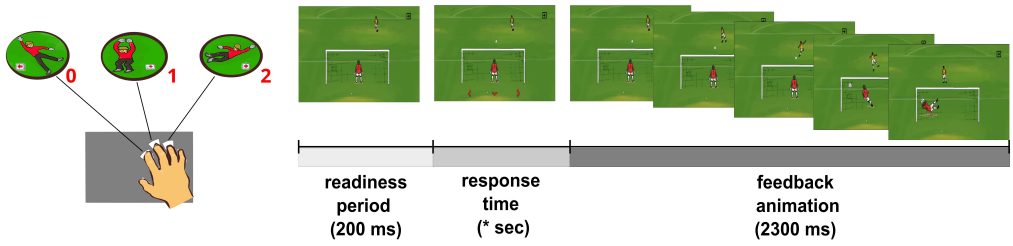


Figure 1. The Goalkeeper Game task. (Left) Key/finger/choice mapping: right index finger (0), right middle finger (1), and right ring finger (2). (Right) Duration of each trial section and game screen samples. The readiness period section is characterized by both, the goalkeeper and penalty taker, at rest. The response time section is characterized by the appearance of three arrows at the bottom of the screen informing the goalkeeper to inform his/her prediction. No maximum duration was set for this period. The button press triggered the end of the section. The last section corresponded to the feedback animation ($2300ms$).

$$P(a \mid w) = \begin{array}{c} \\ \text{w} = 0 \\ \text{w} = 01 \\ \text{w} = 11 \\ \text{w} = 21 \\ \text{w} = 2 \end{array} \begin{array}{c} \begin{array}{ccc} \text{a} = 0 & \text{a} = 1 & \text{a} = 2 \end{array} \\ \begin{pmatrix} 0 & 0.3 & 0.7 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \end{array} \tag{1}$$

Sequence example: ...0 2 1 0 1 1 0 2 1 0 ...

On top of the matrix 1, the symbols 0, 1, and 2 represent the penalty taker choices. Each table row presents a context on the far left. The model presents a total of five contexts. At any step of the sequence, the next symbol of the sequence can be predicted by identifying in which context the step ends. For a given context, the values under each penalty taker choice (0,1,2) indicate the probability associated with this symbol being the one after the context. For example, if the current step ends with 01 (that is, left, center), the next choice of the penalty taker is always 1 (center). On the other hand, if the current step ends in the context 0, the next choice can be either 1 (center) with probability of 0.3 or 2 (right) with probability of 0.7. Therefore, we say that the context is stochastic, since it is impossible to predict with certainty the next choice. A context is described as the minimum amount of past symbols necessary to predict the next symbol. A penalty taker sequence generated by the model is presented as an example bellow the probability matrix.

2.2 Context-tree inference of internal predictive structure.

In the goalkeeper game, each player's sequence of choices embodies an adaptive process of learning and prediction, an implicit attempt to model the opponent's regularities and adjust future actions accordingly. To infer these latent internal models from behavior, we employ a real-time context-tree estimation algorithm, which estimates the context-tree structure underlying the observed sequences.

From the perspective of human statistical learning, each choice reflects a hypothesis about the opponent's hidden generative process. The real-time context-tree estimation algorithm thus serves as a reverse-engineering tool to estimate this latent predictive structure from observed behavior. Unlike conventional batch-based context-tree estimation that requires a fixed dataset, the real-time approach updates its structure and probabilities incrementally, allowing us to trace how the player's inferred model evolves trial-by-trial.

Context-tree estimation

Let $\mathcal{A} = \{a_1, \ldots, a_K\}$ be a finite alphabet of size K. We observe a discrete symbolic time series $x_1, x_2, \ldots$ taking values in A. Denote the history before time t as

$$h_t = (x_1, x_2, \ldots, x_{t-1}).$$

We posit that x_t is generated by a variable-length Markov chain, meaning that its conditional distribution depends on a suffix (context) of the recent history, up to a maximum context length D.

In a context tree each leaf corresponds to a suffix $c \in \mathcal{A}^{\leq D}$. At node c, we maintain counts $n_c(a)$ of how many times symbol a followed that context, and $n_c = \sum_{a \in \mathcal{A}} n_c(a)$. The conditional probability estimate is

$$\hat{p}(a \mid c) = \frac{n_c(a) + \alpha}{n_c + \alpha K},$$

where $\alpha > 0$ is a pseudocount parameter. This additive (Laplace/Dirichlet) smoothing ensures that even unobserved symbols retain a small nonzero probability, preventing zero-likelihood assignments to future sequences. The estimator can be interpreted as the posterior mean under a symmetric Dirichlet–multinomial model with prior parameter α, providing a Bayesian regularization that stabilizes probability estimates when data are sparse and allows the model to adapt as new contexts emerge [12, Ch. 12].

The goal of the real-time context-tree estimation algorithm is to incrementally (i) update counts, (ii) dynamically build and prune the tree of active contexts, and (iii) perform prediction in a streaming fashion.

Count Updates and Forgetting: At each time step t, after observing the new symbol x_t, the algorithm updates its internal statistics to reflect the most recent observation. It does so by considering all suffixes of the preceding sequence up to the maximum depth D. Each suffix

$$c = x_{t-\ell} x_{t-\ell+1} \cdots x_{t-1}, \qquad \ell = 0, 1, \ldots, D,$$

represents a potential predictive context. For every such context, the algorithm keeps track of the total number of times it has been observed, n_c, and how often each possible next symbol $a \in \mathcal{A}$ followed it, denoted $n_c(a)$. These quantities are incremented according to

$$n_c \leftarrow n_c + 1, \qquad n_c(x_t) \leftarrow n_c(x_t) + 1.$$

To ensure adaptability to nonstationary data, the algorithm applies an exponential forgetting mechanism with a decay factor $\gamma \in (0, 1]$ before each update:

$$n_c \leftarrow \gamma\, n_c, \qquad n_c(a) \leftarrow \gamma\, n_c(a) \quad \forall a \in \mathcal{A}.$$

This step progressively downweights older observations, effectively implementing an exponential moving average over time. By emphasizing recent events, the model remains responsive to changes in the underlying data distribution, allowing it to track shifts in participants' behavior or evolving environmental regularities.

Dynamic Context Selection: Instead of predefining all suffixes up to length D as predictive contexts, the real-time context-tree estimation algorithm grows and prunes the tree adaptively based on how informative each context is relative to its parent based on a divergence criterion.

Let c be a candidate leaf whose parent is c_{par}, the suffix obtained by dropping the leftmost (most distant-past) symbol of c. We compare their predictive distributions using the Kullback–Leibler divergence

$$D_{\text{KL}}\big(\hat{p}(\cdot \mid c) \,\|\, \hat{p}(\cdot \mid c_{\text{par}})\big) = \sum_{a \in \mathcal{A}} \hat{p}(a \mid c) \log \frac{\hat{p}(a \mid c)}{\hat{p}(a \mid c_{\text{par}})},$$

We promote node c to an *active predictive context* if its divergence exceeds a threshold

$$\epsilon(n_c, c) = c_0 \sqrt{\frac{(K-1) \ln n_c}{n_c}},$$

The threshold $\epsilon(n_c, c)$ follows a standard finite-sample rationale: the empirical KL divergence of a multinomial estimator concentrates at order $(\ln n_c)/n_c$ up to constants[13], while context-tree methods typically rely on confidence radii scaling as $\sqrt{\ln n_c / n_c}$ for distances between empirical probability vectors[14]. Exponential deviation bounds for context-tree estimation further support the use of data-dependent pruning rules of this type[15]. Thus, $\epsilon(n_c, c)$ serves as a conservative, theoretically motivated heuristic rather than an optimal choice. The constant c_0 balances sensitivity against over-splitting, ensuring that a context is promoted only when D_{KL} exceeds this small threshold and the tree updates only in response to meaningful differences in predictive distributions.

The context tree estimator modifies its structure through two complementary mechanisms: growing and pruning. These operations allow the model to adapt its effective memory length based solely on statistically supported changes in empirical conditional distributions.

Growing occurs when a longer candidate context provides predictive information not captured by its parent. For a candidate $c = ax$, the estimator compares $\hat{P}(\cdot \mid c)$ with $\hat{P}(\cdot \mid x)$ using an information-theoretic discrepancy measure, typically the empirical Kullback–Leibler divergence. If this divergence exceeds the threshold $\epsilon(n_c, c)$, the longer context encodes a significant refinement of the transition structure and is promoted into the tree.

Pruning removes contexts whose additional length no longer improves prediction. After each symbol update, the estimator reassesses all active contexts: if the divergence between $\hat{P}(\cdot \mid c)$ and $\hat{P}(\cdot \mid x)$ falls below $\epsilon(n_c, c)$, or if the context lacks sufficient data for a reliable estimate, then c is deemed indistinguishable from its parent and is removed. Pruning may propagate upward until reaching an ancestor that remains statistically justified.

Together, these mechanisms produce an adaptive and parsimonious model: the tree expands when the data reveal longer-range dependencies and contracts when such dependencies diminish or lose statistical support.

Stability and commitment: To avoid oscillation, a node must exceed the threshold for a minimum number of consecutive updates before it is committed as an active context. One may also require a minimum support condition, e.g. $n_c \geq \beta n_{\min}$. Once committed, all ancestors of c are unmarked so as to enforce the suffix (no-redundancy) property: no two active predictive contexts should lie on the same root-to-leaf path.

If later the divergence of a committed node drops and remains below the threshold for sufficiently many steps, the node may be demoted (pruned) to prevent overfitting.

This dynamic adaptation ensures the model maintains a compact set of contexts that flexibly adjusts to the complexity present in the data stream.

Prediction: At time t, to predict x_t, we examine the current history h_t. We find the active context c^* that matches a suffix of h_t. The predictive distribution is then

$$\hat{p}(x_t \mid h_t) = \hat{p}(x_t \mid c^*) = \frac{n_{c^*}(x_t) + \alpha}{n_{c^*} + \alpha K}.$$

We always use the symbol with highest probability as the prediction.

Model validation: After training the real-time context-tree estimation algorithm on each participant's sequence, we evaluated its predictive validity by comparing the model's one-step-ahead predictions with the actual subsequent choices. For each trial t, the model generated a predictive distribution $\hat{p}(x_t \mid h_t)$, and accuracy was defined as the proportion of trials in which the most probable symbol matched the observed outcome. Across participants, the real-time context-tree estimation algorithm achieved a mean prediction accuracy of 0.75 ± 0.08. Given the stochastic and nonstationary participant behavior during the experiment, such performance indicates that the inferred context trees capture consistent and adaptive dependencies in players' decision processes. To further assess behavioral relevance, we used the estimated transition probabilities to compute the entropy rate of the corresponding variable length memory process. This quantity reflects the uncertainty inherent in each participant's predicted choices. Participants with higher performance scores tended to exhibit lower model-derived entropy, establishing a direct link between the structure inferred by the real-time context-tree estimation algorithm and observable behavioral proficiency; these relationships are quantified in the Results section.

The divergence-based context selection criterion follows classical variable-order Markov approaches employing minimum description length principles[16], while related real-time context-tree estimation formulations have been studied under mistake-bound guarantees[17]. Performance depends on α, c_0, γ (for non-stationary behavior), and minimum-support thresholds. Here, parameters were hand-tuned to reproduce the ground-truth generator and fixed across analyses; no formal hyperparameter search was performed.

3 Results

Having described the real-time context-tree estimation algorithm in the previous section, we now apply it to reconstruct the internal predictive models of the participants in the Goalkeeper Game. Each subject's behavioral sequence, encoded as a stream of discrete symbols corresponding to the chosen target, was processed through the real-time context-tree estimation algorithm described. The inferred tree represents the participant's internal model of the task's temporal structure, summarizing the predictive contingencies between past and future actions.

The estimation was performed using the same hyperparameters across subjects to ensure comparability. The smoothing coefficient was set to $\alpha = 0.05$, the minimum-support threshold to $c_0 = 5$, and the non-stationarity rate to

$\gamma = 0.95$, defining how quickly past statistics decay in influence as new evidence arrives. These parameters were chosen based on sensitivity analyses showing that they robustly recover the ground-truth tree under similar stochastic conditions (Blue trace in Fig. 2 A). The algorithm maintained a fixed maximum context depth of $L_{\max} = 5$, allowing the emergence of long predictive dependencies when sufficiently supported by the data.

At each trial, the model updated transition probabilities and reorganized the tree structure to accommodate statistical changes in the sequence. This produced, for each subject, a temporal trajectory of context trees that reflects the evolution of their internal predictive model throughout the task. In this interpretation, each inferred tree corresponds to a snapshot of the subject's latent cognitive state—how they represent and anticipate the opponent's behavior at that point in learning.

By tracking the entropy of these inferred trees and the appearance or pruning of specific contexts, we can quantify how participants refine, stabilize, or reorganize their internal models over time. In the following subsections we analyze these dynamics, comparing the entropy trajectories of subjects with the stationary entropy of the ground-truth process and exploring how the structural complexity of the inferred models relates to behavioral performance.

3.1 Theoretical predictions

To establish a theoretical benchmark for the participants' behavior, we first computed the entropy rate and the expected accuracy of the process that generates the kicker's choices. The kicker's sequence was produced by a context tree defined over the alphabet $A = \{0,1,2\}$, with contexts (0), (0,1), (1,1), (2,1), and (2), and transition probabilities given by Matrix 1 defined in section 2.1.

Among these contexts, only the context (0) is stochastic; all others are deterministic. The stationary distribution over contexts can be analytically calculated as

$$\pi_{(0)} = \tfrac{1}{3}, \quad \pi_{(0,1)} = \tfrac{1}{10}, \quad \pi_{(1,1)} = \tfrac{1}{10}, \quad \pi_{(2,1)} = \tfrac{7}{30}, \quad \pi_{(2)} = \tfrac{7}{30}.$$

Since the only context with nonzero entropy is (0), whose conditional distribution is $P(\cdot \mid (0)) = (0.3, 0.7)$, and all other contexts are deterministic, the entropy rate is

$$H = \sum_c \pi_c H(P(\cdot \mid c)) = \pi_{(0)}(-0.3\log_2 0.3 \; - \; 0.7\log_2 0.7)\,.$$

Using $\pi_{(0)} = \frac{1}{3}$ gives

$$H = \frac{1}{3}(-0.3\log_2 0.3 \; - \; 0.7\log_2 0.7) \approx 0.29 \text{ bits per symbol.}$$

This value represents the *ground truth entropy* shown in Fig. 2A (gray dashed line in the upper plot) and corresponds to the minimum achievable uncertainty per trial in the kicker's sequence. Note that the deterministic transitions do not contribute to this entropy.

The *optimal strategy*, assuming full knowledge of the kicker's transition model, consists in always predicting the most probable outcome for each context:

$$(0)\to 2, \quad (0,1)\to 1, \quad (1,1)\to 0, \quad (2,1)\to 0, \quad (2)\to 1.$$

Under the stationary distribution of contexts, the expected accuracy of this strategy is

$$S^* = \sum_c \pi_c \max_a P(a|c) = \tfrac{1}{3}\times 0.7 + \tfrac{2}{3}\times 1 = 0.90.$$

This theoretical upper bound is depicted in Fig. 2A (green dashed line in the bottom plot) as the *optimal strategy score*.

The structure of the generator also admits a coarse-grained description in terms of a new alphabet

$$\mathcal{A}' = \{A, B\}, \quad A \equiv (0,1,1), \quad B \equiv (0,2,1).$$

In this representation, each three-symbol block of the original process corresponds to one symbol of $\mathcal{A}'$: block A occurs with probability $p_A = 0.3$ and block B with probability $p_B = 0.7$. Since consecutive blocks are independent, the entropy per block in this reduced alphabet is

$$H(\mathcal{A}') = -p_A \log_2 p_A - p_B \log_2 p_B \approx 0.881 \text{ bits per block.}$$

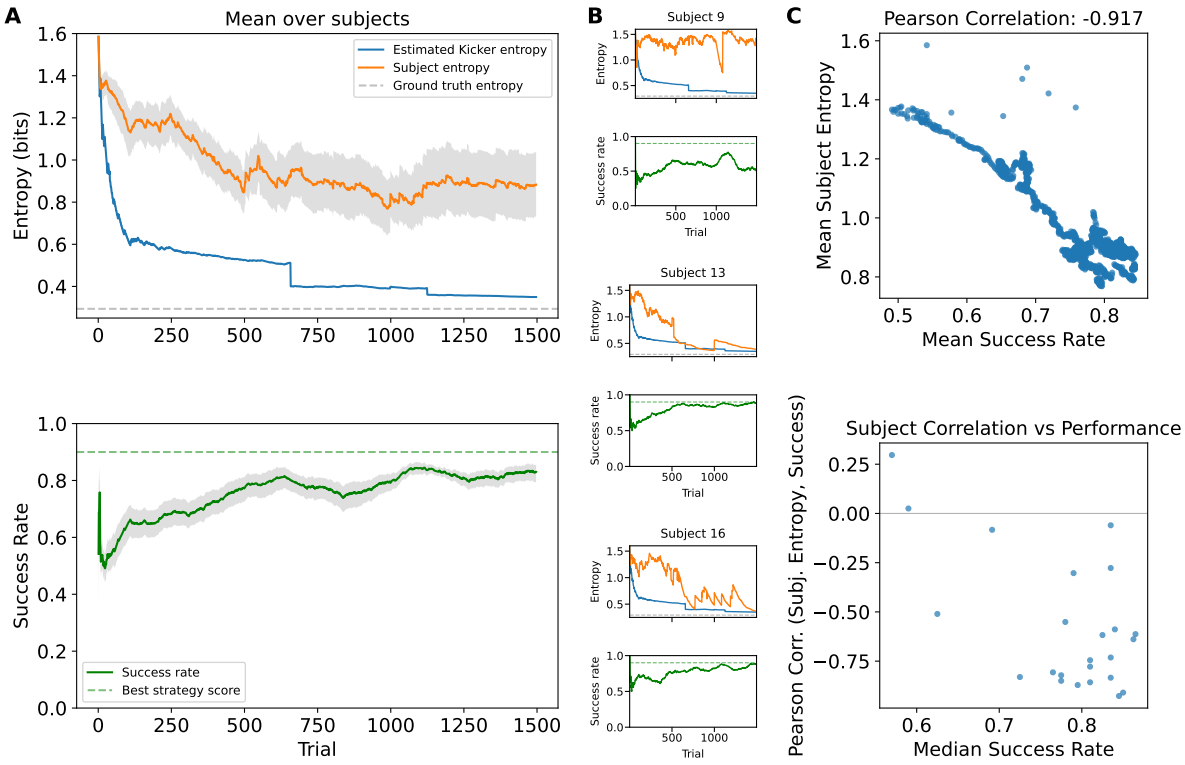


Figure 2. Relationship between entropy dynamics and performance during learning. **(A)** Mean subjects' entropy and entropy estimated by a perfect-memory kicker model across trials. Shaded areas indicate 95% bootstrap confidence intervals across participants, and the dashed line marks the theoretical steady-state entropy of the ground-truth context tree. **(B)** Example trajectories for three participants (Subjects 9, 13, and 16): top panels show the evolution of subject entropy and model-estimated entropy with the ground-truth reference; bottom panels show each participant's success rate and the best-strategy benchmark (dashed line). **(C)** Relation between uncertainty and performance. Top: correlation between the cohort-averaged subject entropy and mean success rate. Bottom: per-subject Pearson correlation between entropy and success rate as a function of median success rate, showing that higher performance is associated with lower entropy. The success rate at each trial reflects the proportion of correct responses over the preceding 200 trials (or cumulatively for the first 200 trials).

Because each block of $\mathcal{A}'$ corresponds to three symbols in the original alphabet $\mathcal{A} = \{0,1,2\}$, the corresponding entropy rate per original symbol is

$$H = \frac{H(\mathcal{A}')}{3} \approx 0.29 \text{ bits per symbol.}$$

Thus, the entropy rate of the original process equals 1/3 of the binary entropy associated with the stochastic choice between the two motifs A and B. This equivalence confirms that all residual uncertainty in the kicker's sequence arises from the probabilistic alternation between these two stereotyped patterns.

Together, these theoretical predictions provide normative references for the empirical results: participants' estimated entropies in Fig. 2A converge toward the *ground truth entropy* of the kicker, while their *average success rates* approach the *optimal strategy score* of 0.9. This correspondence strengthens the validation of the behavioral adaptation captured by the participants' inferred context-trees, indicating that participants progressively internalize the underlying structure of the kicker's sequence.

3.2 Learning Dynamics and Entropy Reduction

The empirical trajectories of entropy and performance, shown in Fig. 2, reveal how participants progressively refined their internal statistical models of the kicker's sequence. Each participant's behavior was modeled through the context trees estimated from their observed choices. The entropy of these subject-estimated context trees quantifies

the uncertainty in their predictive models and thus provides a direct measure of the information acquired during the learning process.

Fig. 2A presents the mean subject entropy (orange with a 95% CI) across trials together with the estimated entropy obtained by applying the same inference procedure to the kicker sequence (blue), the theoretical ground truth entropy of the generator (gray dashed line). The blue curve serves as a reference for the expected entropy trajectory when the learning algorithm is applied to the true generating process itself, illustrating the lower bound that subjects can asymptotically approach as information is accumulated through repeated exposure. The mean subject entropy gradually approaches the theoretical limit of approximately 0.29 bits per symbol, while the average success rate steadily approaches the optimal strategy score of $S^* = 0.9$. This joint evolution demonstrates that behavioral adaptation can be described as a process of entropy matching, whereby uncertainty about the kicker's statistics is progressively minimized through information flow.

Fig. 2B illustrates the diversity of individual learning trajectories by showing three representative participants. Subject 9 exhibits persistently high entropy and a persistent low success rate, indicative of a low performer who fails to learn the statistical pattern of the kicker. Subject 13, by contrast, displays a steady decrease in entropy accompanied by a monotonic improvement in performance, representing a consistent learner and a stabilization of performance close to the optimal strategy. Subject 16 shows an intermediate, saw-tooth pattern, with overall entropy reduction with sporadic discontinuities. This behavior is correlated with a gradual improvement in success rate. Together, these examples highlight that although individual trajectories may differ in their dynamics, successful learning consistently manifests as decreasing entropy and increasing predictive accuracy.

Fig. 2C summarizes the relationship between entropy and performance across all subjects. The upper panel shows that the correlation between mean entropy and mean success rate is strongly negative ($r = -0.917$), confirming that participants who maintained lower entropy in their estimated context trees achieved higher overall performance. Moreover, this inverse relationship becomes even stronger for subjects with higher median success rates, suggesting that efficient learners not only achieve better scores but also maintain more predictable and internally consistent context-tree structures, indicative of a stable set of transition patterns in their learned models.

Overall, these results support the view that learning in this task unfolds through two coupled processes operating at different time scales: a faster refinement process, in which the transition probabilities within the existing internal context tree are adjusted to match the observed sequence, and a slower structural process involving the growth and pruning of contexts in the tree itself. The latter is expressed as intermittent gaps and discontinuities in the entropy trajectory, a phenomenon that will be examined in detail in the following discussions.

3.3 Structure of Entropy Updates

Consider H_t as the entropy of the context tree estimated at time t. Fig. 3A characterizes the statistical structure of entropy increments $\Delta H_t = H_{t+1} - H_t$ across all participants. Each increment quantifies how much the uncertainty of the inferred context tree changes between consecutive trials. We classified ΔH_t values as refinements when entropy changes resulted from probability updates within an unchanged tree structure, and as transitions when the change reflected a structural reorganization involving additions, removals, or splits of predictive contexts. Transitions were comparatively rare, occurring in only a small fraction of the trials ($f_{\text{trans}} = 321/35904 \approx 0.9\%$), indicating that large-scale reorganizations of the internal model are exceptional events embedded within a background of continuous refinements. The histograms show the empirical probability densities of both classes on a logarithmic scale, highlighting their distinct scales and tail behaviors. Both distributions are heavy-tailed, revealing that entropy evolves through a mixture of frequent small fluctuations and occasional large adjustments. Refinements form a sharply peaked, positively skewed distribution, showing that most local updates slightly decrease entropy and, therefore, correspond to incremental exploitation of the current context tree, where participants fine-tune predictions based on accumulated evidence. In contrast, transitions display a broader and more symmetric distribution, consistent with abrupt exploration events in which the internal structure is reorganized to accommodate new hypotheses about the opponent's behavior. The coexistence of these two regimes, continuous exploitation through probabilistic refinements and rare exploration through structural transitions, demonstrates that learning proceeds through punctuated entropy dynamics rather than gradual convergence.

Fig. 3B focuses on the temporal evolution of entropy during refinement periods—updates in which the tree structure remains unchanged—after explicitly excluding all transition events. Because participants' behavior is intrinsically non-stationary, the session was divided into ten non-overlapping equal segments (W1–W10) to capture slow trends in refinement dynamics across the learning process. Within each segment, entropy values corresponding solely to refinement increments were pooled across all participants ($N = 3558 \pm 17$), yielding aggregate distributions that represent the typical state of the internal model at each learning stage. This segmentation procedure isolates the progressive adaptation of refinement statistics from the abrupt structural reorganizations analyzed in Panel

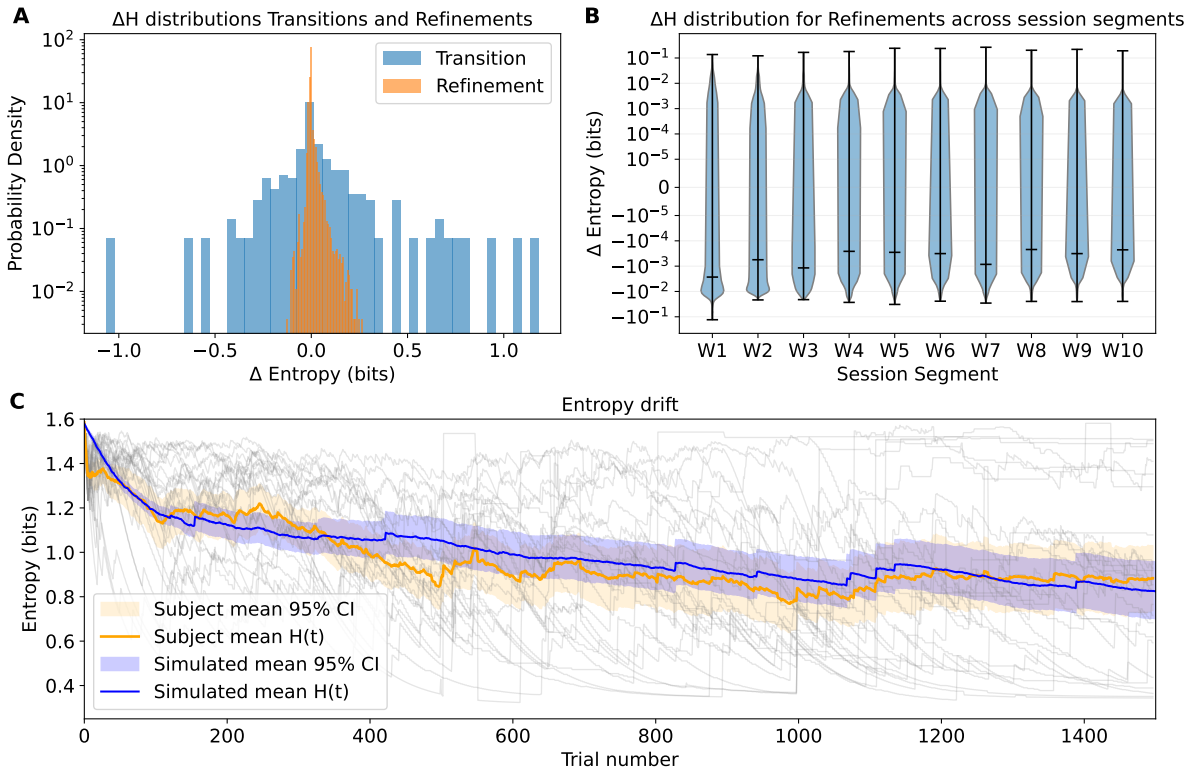


Figure 3. Entropy dynamics across learning and simulation. **(A)** Distributions of entropy change (ΔH) during context-tree transitions and refinements, plotted on a logarithmic probability scale. The distributions differ in dispersion: transition events exhibit a broader spread of ΔH values, while refinements are more concentrated around small magnitudes. The refinement distribution is also asymmetric and skewed toward negative ΔH values, indicating that entropy reductions are more common during refinement, whereas the transition distribution shows no significant asymmetry. **(B)** ΔH distributions for refinements across session segments, plotted on a symmetric logarithmic scale (symlog) in ΔH to capture both large positive and negative deviations. The amplitude and spread of entropy updates diminish over successive segments, suggesting progressive stabilization of the learned context tree. **(C)** Entropy drift over trials: gray traces represent individual subjects, the orange line and shaded region denote the subject mean and 95% bootstrap confidence interval, the blue line and shaded region indicate the simulated median and its 95% confidence interval.

A. The resulting distributions show a clear temporal evolution: early in the session (W1–W3), refinement entropy is broadly distributed and positively skewed, reflecting a highly variable exploratory regime in which participants are still calibrating their probabilistic expectations. As learning proceeds, the distributions become narrower and shift toward near-zero entropy increments, marking the emergence of a stable exploitation regime characterized by fine-grained adjustments around a consolidated internal model. Despite this overall stabilization, a residual dispersion persists even in late segments, indicating that small-scale exploratory corrections remain embedded within otherwise stable refinement dynamics. By focusing exclusively on non-transition updates, this analysis reveals how gradual entropy stabilization emerges from the internal exploitation of a fixed predictive structure, while the larger exploratory reorganizations remain segregated into the transition events quantified elsewhere.

To characterize how participants gradually adjusted their predictive models without undergoing structural reorganizations, we analyzed the refinement increments of entropy, corresponding to epochs in which the inferred context-tree remained unchanged between consecutive trials. The distribution of ΔH_t during refinements was sharply peaked and negatively skewed, showing that most updates slightly reduced entropy. To quantify how these fluctuations changed over time, the session was divided into ten equal segments, and the first three central moments (mean, standard deviation, skewness) were estimated in each segment. Weighted least-squares fits with exponential relaxation functions were then applied to obtain smooth, time-dependent curves $\mu(t)$, $\sigma(t)$, and skew(t), capturing the non-stationary evolution of refinement statistics throughout learning. These moment trajectories were subsequently used to parameterize a time-varying shifted-Gamma model of the entropy increments, in which the mean, variance, and skewness evolve continuously with session time (for more details on the parameter time dependence model see appendix A).

In contrast to refinements, transitions correspond to abrupt reorganizations of the inferred context tree, in which one or more predictive contexts were added, removed, or split. Although transitions occurred much less frequently ($N_{\text{trans}}/N_{\text{ref}} \approx 3.2 \times 10^{-2}$), their entropy increments ΔH_t exhibited a broader and more symmetric distribution, often producing large positive or negative deviations relative to the refinement baseline. These events therefore represent moments when participants restructured their internal predictive models rather than fine-tuning existing ones. To model the temporal spacing of these reorganizations, we analyzed the empirical distribution of the inter-transition intervals Δt, defined as the number of trials separating consecutive transition events.

Model comparison using maximum-likelihood estimation and Akaike Information Criteria (AIC) demonstrated a clear ranking among the candidate renewal models. The Weibull renewal process[18,19] provided the best fit to the data ($AIC_{Weibull} = 3185.73$, shape $\beta = 0.75$, scale $\lambda = 70.49$). The next-best alternative was the gamma distribution ($\text{AIC}_{\text{Gamma}}$=3192.18, $k = 0.66$, $\theta = 128.76$), followed by the log-normal (AIC_{LN}=3197.11, log-likelihood = -1596.56, $\sigma = 1.57$, scale = 33.38).

In contrast, simpler memoryless dynamics described by the exponential distribution performed substantially worse (AIC_{Exp}=3231.56, log-likelihood = -1614.78, $\lambda = 84.52$), indicating clear deviations from a Poisson process. Heavy-tailed alternatives also showed inferior support: the Lévy-stable model yielded AIC_{Levy}=3300.25 (log-likelihood = -1647.12), and a Pareto (power-law tail) fit performed worst overall ($\text{AIC}_{\text{Pareto}}$=3425.15, log-likelihood = -1711.58, $\alpha = 1.29$).

These results provide substantial evidence in favor of the Weibull model. Collectively, these results indicate that waiting times are best characterized by a stretched-exponential renewal process with decreasing hazard rate ($\beta < 1$), rather than by scale-free or memoryless dynamics.

This sub-exponential shape indicates a heavy-tailed, history-dependent process in which short intervals between transitions are common but long pauses occasionally occur. In the simulations, transition increments were drawn from a stationary shifted-Gamma distribution parameterized by the empirical mean, variance, and skewness of ΔH_t during transitions, while their timing was governed by the fitted Weibull renewal process. Together, these components allowed the model to reproduce both the smooth entropy drift driven by frequent refinements and the intermittent discontinuities caused by rare structural transitions in the participants' internal models.

Fig. 3C summarizes the comparison between the empirical and simulated entropy trajectories. The grey transparent traces correspond to the entropy $H(t)$ trajectories of individual participants, illustrating the variability in how entropy evolved across subjects. The orange curve represents the mean empirical trajectory, with shaded areas indicating 95% bootstrap confidence intervals. The blue curve shows the mean trajectory obtained from simulations of the non-stationary stochastic model that integrates both refinement and transition processes. The close correspondence between simulated and empirical curves demonstrates that the model captures the main features of entropy evolution across the experiment, including the gradual decline driven by refinements and the sporadic jumps associated with structural transitions.

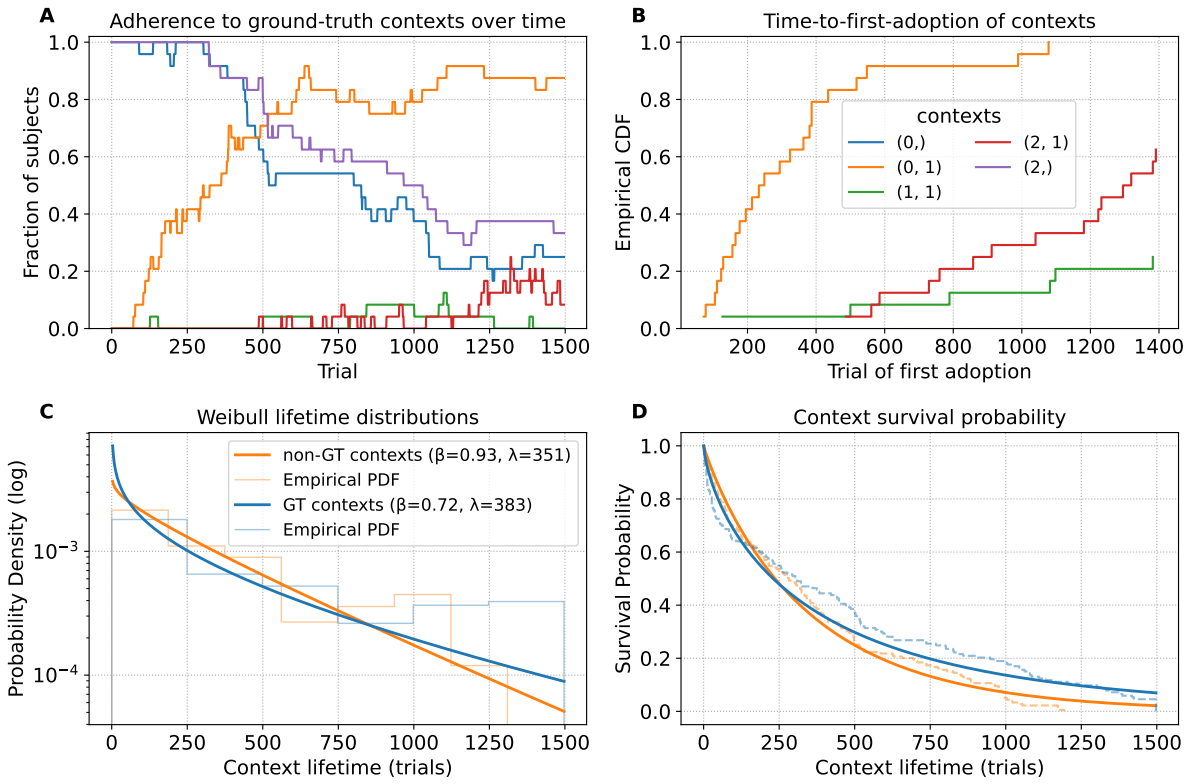


Figure 4. Adoption and persistence of contexts across the cohort. (**A**) Empirical cumulative distribution function (CDF) of time-to-first-adoption for contexts with length greater than one, normalized by cohort size (colors indicate contexts; legend shared across panels A and B). (**B**) Fraction of participants whose learned tree contains each ground-truth (GT) context over trials (**C**) Context lifetime probability density on a logarithmic scale: empirical stair histograms with corresponding Weibull fits. GT contexts exhibit heavier tails (fit example: $\beta \approx 0.72$, $\lambda \approx 383$) compared to non-GT contexts ($\beta \approx 0.93$, $\lambda \approx 351$), indicating sub-exponential lifetimes ($\beta < 1$). (**D**) Survival functions $S(t) = \Pr[T \geq t]$: empirical complements of the ECDF with Weibull survival fits, highlighting persistent contexts with slow decay.

3.4 Context Adoption and Persistence

To characterize the temporal dynamics of model adaptation, we examined how participants discovered and maintained predictive contexts over the course of the session. These results are summarized in Fig. 4. The cumulative distribution of first-adoption times for the length-2 contexts is shown in Fig. 4A. The context $(0,1)$ was typically discovered at early stages, whereas $(1,1)$ and $(2,1)$ emerged later in the session, with $(2,1)$ being ultimately adopted by a larger fraction of participants than $(1,1)$. The fraction of participants whose inferred trees contained the ground-truth (GT) contexts gradually increased throughout the experiment, as shown in Fig. 4B, indicating that most participants progressively converged toward the optimal set of predictive dependencies.

As suggested by the model of time intervals between context-tree transition events, context lifetimes are well described by a Weibull distribution. To examine the temporal stability of the inferred contexts, we separated those belonging to the ground-truth (GT) context tree from those that did not. The resulting lifetime distributions are shown in Fig. 4C. In both cases, the empirical data were well fitted by Weibull models but with distinct shape parameters, indicating different temporal regimes for GT and non-ground-truth (non-GT) contexts. Specifically, the fits yielded shape parameters $\beta_{\mathrm{GT}} = 0.72$ and $\beta_{\mathrm{nonGT}} = 0.93$, with corresponding scale parameters $\lambda_{\mathrm{GT}} = 383$ and $\lambda_{\mathrm{nonGT}} = 351$. The subunitary values of β imply decreasing hazard rates, meaning that the longer a context persists, the less likely it is to be replaced. This effect is stronger for GT contexts, consistent with their greater consolidation and stability once participants adopt them. This interpretation aligns with the survival curves shown in Fig. 4D, where GT contexts exhibit noticeably longer tails than non-GT contexts.

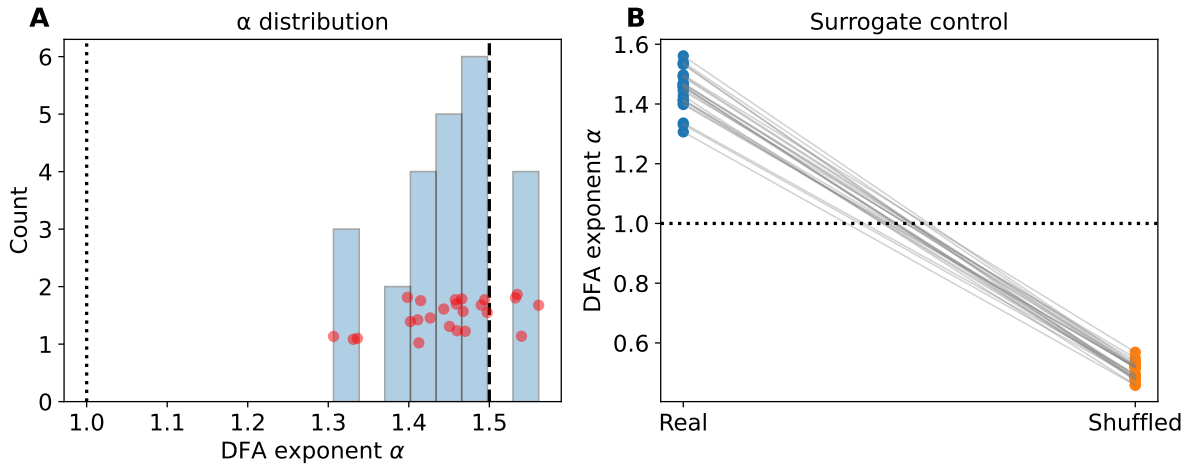


Figure 5. Scaling properties of entropy dynamics assessed by detrended fluctuation analysis (DFA). (**A**) Distribution of DFA scaling exponents (α) computed from the entropy time series of each participant. Individual subject values are overlaid (red dots). Vertical dotted and dashed lines mark reference slopes $\alpha = 1$ (often associated with 1/f–like critical fluctuations) and $\alpha = 1.5$ (Brownian boundary), respectively. Exponents cluster tightly around $\alpha \approx 1.3 - 1.5$, indicating strongly persistent, nonstationary dynamics rather than critical $1/f$ scaling. (**B**) Surrogate control analysis. For each participant, the DFA exponent computed from the original entropy trace ("Real") is compared to that obtained after temporal shuffling ("Shuffled"). Shuffling collapses α toward values consistent with uncorrelated noise, confirming that the observed persistence reflects genuine temporal structure. The horizontal dotted line indicates $\alpha = 1$ for reference.

3.5 Scaling analysis of entropy dynamics

To evaluate whether the entropy dynamics of the inferred context trees exhibit signatures compatible with self-organized criticality (SOC), we applied detrended fluctuation analysis (DFA) to the entropy time series $S(t)$ of each participant. DFA quantifies long-range temporal correlations through the scaling relation $F(s) \sim s^{\alpha}$, where $\alpha = 1$ is commonly associated with $1/f$–like critical fluctuations, $\alpha = 0.5$ with uncorrelated noise, and $1 < \alpha < 1.5$ with persistent, nonstationary dynamics characteristic of integrated stochastic processes.

Fig. 5 presents the distribution of DFA exponents across participants. Exponents clustered tightly within the range $\alpha \approx 1.3{-}1.5$, with a population mean near $\alpha \approx 1.4$ and a minimum value close to $\alpha \approx 1.3$. Only a small subset of subjects exhibited exponents exceeding 1.5, and critically, none showed values near $\alpha = 1$. Thus, the entropy dynamics are systematically shifted away from the regime typically associated with critical $1/f$ scaling.

Fig. 5 presents the surrogate control analysis comparing original and temporally shuffled entropy traces. The quality of the log–log linear fits was consistently high across participants. For the original entropy traces, the DFA regressions yielded a mean $R^2 = 0.9935$, with a minimum $R^2 = 0.9825$ and a 95% confidence interval for the mean of $[0.9917, 0.9953]$. For shuffled surrogate traces, the mean R^2 was similarly high ($R^2 = 0.9951$, minimum $R^2 = 0.9860$), indicating that linear scaling can still be obtained after destroying temporal correlations. However, shuffling systematically reduced the scaling exponent toward values consistent with uncorrelated noise, demonstrating that the elevated exponents in the real data arise from genuine temporal structure rather than fitting artifacts.

The observed regime $1 < \alpha < 1.5$ for most subjects is compatible with a random walk with memory, i.e., an integrated process with persistent increments (fractional Brownian motion-like dynamics). In this range, the dynamics correspond to subdiffusive fractional Brownian motion, since the mean-square entropy fluctuations scale as

$$\langle \Delta S^2(\tau) \rangle \sim \tau^{2(\alpha-1)},$$

with diffusion exponent $\gamma = 2(\alpha - 1) < 1$. This regime produces strong persistence and nonstationary scaling without requiring critical $1/f$ organization. Therefore, although entropy dynamics exhibit robust long-range correlations, their scaling properties are not consistent with a simple self-organized critical interpretation based on $1/f$-type critical fluctuations. Instead, they are better described as subdiffusive, persistent accumulation dynamics, consistent with a gradual refinement process punctuated by occasional structural reorganizations.

4 Discussion

The learning dynamics revealed by the context-tree inference demonstrate that participants adapt their internal predictive models through two intertwined yet distinct processes: refinement and transition. Refinement corresponds to continuous, small-scale adjustments to next-symbol probabilities within an unchanged tree structure, whereas transitions represent discrete reorganizations in which the topology of the inferred model is modified by adding or removing contexts. The predominance of refinement events and the rarity of transitions indicate that human learning in this task proceeds primarily through gradual exploitation of existing predictive structures, punctuated by occasional exploratory restructurings.

From a computational perspective, this dual-process organization mirrors the well-established exploration-exploitation trade-off. During refinements, participants exploit accumulated information to fine-tune their probabilistic expectations, reducing entropy and stabilizing predictions. Transitions, in contrast, correspond to exploration events: abrupt shifts in the internal model triggered when the current structure fails to accommodate new evidence. The alternation between these two regimes enables the learner to remain both stable and flexible, preserving reliable representations while maintaining the capacity to restructure when prediction errors accumulate. This dynamic is reminiscent of drift-diffusion models of decision making, in which evidence accumulates to a decision threshold; recent work shows that random exploration can arise from changes in the signal-to-noise ratio or threshold that controls evidence accumulation[20]. The exploration–exploitation dilemma is fundamental across psychology, neuroscience of decision-making, and machine learning; tasks such as multi-armed bandit problems are widely used to study how subjects balance reward maximization and information gain[21,22]. In reinforcement-learning contexts, exploration is favoured under low knowledge and high uncertainty, whereas exploitation is favoured when knowledge is high and uncertainty is low, and exploration and exploitation represent endpoints on a continuum from broad, noisy information-seeking to narrow, efficient reward-seeking[23]. Developmental evidence indicates that children initially explore broadly and shift towards exploitation as knowledge accumulates[24].

Consistent with this interpretation, the entropy time series exhibits anomalous diffusion in the subdiffusive regime. Detrended Fluctuation Analysis yielded exponents $1 < \alpha < 1.5$, which, for fractional Brownian motion, correspond to a diffusion exponent $\gamma = 2(\alpha - 1) < 1$. Hence, mean-square entropy fluctuations grow sublinearly with time, $\langle \Delta S^2(\tau) \rangle \sim \tau^{2(\alpha-1)}$, indicating persistent yet constrained accumulation dynamics. Although entropy fluctuations are nonstationary and long-range correlated, their dispersion increases more slowly than in Brownian motion. This behaviour reflects gradual, memory-dependent consolidation rather than ballistic or rapidly expanding exploration.

This subdiffusive informational dynamics stands in contrast with anomalous diffusion models developed for random searches and biological foraging. In those settings, heavy-tailed step-length distributions, such as Lévy flights, can optimize search efficiency when targets are sparse and revisitable[25]. Such Lévy-type movements generate superdiffusive spreading, with $\langle x^2(t) \rangle \propto t^\gamma$ and $\gamma > 1$, due to rare but long relocations that accelerate spatial coverage. Intermittent search strategies further combine slow detection phases with fast relocation phases to minimize search time[26]. More recent analyses show that the optimal Lévy exponent depends on environmental scale and prey distribution, with Gaussian walks favoured in large homogeneous domains and Lévy strategies favoured in smaller or patchy environments[27]. Even critiques of the Lévy-flight hypothesis emphasize the environmental conditions under which Brownian motion may outperform superdiffusive searches[28]. Thus, while optimal random searches often rely on superdiffusive exploration, the entropy evolution observed here follows a subdiffusive regime, reflecting restrained and history-dependent refinement rather than rare long informational jumps.

Further supporting this claim, the inter-transition intervals follow a sub-exponential Weibull law with shape parameter $\beta \approx 0.75$. The superiority of the Weibull renewal model over exponential, gamma, and power-law alternatives indicates that context-tree reorganizations are not memoryless. Instead, they follow a heavy-tailed waiting-time law, meaning that changes in the participants' predictive strategies occur in intermittent updates shaped by the system's past evolution. These renewal statistics imply temporal correlations: the longer the system remains in a given configuration, the lower the probability of reorganization. Such behaviour is consistent with consolidation and with reinforcement-learning dynamics in which exploration decreases as confidence in the current model increases, reinforcing the picture of learning as a predominantly exploitative, subdiffusive accumulation process punctuated by rare exploratory restructuring events.

The temporal organization of entropy changes suggests an underlying evidence-accumulation mechanism. Refinements progressively integrate predictive information, effectively loading the internal model until a stability threshold is reached, at which point a discrete context-tree transition occurs. In this interpretation, refinements correspond to exploitation, while transitions represent exploratory reorganisation when prediction errors accumulate. The separation between slow accumulation and rapid reorganisation resembles the evidence-accumulation dynamics

observed in decision making and is consistent with drift-diffusion models of explore–exploit choices[20].

Although heavy-tailed waiting times and long-range correlations are sometimes linked to near-critical systems, our detrended-fluctuation analysis (DFA) shows that the scaling exponent α is about 1.4. In DFA, $\alpha \approx 0.5$ corresponds to uncorrelated white noise, $\alpha \approx 1$ indicates pink $(1/f)$ noise typical of critical states, and values between 1 and 1.5 reflect an under-diffusive fractional Brownian motion with persistent positive correlations. Moreover, $\alpha > 1$ signals reduced complexity and stronger persistence. The fact that α is closer to 1.4 rather than 1 therefore rules out classic self-organised criticality; instead, it points to persistent, Brownian-like exploitation[29].

The Weibull waiting-time distribution we observe has a shape parameter less than unity, implying a decreasing hazard rate—once a context is adopted, the probability of switching declines over time. Such decreasing hazard rates and heavy-tailed statistics can arise naturally from memory effects, consolidation and evidence accumulation rather than from criticality[30]. In this view, refinements act as microscopic loading events that increase predictive tension, while context-tree transitions serve as relaxation events that reorganize the internal model. Long-lived contexts function as attractors in the space of internal models, stabilizing performance while permitting occasional exploratory restructurings.

Thus, the learning dynamics in the Goalkeeper Game reflect a balance between exploitation and exploration. Refinements exploit accumulated information, whereas transitions represent exploratory shifts triggered when prediction errors can no longer be accommodated within the existing model. The heavy-tailed timing of transitions and the persistent dynamics revealed by the α exponent suggest that participants' internal models evolve through punctuated equilibria driven by consolidation and evidence accumulation rather than by self-organised critical dynamics[31–34].

Altogether, the results indicate that the adaptation of internal predictive models during human learning is neither purely continuous nor purely discrete, but rather governed by an interplay between smooth probabilistic refinements and abrupt structural transitions. This dynamic resembles a balance between evidence accumulation and world model reorganization, reflecting the broader cognitive principle of exploration–exploitation trade-offs. The heavy-tailed timing and magnitude of these events suggest that participants' internal models evolve through punctuated equilibria rather than gradual convergence. This adaptive strategy may generalize to other domains of cognitive and motor learning where stability and flexibility must coexist.

5 Conclusion

In this work, we introduced a real-time context-tree inference framework to reconstruct the internal predictive models underlying human learning in the Goalkeeper Game. By analyzing the sequence of participants' choices, we revealed that adaptive behavior emerges from two coupled processes acting on distinct time scales. Frequent refinements correspond to incremental updates of transition probabilities within an unchanged tree structure, reflecting the gradual exploitation of acquired knowledge. In contrast, rare transitions correspond to structural reorganizations of the inferred context tree, in which new predictive contexts are added or pruned. Together, these processes capture the continuous–discrete duality of human learning: smooth probabilistic adjustments punctuated by abrupt representational shifts.

Quantitatively, refinements dominated the entropy dynamics, driving a progressive decrease in uncertainty that paralleled the increase in predictive accuracy. Transitions were infrequent but followed a sub-exponential Weibull waiting-time distribution, indicating a bursty, history-dependent process of model reorganization. Participants with lower model entropy achieved higher success rates, confirming that the inferred context trees faithfully reflect the internalization of the kicker's statistical structure. The temporal pattern of refinements and transitions thus provides a direct window into the microdynamics of evidence accumulation and structural adaptation during learning.

Conceptually, the coexistence of refinement and transition processes mirrors the exploration–exploitation trade-off that characterizes adaptive cognition. Refinements exploit the accumulated evidence to stabilize predictions, while transitions instantiate exploratory restructuring events triggered when prediction errors can no longer be accommodated within the current model. The heavy-tailed statistics of transition intervals and the decreasing hazard rates of context lifetimes suggest that these reorganizations are not memoryless but governed by consolidation mechanisms reminiscent of critical-like dynamics. In this view, refinements accumulate informational "tension" that is intermittently released through structural transitions—an analogy to energy buildup and relaxation cycles observed in self-organized critical systems.

Methodologically, the real-time context-tree estimation approach provides a transparent and interpretable tool for modeling the evolving internal representations that guide behavior. Unlike batch-based models, it tracks the learner's predictive structure trial-by-trial, enabling a fine-grained analysis of how uncertainty, stability, and flexibility

coevolve during adaptive behavior. This framework can be readily extended to other domains where predictive structures change dynamically, including motor learning, decision-making, and joint coordination.

Data Availability Statement

The data and the code of the algorithms used in the analysis are available at the following repository: https://github.com/italoivo/Goalkeeper_game Instructions about the use of the algorithms are presented in README files included in the repository.

Funding

This work is part of the activities of the São Paulo Research Foundation (FAPESP)'s Research, Innovation and Dissemination Center for Neuromathematics (NeuroMat, grant 2013/ 07699-0). I.L.D.P was supported by the FAPESP (2025/09647-4) fellowship. P.R.C.P. was supported by the FAPESP fellowship (2022/00699-3). C.D.V. is supported by FAPERJ (grants CNE 202.785/2018, E-26/010.002418/2019 and CNE E-26/204.076/2024), CNPq grants (310397/2021-9; 407092/2023-4,408389/2024-9), and INCT CNPq 408590/2024-6. FL is supported by FAPESP (thematic project 2023/13453-5), CNPq (grants 441884/2023-7 and 305688/2025-1), Capes MathAmsud (88881.179479/2025-01), and INCT CNPq 408590/2024-6.

Author contributions statement

I.I.L.D.P. developed the real-time context-tree inference framework, implemented all computational analyses, performed the simulations, generated all figures, and wrote the first draft of the manuscript. P.R.C.P. contributed to the implementation of the goalkeeper game experimental setup, coordinated data acquisition, and assisted in data preprocessing and interpretation. P.S.A. contributed to experimental design, participant recruitment, and oversight of data collection at UFRJ. F.A.N. provided methodological guidance on variable-length Markov models, advised on statistical modeling, and contributed to analytical tool development. P.B. provided methodological guidance and contributed to the interpretation of the detrended fluctuation analysis results. C.D.V. supervised all behavioral and experimental procedures, contributed to conceptual framing, and provided critical revisions. F.L. co-conceived the study, supervised the development of the inference algorithm, provided theoretical guidance on probabilistic context trees, and contributed extensively to data interpretation and manuscript revision. All authors reviewed and approved the final manuscript.

References

1. Wang, Y., Lu, L. & Wu, M. Statistical learning across cognitive and affective domains: a multidimensional review. *Front. Integr. Neurosci.* **19**, 1460471, DOI: 10.3389/fnint.2025.1460471 (2025). Published 9 May 2025.

2. Chander, A. & Aslin, R. N. Expectation adaptation for rare cadences in music: Item order matters in repetition priming. *Cognition* **240**, 105601, DOI: 10.1016/j.cognition.2023.105601 (2023). Available online 19 Aug 2023.

3. Rissanen, J. A universal data compression system. *IEEE Transactions on Inf. Theory* **29**, 656–664, DOI: 10.1109/TIT.1983.1056741 (1983).

4. Harrison, P. M. C., Bianco, R., Chait, M. & Pearce, M. T. Ppm-decay: A computational model of auditory prediction with memory decay. *PLOS Comput. Biol.* **16**, e1008304, DOI: 10.1371/journal.pcbi.1008304 (2020). Published Nov 4 2020.

5. Schr"oger, E., Roeber, U. & Coy, N. Markov chains as a proxy for the predictive memory representations underlying mismatch negativity. *Front. Hum. Neurosci.* **17**, 1249413, DOI: 10.3389/fnhum.2023.1249413 (2023). Published 13 Sep 2023.

6. Lima Dias Pinto, I. I., Zhou, Z., Garcia, J. O. & Srinivasan, R. Symbolic dynamics of joint brain states during dyadic coordination. *Chaos: An Interdiscip. J. Nonlinear Sci.* **35**, DOI: 10.1063/5.0234902 (2025).

7. Hernández, N. *et al.* Retrieving the structure of probabilistic sequences of auditory stimuli from eeg data. *Sci. Reports* **11**, 3520, DOI: 10.1038/s41598-021-83119-x (2021).

8. Cabral-Passos, P. R. *et al.* Retrieving the structure of probabilistic sequences from eeg data during the goalkeeper game. *arXiv preprint arXiv:2510.06344* DOI: 10.48550/arXiv.2510.06344 (2025).

9. Cabral-Passos, P. R., Galves, A., Garcia, J. E. & Vargas, C. D. Response times are affected by mispredictions in a stochastic game. *Sci. Reports* **14**, 8446, DOI: 10.1038/s41598-024-58203-7 (2024).

10. Hernández, N., Galves, A., García, J. E., Gubitoso, M. D. & Vargas, C. D. Probabilistic prediction and context tree identification in the goalkeeper game. *Sci. Reports* **14**, 15467, DOI: 10.1038/s41598-024-66009-w (2024).

11. Brockbank, E. & Vul, E. Repeated rock, paper, scissors play reveals limits in adaptive sequential behavior. *Cogn. Psychol.* **151**, 101654, DOI: 10.1016/j.cogpsych.2024.101654 (2024). Available online 23 Apr 2024.

12. Manning, C. D. *Introduction to information retrieval* (Syngress Publishing,, 2008).

13. Mardia, J., Jiao, J., Tánczos, E., Nowak, R. D. & Weissman, T. Concentration inequalities for the empirical distribution of discrete distributions: beyond the method of types. *Inf. Inference: A J. IMA* **9**, 813–850, DOI: 10.1093/imaiai/iaz025 (2020).

14. Belloni, A. & Oliveira, R. I. Approximate group context tree. *The Annals Stat.* **45**, 355 – 385, DOI: 10.1214/16-AOS1455 (2017).

15. Galves, A., Maume-Deschamps, V. & Schmitt, B. Exponential inequalities for vlmc empirical trees. *ESAIM: Probab. Stat.* **12**, 219–229, DOI: 10.1051/ps:2007035 (2008).

16. Gabadinho, A. & Ritschard, G. Analyzing state sequences with probabilistic suffix trees: The pst r package. *J. statistical software* **72**, 1–39, DOI: 10.18637/jss.v072.i03 (2016).

17. Kim, D. & Walder, C. Self-bounded prediction suffix tree via approximate string matching. In *International Conference on Machine Learning*, 2659–2667, DOI: 10.48550/arXiv.1802.03184 (PMLR, 2018).

18. Weibull, W. A statistical distribution function of wide applicability. *J. applied mechanics* DOI: 10.1115/1.4010337 (1951).

19. Almalki, S. J. & Nadarajah, S. Modifications of the weibull distribution: A review. *Reliab. Eng. & Syst. Saf.* **124**, 32–55, DOI: 10.1016/j.ress.2013.11.010 (2014).

20. Feng, S. F., Wang, S., Zarnescu, S. & Wilson, R. C. The dynamics of explore–exploit decisions reveal a signal-to-noise mechanism for random exploration. *Sci. Reports* **11**, DOI: 10.1038/s41598-021-82530-8 (2021).

21. Harlé, K. M. *et al.* Altered statistical learning and decision-making in methamphetamine dependence: Evidence from a two-armed bandit task. *Front. Psychol.* **6**, DOI: 10.3389/fpsyg.2015.01910 (2015).

22. Hills, T. T., Todd, P. M., Lazer, D., Redish, A. D. & Couzin, I. D. Exploration versus exploitation in space, mind, and society. *Trends Cogn. Sci.* **19**, 46–54, DOI: 10.1016/j.tics.2014.10.004 (2015).

23. Auer, P. Using confidence bounds for exploitation-exploration trade-offs. *J. machine learning research* **3**, 397–422 (2002).

24. Kim, S. & Carlson, S. M. Understanding explore-exploit dynamics in child development: current insights and future directions. *Front. Dev. Psychol.* **2**, DOI: 10.3389/fdpys.2024.1467880 (2024).

25. Viswanathan, G. M. *et al.* Optimizing the success of random searches. *Nature* **401**, 911–914, DOI: 10.1038/44831 (1999).

26. Bénichou, O., Loverdo, C., Moreau, M. & Voituriez, R. Intermittent search strategies. *Rev. Mod. Phys.* **83**, 81–129, DOI: 10.1103/RevModPhys.83.81 (2011).

27. Dipierro, S., Giacomin, G. & Valdinoci, E. The lévy flight foraging hypothesis: comparison between stationary distributions and anomalous diffusion. *J. Phys. A: Math. Theor.* **56**, 485601, DOI: 10.1088/1751-8121/ad01ff (2023).

28. Palyulin, V. V., Chechkin, A. V. & Metzler, R. Lévy flights do not always optimize random blind search for sparse targets. *Proc. Natl. Acad. Sci.* **111**, 2931–2936, DOI: 10.1073/pnas.1320424111 (2014).

29. Lovecchio, E., Allegrini, P., Geneston, E., West, B. J. & Grigolini, P. From self-organized to extended criticality. *Front. Physiol.* **3**, 98, DOI: 10.3389/fphys.2012.00098 (2012).

30. Abaimov, S. G., Turcotte, D. L., Shcherbakov, R. & Rundle, J. B. Recurrence and interoccurrence behavior of self-organized complex phenomena. *Nonlinear Process. Geophys.* **14**, 455–464, DOI: 10.5194/npg-14-455-2007 (2007).

31. Lombardi, F., Pepić, S., Shriki, O., Tkačik, G. & De Martino, D. Statistical modeling of adaptive neural networks explains co-existence of avalanches and oscillations in resting human brain. *Nat. Comput. Sci.* **3**, 254–263, DOI: 10.1038/s43588-023-00410-9 (2023).

32. Lombardi, F. *et al.* Beyond pulsed inhibition: Alpha oscillations modulate attenuation and amplification of neural activity in the awake resting state. *Cell Reports* **42**, 113162, DOI: 10.1016/j.celrep.2023.113162 (2023).

33. Wang, J. W. J. L., Lombardi, F., Zhang, X., Anaclet, C. & Ivanov, P. C. Non-equilibrium critical dynamics of bursts in theta and delta rhythms as fundamental characteristic of sleep and wake micro-architecture. *PLOS Comput. Biol.* **15**, e1007268, DOI: 10.1371/journal.pcbi.1007268 (2019).

34. Huo, C., Lombardi, F., Blanco-Centurion, C., Shiromani, P. J. & Ivanov, P. C. Role of the locus coeruleus arousal promoting neurons in maintaining brain criticality across the sleep–wake cycle. *The J. Neurosci.* **44**, e1939232024, DOI: 10.1523/JNEUROSCI.1939-23.2024 (2024).

Appendix

A Parametric fitting of per-window moments

This appendix describes the procedure used to estimate per-window moments and fit their temporal trajectories. The time axis was divided into equally spaced windows with centers $t \in [0,1]$, and for each window the sample mean, standard deviation, and skewness were computed. Each trajectory was then modeled as a single-exponential relaxation of the form

$$y(t) = y_\infty + (y_0 - y_\infty)e^{-kt}, \quad k > 0,$$

where y_0 and y_∞ are the initial and asymptotic values, and k is the relaxation rate. This functional form was applied consistently to all moments, yielding:

$$\mu(t) = \mu_\infty + (\mu_0 - \mu_\infty)e^{-k_\mu t}, \tag{2}$$

$$\sigma(t) = \sigma_\infty + (\sigma_0 - \sigma_\infty)e^{-k_\sigma t}, \tag{3}$$

$$s(t) = s_\infty + (s_0 - s_\infty)e^{-k_s t}. \tag{4}$$

Parameters were estimated by weighted nonlinear least squares using the `scipy.optimize.least_squares` routine with inverse-variance weights derived from the estimated standard errors of each window. Initialization values were taken directly from the data: $y_0^{(0)}$ as the first-window value, $y_\infty^{(0)}$ as the last-window value (or the median of the last third of windows for σ), and $\log k^{(0)} = 0$ corresponding to $k = 1$. To enforce positivity, rate parameters were fitted in log space. For $s(t)$, small numerical floors were applied to ensure positivity throughout the fitting. All optimizations were performed on normalized window centers $t \in [0,1]$.

The fitted parameter values obtained from the weighted least-squares estimation are summarized in Table 1:

Table 1. Fitted parameter values

Moment	y_∞	y_0	k
Mean (μ)	-4.29×10^{-4}	-6.08×10^{-3}	19.8
Std. dev. (σ)	0.0087	0.0255	2.03
Skewness (s)	11.5	5.0×10^{-9}	1.83

These values indicate a rapid convergence of the mean toward zero, a progressive reduction of variability, and a marked increase in skewness over time, reflecting the gradual stabilization and asymmetry buildup in the data distribution.