Preprints.org

Article

The Rosencrantz Coin: Predictability and Structure in Non-Ergodic Dynamics. From Recurrence Times to Temporal Horizons

Dimitri Volchenkov [*]

Article

The Rosencrantz Coin: Predictability and Structure in Non-Ergodic Dynamics. From Recurrence Times to Temporal Horizons

Dimitri Volchenkov [ORCID]

Texas Tech University, Department of Mathematics and Statistics, 1108 Memorial Circle, Lubbock, TX 79409 USA;
dimitri.volchenkov@ttu.edu

Abstract: We examine the Rosencrantz coin that can "stick" in states for extended periods. Non-ergodic dynamics is highlighted by logarithmically growing block lengths in sequences. Traditional entropy decomposition into predictable and unpredictable components fails due to the absence of stationary distributions. Instead, sequence structure is characterized by block probabilities and Stirling numbers of the second kind, peaking at block size $n/\log n$. For large n, combinatorial growth dominates probability decay, creating a deterministic-like structure. This approach shifts the focus from predicting states to predicting temporal horizons, providing insights into systems beyond traditional equilibrium frameworks.

Keywords: entropy decomposition; characteristic times; non-ergodic dynamics

1. Introduction

Predicting the future state of a system is a fundamental challenge across various fields, from physics and information theory to economics and cognitive science [1]. Simple systems, such as a biased coin, can encapsulate profound insights into uncertainty, entropy, and information dynamics. In traditional ergodic systems, where the dynamics ensure that time averages converge to ensemble averages, predictability is often assessed using entropy and its decomposition into predictable and unpredictable components [2]. These components are characterized by measures such as recurrence time, residence time, and repetition time, which capture the structure of state sequences and quantify the degree of uncertainty (see Section 2).

However, many real-world systems exhibit non-ergodic behavior, where the assumption of visiting all possible states uniformly over time breaks down. In such systems, conventional entropy-based methods become inadequate because there is no stationary distribution to describe long-term behavior. Instead, non-ergodic systems are marked by prolonged persistence in certain states, leading to sequences dominated by blocks of repeated outcomes. A striking example of such a system can be found in Tom Stoppard's play *Rosencrantz and Guildenstern Are Dead*, where Rosencrantz experiences an improbable streak of 92 consecutive heads [3] that challenge traditional probabilistic frameworks. This anomaly prompts deeper exploration into non-ergodic dynamics, where long-term correlations lead to prolonged persistence in specific states. The Rosencrantz coin model (Section 3) provides a useful abstraction for studying systems with long-term correlations and memory effects. In this model, the barrier to switching states may remain constant or change intermittently, governed by a stochastic threshold. This behavior results in block-like structures within the sequences, where the lengths of these blocks increase logarithmically over time, leading to sequences dominated by blocks of persistent states. The model offers insights into the interplay between randomness and determinism, highlighting how predictable patterns emerge even in fundamentally stochastic processes (see Section 4).

To analyze these patterns, we explore the combinatorial structures governing the emerging block sequences, particularly focusing on the Stirling numbers of the second kind, which count the ways to partition sequences into non-empty blocks. The Stirling numbers reveal that for sufficiently long sequences, the most probable partitioning involves blocks of length approximately $\ln n$, where n is the sequence length. This combinatorial insight compensates for the lack of a stationary distribution and enables meaningful predictions of the system's temporal behavior (Section 4).

Our analysis shows that in a non-ergodic system, the balance between combinatorial growth and the exponential decay of block probabilities results in a deterministic-like structure within a fundamentally random process. Instead of predicting individual outcomes, the focus shifts to predicting the length of sequences based on the observed block size. The concept of a *temporal horizon* — the characteristic length of time over which predictable patterns emerge — becomes central in non-ergodic systems. Instead of predicting the next state, the focus shifts to predicting the duration of structured behavior based on observed block sizes (Section 4). The logarithmic utility of time for prediction the future, which reflects diminishing "returns" on prediction, further connects these ideas to hyperbolic time discounting models often found in human and animal decision-making (Section 5). This novel approach offers a deeper understanding of the dynamics of non-ergodic systems, bridging concepts from information theory, combinatorics, and stochastic processes.

The paper is organized as follows: Section 2 discusses entropy decomposition in stationary Markov chains, highlighting predictable and unpredictable components and introducing the notions of characteristic times. Section 3 introduces the Rosencrantz coin model and explores its stochastic dynamics. Section 4 delves into the structure and dynamics of sequences in the non-ergodic Rosencrantz coin model, focusing on the combinatorial properties of block partitions. Section 5 presents a detailed discussion of the findings, and Section 6 concludes the paper by summarizing the key insights and implications for non-ergodic systems.

2. Decomposition of Entropy in Stationary Markov Chains – Predictable and Unpredictable Information

An outcome of each coin flip, governed by the transition matrix $\begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}$, encapsulates a single bit of information. The probability of state repetition, $0 \leq p \leq 1$ determines how this information is divided into predictable and unpredictable components [2]. When $p = 1$, the sequence becomes perfectly predictable, locking outcomes into stationary patterns like $H, H, \ldots$ (*heads*) or $T, T, \ldots$ (*tails*). Similarly, when $p = 0$, the sequence alternates deterministically: $H, T, H, T, \ldots$. In contrast, $p = 1/2$ corresponds to complete randomness, making future outcomes wholly unpredictable.

For long sequences ($n \gg 1$) generated by an N-state, irreducible, and recurrent Markov chain $\{X_t\}_{t \geq 0}$ with a stationary transition matrix $T_{ks} = \Pr[X_t = s \mid X_{t-1} = k]$, the relative frequency n_k/n of visits to state k converges to the stationary distribution π_k, such that $\sum_{k=1}^{N} \pi_k T_{ks} = \pi_s$ for all s. The probability of each specific sequence $\{i_1, i_2, \ldots, i_n\}$ becomes extremely small for large n, decreasing as $\pi_{i_1}^{n\pi_{i_1}} \pi_{i_2}^{n\pi_{i_2}} \ldots \pi_{i_n}^{n\pi_{i_n}}$. Consequently, the number of distinct sequences consistent with the stationary distribution π decreases exponentially with n, viz.,

$$\mathfrak{N}_\pi(n) = \frac{n!}{(n\pi_1)! \ldots (n\pi_N)!} \simeq e^{-n\left[\sum_{k=1}^{N} \pi_k \ln(n\pi_k) - \ln n\right]} = \exp(-nH_\pi), \quad H_\pi \equiv -\sum_{k=1}^{N} \pi_k \ln \pi_k \quad (1)$$

where H_π is the *decay rate* at which the number of different sequences constrained by the stationary distribution π decreases with sequence length n. Up to a factor of $1/\ln N$, the decay rate H_π, corresponds to the Boltzmann-Gibbs-Shannon entropy quantifying the uncertainty of a Markov chain's state in equilibrium:

$$H(X_t) = -\sum_{k=1}^{N} \pi_k \log_N \pi_k = \sum_{k=1}^{N} \frac{\log_N \mathcal{R}_k}{\mathcal{R}_k}, \quad \mathcal{R}_k \equiv \frac{1}{\pi_k} = \lim_{n \to \infty} \frac{n}{n_k}. \quad (2)$$

The inverse frequency $\mathcal{R}_k$ in (2) is the expected *recurrence time* of sequence returns to state k. Thus, $\ln \mathcal{R}_k$ can be termed the *utility function* of recurrence time, quantifying the reduction in diversity of state sequences caused by the repetition of state k in the most likely sequence patterns corresponding to the stationary distribution π.

The entropy (2) serves as a foundation for decomposing the system's total uncertainty into predictable and unpredictable components [2,4,5]. To analyze the information dynamics, we add and subtract the following conditional entropy quantities to $H(X_t)$, grouping the resulting terms into distinct informational quantities:

$$
\begin{aligned}
H(X_t) \quad &= H(X_t) \underbrace{\pm H(X_{t+1}|X_t) \pm H(X_t|X_{t-1})}_{E_{\text{ex}}(X_t)} \underbrace{\pm H(X_{t+1}|X_{t-1})}_{I(X_{t+1},X_t|X_{t-1})} \\
&= \overbrace{(H(X_t) - H(X_{t+1}|X_t))}^{} + \overbrace{(H(X_{t+1}|X_{t-1}) - H(X_t|X_{t-1}))}^{} \\
&\qquad\qquad\qquad\;\underbrace{\phantom{(H(X_t) - H(X_{t+1}|X_t))}}_{H(X_t|X_{t+1},X_{t-1})} \\
&+ \overbrace{(H(X_{t+1}|X_t) + H(X_t|X_{t-1}) - H(X_{t+1}|X_{t-1}))}^{}.
\end{aligned}
\tag{3}
$$

The excess entropy $E_{\text{ex}}(X_t)$ quantifies the influence of past states on the present and future states, capturing the system's structural correlations and predictive potential. The conditional mutual information $I(X_{t+1}, X_t|X_{t-1})$, represents the information shared between the current state X_t and the future state X_{t+1}, independent of past history. It vanishes for both fully deterministic and completely random systems, reflecting the absence of predictive utility in these extremes. The sum of the excess entropy and the conditional mutual information shown in the second line of (3) represents the *predictable information* in the system [2,4]. This component quantifies the portion of the total uncertainty in the future state X_{t+1} that can be resolved using information about the current state X_t and the past states within the system's dynamics. In contrast, the *unpredictable* component $H(X_t|X_{t+1}, X_{t-1})$, given in the third line of (3), quantifies the intrinsic randomness in the system that remains unresolved even with full knowledge of its history. The entropy decomposition in (3) is closed, meaning it completely partitions the total entropy $H(X_t)$ into predictable and the unpredictable components. For a fully deterministic system (e.g., $p = 1$ or $p = 0$), the unpredictable component vanishes, $H(X_t \mid X_{t+1}, X_{t-1}) = 0$, making the entire entropy predictable, $H(X_t) = E_{\text{ex}}(X_t)$. Conversely, for a completely random system ($p = 1/2$), the predictable components disappear. In this case, $H(X_t) = H(X_t \mid X_{t+1}, X_{t-1})$, as observations of the prior sequence provide no information for predicting future states ($E_{\text{ex}}(X_t) = 0$). Similarly, attempts to predict the next state by repeating or alternating the current state fail, as $I(X_{t+1}, X_t \mid X_{t-1}) = 0$.

In a finite-state, irreducible Markov chain, the entropy rate quantifies the uncertainty associated with transitions from one state to the next. It is formally given by:

$$
H(X_{t+1}|X_t) = -\sum_{k=1}^{N} \pi_k \sum_{s=1}^{N} T_{ks} \log_N T_{ks} = -\sum_{k=1}^{N} \pi_k \log_N \left(\prod_{s=1}^{N} T_{ks}^{T_{ks}} \right)
\tag{4}
$$

where the second formulation highlights the connection to the *geometric mean* of transition probabilities. The term $\prod_{s=1}^{N} T_{ks}^{T_{ks}}$ can be interpreted as the geometric mean time-averaged transition probability rate (per step) from the state k over an infinitely long observation period:

$$
\prod_{s=1}^{N} T_{ks}^{T_{ks}} = \lim_{n \to \infty} \sqrt[n]{\prod_{s=1}^{N} (T_{ks})^{n_{k,s}}} \equiv Q_k,
\tag{5}
$$

where $n_{k,s}/n$ is the observed frequency of transitions from k to s over the most likely sequence of length n. The frequency converges to the transition probability T_{ks} as $n \to \infty$. The inverse of the geometric mean transition probability rate, $R_k \equiv Q_k^{-1}$, represents the average *residence time* in state k. The excess entropy $E_{\text{ex}}(X_t)$ takes a form similar to the entropy (2):

$$
E_{\text{ex}}(X_t) = H(X_t) - H(X_{t+1}|X_t) = \sum_{k=1}^{N} \pi_k (\log_N \mathcal{R}_k - \log_N R_k) = \sum_{k=1}^{N} \frac{\log_N \mathcal{R}_k / R_k}{\mathcal{R}_k}
\tag{6}
$$

where the ratio of average recurrence and residence times, $\mathcal{R}_k / R_k$, measures the *transience* of state k in a sequence consistent with the stationary distribution π. Low transience implies that typical sequences

are more predictable. Conversely, in the case of maximally random sequences where $\mathcal{R}_k \approx R_k$, states are visited regularly, and the excess entropy (6) does not enhance predictability. The formula for excess entropy E_{ex} remains valid only for non-deterministic processes where all states can be visited and exited with non-zero probability. If the system remains indefinitely in a single state, the excess entropy becomes undefined because the system's behavior carries no uncertainty.

Similarly, the conditional entropy,

$$H(X_{t+1}|X_{t-1}) = -\sum_{k=1}^{N} \pi_k \log_N \prod_{s=1}^{N} \left(T^2\right)_{ks}^{(T^2)_{ks}} \equiv -\sum_{k=1}^{N} \pi_k \log_N Q_k^{(2)} \tag{7}$$

quantifies the uncertainty associated with the statistics of *trigrams* involving two transitions. The transition probability rate,

$$Q_k^{(2)} = \lim_{n \to \infty} \sqrt[n]{\prod_{s=1}^{N} (T^2)_{ks}^{n_{k,\bullet,s}}} = \prod_{s=1}^{N} \left(T^2\right)_{ks}^{(T^2)_{ks}}, \tag{8}$$

represents the asymptotic frequency of trigrams starting with state k. The relative frequency $n_{k,\bullet,s}/n$ converges to the corresponding elements of the squared transition matrix $\left(T^2\right)_{ks}$, over long, most likely sequences conforming the stationary distribution π. The conditional mutual information,

$$I(X_{t+1}, X_t|X_{t-1}) = H(X_{t+1}|X_{t-1}) - H(X_t|X_{t-1}) = \sum_{k=1}^{N} \pi_k \log_N \frac{Q_k}{Q_k^{(2)}}, \tag{9}$$

measures the amount of information shared between the current state X_t and the future state X_{t+1}, independently of the historical context provided by X_{t-1}. In the context of coin tossing, the mutual information (9) emerges from uncertainty in choosing between alternating the present coin side ($p \gtrsim 0$) or repeating the current side ($p \lesssim 1$) when predicting the coin's future state. The mutual information vanishes for the fair coin ($p = 1/2$), where transitions are completely random and independent. It also vanishes for a fully deterministic coin ($p = 1$ or $p = 0$), where past states provide no additional information for predicting future behavior.

When the past and present states of the chain are known, the predictable information about the future state can be expressed as:

$$\mathcal{P}(X_t) = E_{\text{ex}}(X_t) + I(X_{t+1}, X_t|X_{t-1}) = -\sum_{k=1}^{N} \pi_k \log_N \frac{\pi_k}{Q_k^2/Q_k^{(2)}} = \sum_{k=1}^{N} \frac{\log_N \mathcal{R}_k/\mathfrak{R}_k}{\mathcal{R}_k}, \quad \mathfrak{R}_k \equiv \frac{Q_k^{(2)}}{Q_k^2}, \tag{10}$$

The conditional probability $Q_k^2/Q_k^{(2)}$ represents the likelihood that, in a two-step transition starting from k, both steps remain in k. The inverse conditional probability, $\mathfrak{R}_k$, can be interpreted as the average state *repetition time* — the expected time between instances where state k appears consecutively twice. Predictable information, therefore, reflects the balance between recurrences (returns to states) and state repetitions (remaining in the same state). Systems with frequent recurrences and repetitions exhibit high predictability, indicating more organized and regular behavior. In contrast, systems with infrequent recurrences and frequent state changes exhibit low predictability, suggesting a more random and chaotic behavior. According to (3), unpredictable information is then defined as:

$$\mathcal{U}(X_t) = H(X_t) - \mathcal{P}(X_t) = \sum_{k=1}^{N} \pi_k \left(\log_N \mathcal{R}_k - \log_N \frac{\mathcal{R}_k}{\mathfrak{R}_k} \right) = \sum_{k=1}^{N} \frac{\log_N \mathfrak{R}_k}{\mathcal{R}_k}. \tag{11}$$

The logarithm of the state repetition time $\mathfrak{R}_k$, which can also be interpreted as the *utility* of state repetition, emphasizes the exponential growth of unpredictability when repetition times increase. Conversely, lower values of $\log_N \mathfrak{R}_k$ indicate reduced uncertainty and predictable behavior.

Figure 1 provides a detailed visualization of the entropy decomposition (3) for a biased coin modeled as a Markov chain with transition matrix

$$T(p,q) = \begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix}, \tag{12}$$

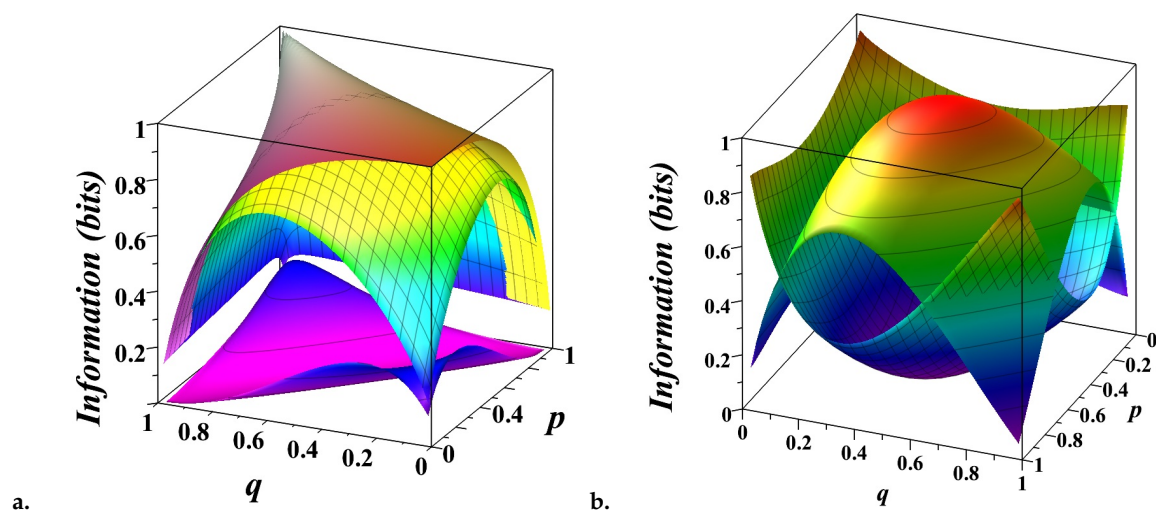where p and q are the probabilities of repeating the current state.

Figure 1. Decomposition of entropy $H(p,q)$ into its components (3) for a biased coin modeled as a Markov chain $T(p,q)$. **a).** The top surface shows the total entropy $H(p,q)$, the middle surface shows the entropy rate (7), and the bottom surface shows the conditional mutual information (9). The gap between the top and middle surfaces illustrate the excess entropy (6). The gap between the middle and the bottom surfaces corresponds to the unpredictable information component (11) of the system. **b.)** The top, convex surface shows the unpredictable component of information, reaching a maximum of 1 bit for a fair coin ($p = q = 1/2$), but reducing to zero for a deterministic coin ($p, q = 0, 1$).

In Figure 1.**a**, the three surfaces illustrate different information quantities as functions of p and q. The top surface represents the total entropy $H(X_t)$, capturing the overall uncertainty in the system's state. For a symmetric chain ($p = q$), the uncertainty reaches its maximum of 1 bit. The entropy decreases when $p \neq q$, as one state becomes more probable than another, reducing overall uncertainty. The middle surface shows the entropy rate $H(X_{t+1}|X_t)$, measuring the uncertainty in predicting the next state given the past states. When $p = 1 - q$, this surface coincides with the top one, indicating that the system behaves like a fair coin with no memory of past states. The bottom surface illustrates the conditional mutual information $I(X_{t+1}, X_t|X_{t-1})$, measures the predictive power of the current state for the next state, independent of past history. When $p = 1 - q$, the bottom surface highlights the tension between two simple prediction strategies — repeating or alternating the current state — reflecting the inherent randomness of a fair coin. The gaps between these surfaces reveal how the total uncertainty is partitioned. The gap between the top and middle surfaces corresponds to the excess entropy $E_{\text{ex}}(X_t)$, capturing the structured, predictable correlations in the system. Together with the space below the bottom surface, these gaps represent the amount of predictable information $\mathcal{P}(X_t)$. The gap between the middle and bottom surfaces represents the unpredictable component of the system, which diminishes as p and q approach deterministic values (0 or 1), reflecting minimal randomness.

In Figure 1.**b**, the decomposition of entropy is presented more intuitively. The top, convex surface shows the unpredictable component, peaking at 1 bit for a fair coin ($p = q = 1/2$) and dropping to zero for deterministic scenarios ($p, q = 0, 1$). The bottom, concave surface represents the predictable component, which grows as the system becomes more deterministic. This figure highlights the delicate balance between randomness and structure, emphasizing how the predictability of the system depends on the interplay between state repetition probabilities.

3. Stochastic Dynamics of Rosencrantz's Coin

In Tom Stoppard's play "*Rosencrantz and Guildenstern Are Dead*" [3], Rosencrantz improbably wins 92 consecutive coin flips, prompting Guildenstern to speculate that their situation may be influenced by supernatural forces. Indeed, he is correct—for their lives are no longer governed by chance but by the will of the king. Here, we present a simple probabilistic model in which such an improbable sequence can naturally arise, where long-term correlations effectively "freeze" the coin in one of its states. Our model builds on previous approaches using stochastic thresholds to study ecological subsistence dynamics and systems near critical instability points [6,7].

In this model, each "coin flip" corresponds to a step in a stochastic process. At each step, a random variable X, representing the energy or capability to transition, is drawn from a probability distribution $\Pr\{X < x\} = F(x)$. Similarly, the barrier height Y is also a random variable with its own distribution $\Pr\{Y < y\} = G(y)$. The process updates X at every step, reflecting the inherent randomness in the transition capability. However, the barrier height Y updates intermittently: with probability $0 \leq \eta \leq 1$, a new Y is drawn from $G(y)$; otherwise, Y retains its current value. The process remains in the current state at step t if $X_t \leq Y_t$, but transitions to another state if $X_t > Y_t$ (see Figure 2).

Figure 2. Illustration of a stochastic process where a dynamic barrier Y and random "energy" X determine state transitions (Rosencrantz's coin).

The total number of consecutive successful coin flips, τ, represents the duration of Rosencrantz's improbable streak — the time the process remains in the state where $X \leq Y$. This framework provides a simple probabilistic explanation for how extraordinary sequences of successes can arise.

The analysis of this model (see [6,7] for details) depends on the specific distributions $F(x)$ and $G(y)$, as well as on the value of the barrier update probability η. When the barrier is maximally volatile ($\eta = 0$), Y is updated at every step. In this regime, the probability of observing τ successful flips is given by

$$P_{\eta=0}(\tau) = \left(\int_0^1 dG(y) dF(y) \right)^\tau \int_0^1 dG(y)(1 - F(y)), \tag{13}$$

which decays exponentially with τ, regardless of the specific form of $F(x)$ and $G(y)$. The expected duration of state in such a sequence is always finite,

$$\sum_{\tau=0}^{\infty} \tau \cdot P_{\eta=0}(\tau + 1) < \infty. \tag{14}$$

For uniform distributions $dF(u) = dG(u) = du$ over $[0, 1]$, the coin behaves like a fair coin, and the probability of τ consecutive successes simplifies to:

$$P_{\eta=0}(\tau) = 2^{-(\tau+1)}, \tag{15}$$

and the expected duration of state equals $\sum_{\tau=0}^{\infty} \tau \cdot 2^{-(\tau+1)} = 1$.

When $\eta = 1$, the barrier Y remains constant throughout the process. In this case, the probability of τ successful flips is:

$$P_{\eta=1}(\tau) = \int_0^1 dG(y) F(y)^{\tau}(1 - F(y)). \tag{16}$$

For uniform distributions $dF(u) = dG(u) = du$ over $[0, 1]$, the probability (16) reduces to the Beta function $B(\tau + 1, 2) = \int_0^1 u^{\tau}(1 - u)du$. Thus, for large τ, the probability of state decays as:

$$P_{\eta=1}(\tau) = B(\tau + 1, 2) = \frac{1}{(\tau+1)(\tau+2)} \simeq \tau^{-2}, \tag{17}$$

The power-law decay in (17) implies that Rosencrantz's coin appears biased toward producing extended streaks of success, as the expected duration of state diverges, $\sum_{\tau=0}^{\infty} \tau \cdot P_{\eta=1}(\tau) = \infty$, indicating that blocks of any length can appear in sufficiently long sequences.

For intermediate values of $0 < \eta < 1$ with uniform distributions $dF(u) = dG(u) = du$ over $[0, 1]$, the barrier Y alternates between stability and variability. This dynamic creates a mixture of exponential and power-law decay for the probability of τ successful flips [6,7]:

$$P_{\eta}(\tau) = \frac{\eta^{\tau}}{(\tau+1)(\tau+2)} + \sum_{k=1}^{\tau} \frac{\eta^{\tau}}{k(\tau-k+1)(\tau-k+2)} \sum_{m=1}^{k} C_{m,k} \left(\frac{1-\eta}{\eta}\right)^m, \tag{18}$$

where the coefficients $C_{m,k}$ are defined by:

$$C_{m,k} = m! \sum_{\mathfrak{P}(k)} \frac{\prod_{s=1}^{m} l_s}{(l_m + 1)\prod_{s=1}^{m-1}[(l_s + 1)(k - \sum_{r=1}^{s} l_r)]}, \tag{19}$$

and $\mathfrak{P}(k)$ represents all integer partitions of k, viz., $\mathfrak{P}(k) = \{\{l_1, \ldots, l_m\} | l_i \geq 1, l_1 + \ldots + l_m = k, l_1 \geq l_2 \geq \ldots \geq l_m\}$.

If $G(y) = 1 - (1 - y)^{\varepsilon}$, $\varepsilon > 0$, the barrier height Y is typically close to one, making transitions rare and sharp. When X remains uniformly distributed, the probability of τ successful flips follows a Zipf-like behavior [7]:

$$P_{\eta=1}(\tau) = \frac{\tau^{-1-\varepsilon}}{\zeta(1+\varepsilon)}, \tag{20}$$

where $\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$ is the Riemann Zeta function. This result reflects the dominance of rare, extended streaks of success, evoking a sense of order emerging from randomness. The Rosencrantz coin can also be described by an integral row-stochastic matrix, viz.,

$$T_{\eta_1,\eta_2}(t) = \begin{pmatrix} P_{\eta_1}(t) & 1 - P_{\eta_1}(t) \\ 1 - P_{\eta_2}(t) & P_{\eta_2}(t) \end{pmatrix}, \tag{21}$$

which captures the transition probabilities over a finite interval t. Unlike the stationary Markov chain $T(p, q)$ (12), which describes instantaneous transitions, $T_{\eta_1,\eta_2}(t)$ reflects the cumulative effect of all probabilities up to time t.

In the absence of correlations ($\eta_{1,2} = 0$), when the height of the energy barrier Y changes at every time step and all transitions are statistically independent, the stationary chain $T(1/2, 1/2)$, corresponding to a fair coin, can be related to the Markov process generator $\mathfrak{G} = 1/\ln 2 (T(1/2, 1/2) - I)$ for the integral transition matrix $T_{0,0}(t)$, where I is the identity matrix. In this case, the integral matrix

$T_{0,0}(t) = \exp(t\mathfrak{G})$ is the solution of the differential equation $\dot{T}_{0,0}(t) = \mathfrak{G}T_{0,0}$, with the initial condition $T_{0,0} = I$. However, when correlated transitions occur ($\eta_{1,2} > 0$), the integral matrix $T_{\eta_1,\eta_2}(t)$ no longer has a straightforward relationship with the instantaneous transition matrix $T(p,q)$.

In the Rosencrantz coin model, the long-time behavior of recurrence times, which describe how long it takes, on average, for the system to return to a given state, depends significantly on the parameter η (see Figure 3). When $\eta_{1,2} = 0$, the barrier height changes dynamically with each step. In this case, the probabilities of coin's states decay as $2^{-(t+1)}$, and the recurrence time quickly converges to 2:

$$\mathcal{R}_k|_{\eta_{1,2}=0}(t) = 2 - \frac{2-t}{2^t} \to_{t\to\infty} 2. \tag{22}$$

This behavior corresponds to a fair coin with a stationary distribution of states $\pi_{1,2} = 1/2$. Here, the ergodic hypothesis holds, meaning the system visits all allowable states with equal frequency over a long enough time. Consequently, time averages and ensemble averages become equivalent.

In contrast, when $\eta_{1,2} = 1$, the barrier height remains fixed, and the recurrence time increases logarithmically. This growth is described by:

$$\mathcal{R}_k|_{\eta_{1,2}=1}(t) = 2\Psi(t+2) + \frac{4}{t+2} - 4 + 2\gamma \tag{23}$$

where $\Psi(x)$ is the digamma function, and γ is Euler's constant. In this scenario, the ergodic hypothesis does not hold because the recurrence times are not constant. The system's behavior becomes non-ergodic, meaning it does not visit all states with the same frequency, and ensemble averages lose their equivalence to time averages.
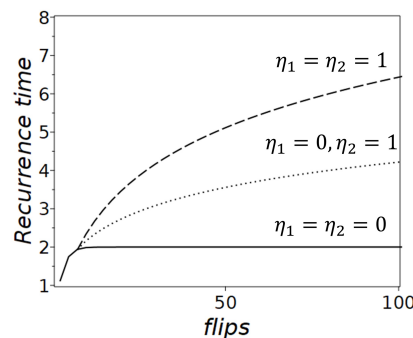

Figure 3. Recurrence times for the Rosencrantz coin model. The curves represent different η_1 and η_2: when $\eta_{1,2} = 0$ (solid line), shows a stable recurrence time of 2, consistent with a fair coin and the ergodic hypothesis; $\eta_1 = 0, \eta_2 = 1$ (dotted line) shows gradually increasing recurrence times; when $\eta_{1,2} = 1$ (dashed line) shows significantly increasing recurrence times.

When correlations are present, the Rosencrantz coin shows increasingly prolonged intervals before returning to a given state, rendering the concept of a stationary distribution invalid. The recurrence time grows without settling into a steady pattern, indicating that time averages diverge from ensemble averages and the system's dynamics are non-ergodic.

4. Structure and Dynamics of Sequences in the Non-Ergodic Rosencrantz Coin Model

To study the sequences generated by the non-ergodic Rosencrantz coin, entropy and its decomposition into predictable and unpredictable components are inapplicable because there is no stationary distribution. Figuratively speaking, the coin "sticks" in each of its states, meaning the structure of observed sequences is characterized by blocks of varying length where the coin remains in the same state for an extended period of time. As we have seen from the analysis of recurrence times (Section 3), the length of such blocks gradually increases, so that increasingly extended streaks of success would be observed in sufficiently long sequences. Because the expected recurrence times grow logarithmically,

the system's dynamics become increasingly non-stationary, and the typical assumptions of equilibrium and steady-state distributions no longer apply. The probability to observe such a sequence, consisting of m consecutive blocks of lengths $n_1, \ldots, n_m$, is characterized by the product $\prod_{i=1}^{m} P_{\eta=1}(n_i)$, where $n = n_1 + \ldots + n_m$ is a partition of n into m terms.

The number of ways to partition a sequence of n elements into m blocks is given by the Stirling numbers of the second kind $S(n, m)$ [8], which satisfy the recurrence relation:

$$S(n + 1, m) = S(n, m - 1) + mS(n, m). \tag{24}$$

The values of $S(n, m)$ are not uniformly distributed across all possible block sizes but exhibit a maximum for specific values of m. Figure 4.**a** shows the normalized Stirling numbers of the second kind, representing the probability of partitioning a sequence of length n into m non-empty blocks. The normalization uses corresponding Bell numbers, $B(n) = \sum_{1 \leq m \leq m} S(n, m)$, which count the total number of partitions of an n-set. This normalization scales large combinatorial values into probabilities, enabling clear graphical representation. The horizontal axis uses a logarithmic scale to depict sequence lengths of 100, 500, 1000, and 5000, while the vertical axis shows the corresponding partition probabilities. The sharp peaks in the distributions highlight the most probable partition configurations for each sequence length.
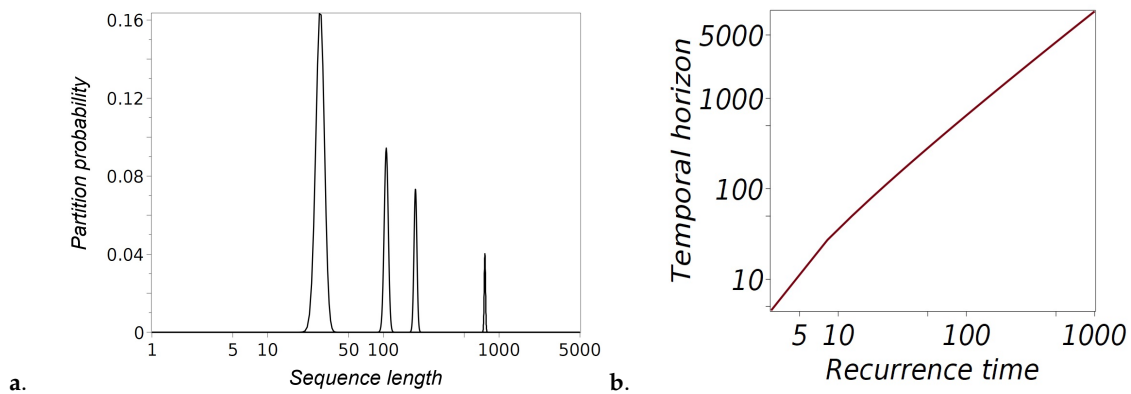

Figure 4. a). Partition probabilities for sequences of lengths 100, 500, 1000, and 5000. The graph displays the Stirling numbers of the second kind, normalized by the corresponding Bell numbers, representing the probability of partitioning a sequence of length n into m non-empty subsets. The sharp peaks, located at $m \simeq O(n/\ln n)$, indicate the maxima of the Stirling numbers, highlighting the most probable partition configurations. The horizontal axis uses a logarithmic scale to effectively capture the wide range of sequence lengths. **b**..) Temporal horizon n as a function of recurrence time $\mathcal{R}$ for a non-ergodic Rosencrantz coin

As n increases, certain partition configurations become increasingly dominant due to the combinatorial growth of the Stirling numbers. These configurations correspond to partitions where the block sizes are roughly balanced, maximizing the number of ways the sequence can be divided.

Taking the natural logarithm of $S(n, m)$, we get:

$$\ln S(n, m) \approx m \, \ln\left(\frac{n}{m}\right) + O(m). \tag{25}$$

By differentiating and setting $d(m \ln(n/m))/dm = 0$, we find the position of the maximum probability to be $m_{\max} = n/\ln n$. A detailed analysis conducted in [9] provides a more accurate estimate:

$$m_{\max} \approx \frac{n}{\ln n} + O\left(\frac{n(\ln \ln n)^{1/2}}{(\ln n)^{3/2}}\right). \tag{26}$$

Substituting this value of $m_{\max}$ back into the asymptotic expression for the logarithm of the maximum Stirling number, we have:

$$\ln S_{\max} \approx n \ln n - n - n \ln \ln n + \frac{2n \ln \ln n}{\ln n}. \tag{27}$$

The above estimate suggests that the system's temporal horizon n is most likely partitioned into blocks of size $\ln n$, representing the typical duration spent in a state before switching. The combinatorial advantage that favors specific block sizes reflects the intrinsic structure determined by the Stirling numbers of the second kind for large n. The estimates in (27) become increasingly accurate as n grows; however, for smaller n, correction terms can significantly influence the approximation (see Figure 4).

According to Rennie and Dobson [9], as n increases, the most likely partition of the sequence into $r = (1 \pm \epsilon)n / \ln n$ blocks approaches the configuration described by (27). The deviation between the maximum Stirling number and the actual Stirling number for this partition is given by:

$$\ln S_{\max} - \ln S(n, r) \approx -n\mathcal{H}, \ \mathcal{H} \approx \frac{(\ln \ln n)^2}{2 \ln n}. \tag{28}$$

The rate of convergence $\mathcal{H}$ to the most likely partition structure in (28) can be interpreted as an entropy-like function - a measure of block related entropy for the non-ergodic system. The value $\mathcal{H}$ varies only slightly over a wide range of n. Asymptotically, a sequence of length n will most likely be divided into $m_{\max} = n / \ln n$ blocks, each of length $\mathcal{R} \approx \ln n$, which is the characteristic recurrence time for the Rosencrantz coin remaining in a particular state. Consequently, the probability of observing a sequence consistent with the configuration of maximum likelihood equals to

$$\left[P_{\eta=1}(\ln n) \right]^{n / \ln n} \approx \exp\left(-\frac{2n \ln \ln n}{\ln n} \right) = \exp\left(-\frac{2n \ln \mathcal{R}}{\mathcal{R}} \right), \tag{29}$$

being compensated by the last correction term in the asymptotic approximation for the maximum Stirling number (27) – for blocks of the most probable size, the power-law decay over time is offset by the combinatorial growth of partitions. The term $2 \ln \mathcal{R} / \mathcal{R}$ in (29) corresponds to the entropy of a stationary coin, as discussed earlier in (2). It captures the balance between the combinatorial growth of the predominant partitions and the algebraic decay of block probabilities (17), which stabilizes the characteristic block size at $\ln n$. For sufficiently large n, the combinatorial increase in the number of predominant partitions compensates for the decay in block probabilities, ultimately dominating the system's behavior. Consequently, the structure of generated sequences resembles a deterministic schedule rather than a purely random process. The deterministic-like pattern emerges because the recurrence time $\mathcal{R}$ aligns with the residence and repetition times, making the sequence's structure predictable. This conclusion fully aligns with the case of a biased coin $T(p, q)$, where the probabilities p and q of repeating a state are close to one. Under these conditions, the coin's behavior becomes almost entirely predictable (see Figure 1).

For such a non-ergodic system, the relevant question is not whether the next state of the coin can be predicted, but whether the *temporal horizon* $\mathfrak{n}$ — the length of the sequence of flips that can be predicted based on the observed characteristic block size $\mathcal{R}$. Solving the equation $\mathcal{R} = \mathfrak{n} / \ln \mathfrak{n}$ gives:

$$\mathfrak{n}(\mathcal{R}) = -\mathcal{R} \ \mathrm{LambertW}_{-1}\left(-\mathcal{R}^{-1} \right), \tag{30}$$

where $\mathrm{LambertW}_{-1}(x)$ refers to the -1-th branch of the Lambert function, defined as the inverse of $ye^y = x$. This branch is real-valued for $-e^{-1} \leq x < 0$ and decreases monotonically towards negative infinity as $x \to 0$ ($\mathcal{R} \to \infty$) along the negative real axis, which forms the branch cut [10]. Figure 4.**b** illustrates the relationship between the recurrence time $\mathcal{R}$ and the temporal horizon $\mathfrak{n}$ as given by (30). The curve demonstrates that as the recurrence time, derived from the optimal block size determined by combinatorial analysis, increases, the temporal horizon grows super-linearly. This relationship

implies that the longer the system remains in a state before switching, the further into the future one can predict its behavior. Solving $\mathcal{R} = \mathfrak{n} / \ln \mathfrak{n}$ via the Lambert function (30) also provides a quantitative prediction of how long structured behavior will persist.

5. Discussion - Temporal Horizons and the Predictive Economy of Non-Ergodic Systems

In the study of non-ergodic systems like the Rosencrantz coin model, the temporal horizon refers to the characteristic duration over which predictable patterns and structures emerge within the system's evolution. Unlike ergodic systems, where time averages converge to ensemble averages, non-ergodic systems fail to visit all possible states uniformly. Consequently, the system's long-term behavior cannot be fully captured by stationary probability distributions.

This limitation shifts the analytical focus to understanding the durations between key events, such as state switches or recurrences. In such cases, combinatorial insights compensate for the absence of a stationary distribution, enabling meaningful temporal predictions in systems governed by deterministic-like patterns amidst randomness. The interplay between the recurrence time and the temporal horizon provides a robust framework for predicting the extent of structured behavior in non-ergodic processes.

Partitioning these sequences into blocks of persistent states reveals a deep connection to combinatorial structures, particularly the Stirling numbers of the second kind. These numbers exhibit a sharp maximum for partitions of size $\ln n$. For sufficiently large n, the combinatorial growth in the number of partitions compensates for the decay in block probabilities $\prod_{i=1}^{m} P_{\eta=1}(n_i)$ for the most likely partitions $n = n_1 + \ldots + n_m$. This balance between increasing block lengths and decreasing block probabilities results in a deterministic-like structure within an inherently stochastic system.

The temporal horizon, defined by the duration over which predictable patterns emerge, offers a new perspective for analyzing non-ergodic processes. The logarithmic recurrence time $\mathcal{R} \approx \ln n$ can be interpreted as the *prediction utility of time* within a conceptual *prediction economy*. The logarithmic utility of time implies that predictable patterns arise over increasingly longer horizons. Although the first derivative $\mathcal{R}'(n) > 0$, indicating that additional time enhances prediction, the second derivative $\mathcal{R}''(n) < 0$ reflects diminishing returns: each subsequent unit of time (or coin flip) adds less predictive value than the previous one. The concave nature of the logarithmic utility highlights this diminishing predictive benefit.

This logarithmic structure seamlessly leads to hyperbolic time discounting in prediction. Hyperbolic discounting prioritizes short-term forecasts with higher certainty over long-term, uncertain ones. The *Arrow-Pratt* measure of risk aversion [11,12] and the *Leland* measure of prudence [13] formalize this idea:

$$-\frac{\mathcal{R}''(n)}{\mathcal{R}'(n)} = \frac{1}{n}, \quad -\frac{\mathcal{R}'''(n)}{\mathcal{R}''(n)} = \frac{2}{n}. \tag{31}$$

These measures describe how knowledge of a block's length improves the ability to predict the temporal horizon. This relationship aligns with the hyperbolic time discounting model, a well-established framework in human [14] and animal [15] intertemporal choice, where immediate rewards are weighted more heavily than future gains. In the Rosencrantz coin model, this perspective underscores the preference for short-term, reliable predictions over long-term, uncertain ones.

6. Conclusion

In this paper, we explored the dynamics of both ergodic and non-ergodic systems through the lens of a coin-flipping model, specifically focusing on the Rosencrantz coin that can "freeze" in one of its states. We compared the behavior of a conventional biased coin, characterized by predictable and unpredictable entropy components, to the non-ergodic Rosencrantz coin, where the absence of a stationary distribution alters the predictability framework.

In the ergodic case, entropy decomposition into predictable and unpredictable components relies on characteristic times such as recurrence, residence, and repetition times. These measures provide insights into the structure and behavior of state sequences generated by stationary Markov chains. However, in the non-ergodic Rosencrantz coin model, where the system can persist in a state for extended periods, traditional entropy decomposition becomes inapplicable.

Our analysis demonstrated that the Rosencrantz coin's non-ergodic dynamics lead to logarithmically increasing block lengths in sequences of persistent states. By leveraging the Stirling numbers of the second kind, we identified the most probable partition size of $\ln n$ and showed that combinatorial growth in the number of partitions compensates for the algebraic decay of block probabilities. This balance results in deterministic-like structures emerging from fundamentally stochastic processes.

The concept of a temporal horizon $\mathfrak{n}$, defined by the recurrence time $\mathcal{R}$, provides a framework for predicting the duration of structured behavior in non-ergodic systems. The logarithmic utility of time implies diminishing predictive returns as the temporal horizon extends, aligning with hyperbolic time discounting models seen in human and animal decision-making.

In conclusion, the Rosencrantz coin model reveals that in non-ergodic systems, predictability is less about individual outcomes and more about understanding the length of sequences governed by persistent state blocks. This study bridges information theory, combinatorics, and stochastic processes, offering a deeper perspective on the dynamics and predictability of systems where conventional equilibrium frameworks fail.

References

1. Álvar Daza, Alexandre Wagemakers, Miguel A. F. Sanjuán; Multistability and unpredictability. *Physics Today* 1 November **2024**; 77 (11): 44–50. https://doi.org/10.1063/pt.kcxv.poxf

2. Volchenkov, D. Memories of the Future. Predictable and Unpredictable Information in Fractional Flipping a Biased Coin. *Entropy* **2019**, *21*, 807. https://doi.org/10.3390/e21080807

3. Stoppard, Tom. Rosencrantz and Guildenstern Are Dead. Page 23. *Grove Press*, **1967**.

4. James, R.G., Ellison, Ch.J. and Crutchfield, J.P., "Anatomy of a Bit: Information in a Time Series Measurement", *CHAOS* 21:3 (**2011**) 037109

5. Volchenkov, D., Infinite Ergodic Walks in Finite Connected Undirected Graphs. Entropy **2021**, 23 (2), 205. https://doi.org/10.3390/e23020205

6. Floriani, E., Volchenkov, D., Lima, R., "A System close to a threshold of instability", *J. of Physics A: Math. General* 36, 4771-4783 (**2003**).

7. Volchenkov, D., "*Survival under Uncertainty an Introduction to Probability Models of Social Structure and Evolution*", Springer Series: Understanding Complex Systems, 240 pages, ISBN 978-3-319-39419-0, Berlin / Heidelberg **2016**.

8. Graham, R.L., Knuth, D.E., Patashnik, O. *Concrete Mathematics*, Addison–Wesley, Reading MA. ISBN 0-201-14236-8, p. 244 **1988**.

9. Rennie, B.C., Dobson, A.J., On stirling numbers of the second kind, *Journal of Combinatorial Theory*, Volume **7**, Issue 2, Pages 116-121, **1969**.

10. Corless, R. M., Gonnet, G. H., Hare, D. E. G., Jeffrey, D. J., & Knuth, D. E. On the Lambert W Function. *Advances in Computational Mathematics*, 5(1), 329–359 (**1996**). DOI: 10.1007/BF02124750

11. Pratt, J.W., Risk Aversion in the Small and in the Large. *Econometrica*, 32(1/2), 122–136 (**1964**). DOI: 10.2307/1913738

12. Arrow, K.J., *Essays in the Theory of Risk-Bearing*. Markham Publishing Co. **1971**

13. Leland, H.E., Saving and uncertainty: the precautionary demand for saving. *Q. J. Econ.* 82, 465–473 (**1968**)

14. Laibson, D., "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics*, 112(2), 443–477 (**1997**). DOI: 10.1162/003355397555253

15. Mazur, J.E., "An Adjusting Procedure for Studying Delayed Reinforcement." In *The Effect of Delay and of Intervening Events on Reinforcement Value* (Eds. Michael L. Commons, James E. Mazur, John A. Nevin, and Howard Rachlin), Series: Quantitative Analyses of Behavior, Volume 5, Lawrence Erlbaum Associates (pp. 55-73)(**1987**).