

Article

# Memories of the Future. Predictable and Unpredictable Information in Fractional Flipping a Biased Coin

Dimitri Volchenkov 

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

Received: 1 August 2019; Accepted: 15 August 2019; Published: 18 August 2019

**Abstract:** Some uncertainty about flipping a biased coin can be resolved from the sequence of coin sides shown already. We report the exact amounts of predictable and unpredictable information in flipping a biased coin. Fractional coin flipping does not reflect any physical process, being defined as a binomial power series of the transition matrix for “integer” flipping. Due to strong coupling between the tossing outcomes at different times, the side repeating probabilities assumed to be independent for “integer” flipping get entangled with one another for fractional flipping. The predictable and unpredictable information components vary smoothly with the fractional order parameter. The destructive interference between two incompatible hypotheses about the flipping outcome culminates in a fair coin, which stays fair also for fractional flipping.

**Keywords:** fractional flipping a biased coin; probability entanglement; destructive interference of information

---

*“A weaker man might be moved to re-examine his faith, if in nothing else at least in the law of probability.”*

Tom Stoppard, “Rosencrantz and Guildenstern Are Dead”, Act 1.

## 1. Introduction

The vanishing probability of winning in a long enough sequence of coin flips features in the opening scene of Tom Stoppard’s play “*Rosencrantz and Guildenstern Are Dead*”, where the protagonists are betting on coin flips. Rosencrantz, who bets on heads each time, has won 92 flips in a row, leading Guildenstern to suggest that they are within the range of supernatural forces. Furthermore, he was actually right, as the king had already sent for them [1].

Although coin-tossing experiments are ubiquitous in courses on elementary probability theory, and coin tossing is regarded as a prototypical random phenomenon of unpredictable outcome, the exact amounts of predictable and unpredictable information related to flipping a biased coin was not discussed in the literature. The discussion on whether the outcome of naturally tossed coins is truly random [2], or if it can be manipulated (and therefore predicted) [3,4] has been around perhaps for as long as coins existed. It is worth mentioning that tossing of a real coin obeys the physical laws and is inherently a deterministic process, with the outcome that, formally speaking, might be determined if the initial state of the coin is known [5].

All in all, the toss of a coin has been a method used to determine random outcomes for centuries [4]. The practice of flipping a coin was ubiquitous for taking decisions under uncertainty, as a chance outcome is often interpreted as the expression of divine will [1]. Individuals who are told by the coin toss to make an important change are reported much more likely to make a change and are happier six months later than those who were told by the coin to maintain the status quo in their lives [6].

If the coin is not fair, the outcome of future flipping can be either (i.) anticipated intuitively by observing the whole sequence of sides shown in the past in search for the possible patterns and repetitions, or (ii.) guessed instantly from the side just showed up. In our brain, the stored routines and patterns making up our experience are managed by the basal *ganglia*, and *insula*, highly sensitive to any change, takes care of our present awareness and might feature the guess on the coin toss outcome [7]. Trusting our gut, we unconsciously look for patterns in sequences of shown sides, *a priori* perceiving any coin as unfair.

In the present paper, we propose the information theoretic study of the most general models for “integer” and fractional flipping a biased coin. We show that these stochastic models are singular (along with many other well-known stochastic models), and therefore their parameters—the side repeating probabilities—cannot be inferred from assessing frequencies of shown sides (see Sections 2 and 4). In Section 3, we demonstrate that some uncertainty about the coin flipping outcome can nevertheless be resolved from the presently shown side and the sequence of sides occurred in the past, so that the actual level of uncertainty attributed to flipping a biased coin can be lower than assessed by entropy. We suggest that the entropy function can therefore be decomposed into the predictable and unpredictable information components (Section 3). Interestingly, the efficacy of the side forecasting strategies (i.) and (ii.) mentioned above can be quantified by the distinct information theoretic quantities—the excess entropy and conditional mutual information, respectively (Section 3). The decomposition of entropy into the predictable and unpredictable information components is justified rigorously at the end of Section 3.

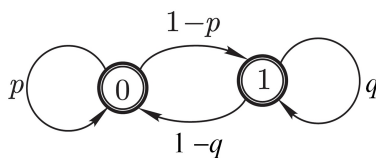
In Section 4, we introduce a backward-shift Markov chain transition matrix generalizing the standard “integer” coin flipping model for fractional order flipping. Namely, the *fractional order Markov chain* is defined as a convergent infinite binomial series in the “integer”-order transition matrix that assumes strong coupling between the chain states (coin tossing outcomes) at different times. The fractional backward shift transition operator *does not reflect any physical process*.

On the one hand, our fractional coin-tossing model is intrinsically similar to the *fractional random walks* introduced recently in [8–12] in the context of Markovian processes defined on networks. In contrast to the normal random walk where the walker can reach in one time-step only immediately connected nodes, the fractional random walker governed by a fractional Laplacian operator is allowed to reach any node in one time-step dynamically introducing a small-world property to the network. On the other hand, our fractional order Markov chain is closely related to the *Autoregressive Fractional Integral Moving Average* (ARFIMA) models [13–15], a fractional order signal processing technique generalizing the conventional integer order models—autoregressive integral moving average (ARIMA) and autoregressive moving average (ARMA) model [16]. In the context of time series analysis, the proposed fractional coin-flipping model resolves the fractional order time-backward outcomes (i.e., *memories* [17–21]) as the moving averages over all *future* states of the chain—that explains the title of our paper. We also show that the side repeating probabilities considered independent of each other in the standard, “integer” coin-tossing model appear to be entangled with one another as a result of strong coupling between the future states in fractional flipping. Finally, we study the evolution of the predictable and unpredictable information components of entropy in the model of fractional flipping a biased coin (Section 5). We conclude in the last section.

## 2. The Model of a Biased Coin

A biased coin prefers one side over another. If this preference is stationary, and the coin tosses are independent of each other, we describe coin flipping by a Markov chain defined by the stochastic transition matrix, viz.,

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



in which the states, ‘heads’ (“0”) or ‘tails’ (“1”), repeat themselves with the probabilities  $0 \leq p \leq 1$  and  $0 \leq q \leq 1$ , respectively. The Markov chain Equation (1) generates the stationary sequences of states, viz.,  $0, 0, 0, \dots$  when  $p = 1$ , or  $1, 1, 1, \dots$  when  $q = 1$ , or  $0, 1, 0, 1 \dots$  when  $q = p = 0$ , but describes flipping a fair coin if  $q = p = \frac{1}{2}$ .

For a symmetric chain,  $q = p$ , the relative frequencies (or densities) of ‘head’ and ‘tail’,

$$\pi_1(p, q) = \frac{1 - q}{2 - p - q} \quad \text{and} \quad \pi_2(p, q) = \frac{1 - p}{2 - p - q}, \tag{2}$$

are equal each other,  $\pi_1(p, p) = \pi_2(p, p) = 1/2$ , and therefore the entropy function, expressing the amount of uncertainty about the coin flip outcome, viz.,

$$H(p, q) = - \sum_{k=1}^2 \pi_k(p, q) \cdot \log_2 \pi_k(p, q), \tag{3}$$

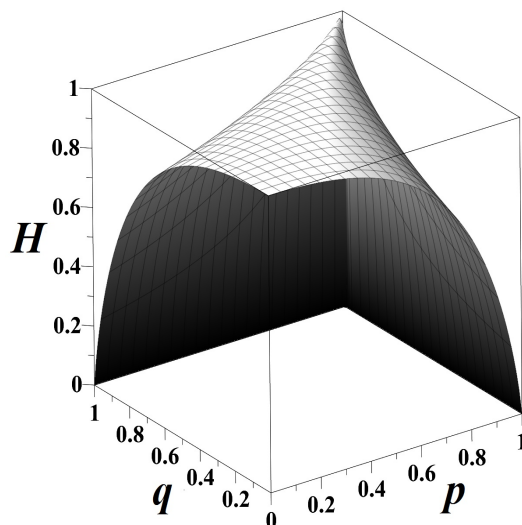
attains the maximum value,  $H(p, p) = 1$  bit, uniformly for all  $0 < p < 1$ . On the contrary, flipping the coin when  $p = 1$  (or  $q = 1$ ) generates the stationary sequences of no uncertainty,  $H(p, q) = 0$  (see Figure 1). In Equation (3) and throughout the paper, we use the following conventions reasonable by a limit argument:  $0 \cdot \log 0 = \log 0^0 = \log 1 = 0$ . The information difference between the amounts of uncertainty on a smooth statistical manifold parametrized by the probabilities  $p$  and  $q$  is calculated using the Fisher information matrix (FIM) [22–24], viz.,

$$g_{p,q} = \sum_{k=1}^2 \pi_k(p, q) \cdot \frac{\partial}{\partial p} \log_2 \pi_k(p, q) \cdot \frac{\partial}{\partial q} \log_2 \pi_k(p, q). \tag{4}$$

However, since  $H(p, p) = 1$  bit, for  $0 < p = q < 1$ , the FIM,

$$\mathbf{g} = \frac{1}{(\ln 2)^2 (2 - p - q)^2} \begin{pmatrix} \frac{1-q}{1-p} & -1 \\ -1 & \frac{1-p}{1-q} \end{pmatrix}, \tag{5}$$

is degenerate (with eigenvalues  $\lambda_1 = 0, \lambda_2 = (p^2 + q^2 + 2(1 - p - q)) / \ln^2(2)(1 - p)(1 - q)(2 - p - q)^2$ ), and therefore the biased coin model Equation (1) is singular, along with many other stochastic models, such as Bayesian networks, neural networks, hidden Markov models, stochastic context-free grammars, and Boltzmann machines [25]. The singular FIM (4) assumes that the parameters of the model,  $p$  and  $q$ , cannot be inferred from assessing relative frequencies of sides in sequences generated by the Markov chain Equation (1).



**Figure 1.** The value of entropy Equation (3) attains maximum (of 1 bit) for the symmetric chain,  $q = p$ , but is zero for the stationary sequences,  $p = 1$ , or  $q = 1$ .

### 3. Predictable and Unpredictable Information in the Model of Tossing a Biased Coin

Although coin tossing is traditionally regarded as a prototypical random experiment of unpredictable outcome, some amount of uncertainty in the model Equation (1) can be dispelled before tossing a coin. Namely, we can consider the entropy function Equation (3) as a sum of the *predictable* and *unpredictable* information components,

$$H(p, q) = P(p, q) + U(p, q), \tag{6}$$

where the predictable part  $P(p, q)$  estimates the amount of *apparent* uncertainty about the future flipping outcome that might be resolved from the sequence of sides shown already, and  $U(p, q)$  estimates the amount of *true* uncertainty that cannot be inferred neither from the past, nor from present outcomes anyway. It is reasonable to assume that both functions,  $P$  and  $U$ , in Equation (6) should have the same form as the entropy function in Equation (3), viz.,

$$P = - \sum_{k=1}^2 \pi_k \cdot \log_2 \varphi_k, \quad U = - \sum_{k=1}^2 \pi_k \cdot \log_2 \psi_k, \quad \varphi_k \psi_k = \pi_k. \tag{7}$$

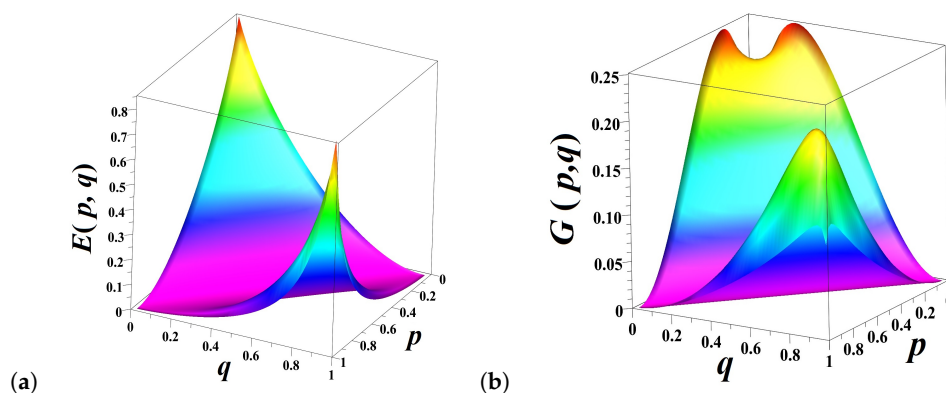
Furthermore, as the more frequent the side, the higher the forecast accuracy, we assume that the partition function  $\varphi_k$  featuring the predicting potential in already shown sequences for forecasting the side  $k$  is obviously proportional to the relative frequency of that side,  $\varphi_k \propto \pi_k$ . Denoting the relevant proportionality coefficient as  $\sigma_k$  in  $\varphi_k = \pi_k \sigma_k$ , we obtain  $\psi_k = \sigma_k^{-1} = \frac{\pi_k}{\varphi_k}$ . Given the already shown sequence of coin sides  $\overleftarrow{S}_t = S_{t-1}, S_{t-2}, S_{t-3}, \dots$ , the average amount of uncertainty about the flipping outcome is assessed by the entropy rate [24] of the Markov chain Equation (1), viz

$$H(S_t | \overleftarrow{S}_t) = H(S_t | S_{t-1}) = - \sum_{k=1}^2 \pi_k \sum_{r=1}^2 T_{kr} \log_2 T_{kr}, \quad \text{where } \mathbf{T} = \begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix}, \tag{8}$$

and therefore, the *excess entropy* [25–27], quantifying the apparent uncertainty of the flipping outcome that can be resolved by discovering the repetition, rhythm, and patterns over the whole (infinite) sequence of sides shown in the past,  $\overleftarrow{S}_t$ , equals

$$E(p, q) \equiv H(S_t) - H(S_t|S_{t-1}) = - \sum_{k=1}^2 \pi_k \cdot \left( \log_2 \pi_k - \sum_{r=1}^2 T_{kr} \log_2 T_{kr} \right). \tag{9}$$

The excess entropy  $E(p, q)$  attains the maximum value of 1 bit over the stationary sequences but equals zero for  $q = 1 - p$  (see Figure 2a).



**Figure 2.** (a)  $E(p, q)$ , the apparent uncertainty of the flipping outcome that can be resolved by discovering possible patterns and repetitions in the infinite sequence of shown sides; (b)  $G(p, q)$ , the mutual information between the present state and the future state conditioned on the past measuring the efficacy of forecast of the coin toss outcome from the present state alone.

Moreover, the next flipping outcome can be guessed from the present state alone, and the level of accuracy of such a guess can be assessed by the *mutual information* between the present state and the future state conditioned on the past state  $I(S_t; S_{t+1}|S_{t-1})$  [25,28], viz.,

$$\begin{aligned} G(p, q) &\equiv I(S_t; S_{t+1}|S_{t-1}) \\ &= H(S_{t+1}|S_{t-1}) - H(S_t|S_{t-1}) = \sum_{k=1}^2 \pi_k \sum_{r=1}^2 (T_{kr} \log_2 T_{kr} - T_{kr}^2 \log_2 T_{kr}^2). \end{aligned} \tag{10}$$

The mutual information (10) is a component of the entropy rate (9) growing as  $p, q \geq 0$  and  $p, q \lesssim 1$ . For  $q = 1 - p$ , the rise of *destructive interference* between two incompatible hypotheses on

- (i) alternating the present side at the next tossing (if  $p, q > 0$ ), or
- (ii) repeating the present side at the next tossing (when  $p, q < 1$ )

causes the attenuation and cancellation of mutual information (10) (Figure 2b).

By summing (9) and (10), we obtain the amounts of predictable and unpredictable information, respectively:

$$\begin{aligned} P(p, q) &= E(p, q) + G(p, q) = H(S_t) - H(S_t|S_{t-1}) + I(S_t; S_{t+1}|S_{t-1}), \\ U(p, q) &= H(p, q) - P(p, q) = H(S_t|S_{t-1}) - I(S_t; S_{t+1}|S_{t-1}) \\ &= H(S_t|S_{t+1}; S_{t-1}), \end{aligned} \tag{11}$$

where  $H(S_t|S_{t+1}; S_{t-1})$  is the entropy of the present state conditional on the future and past states of the chain. The latter conditional entropy is naturally expressed via the entropy of the future state

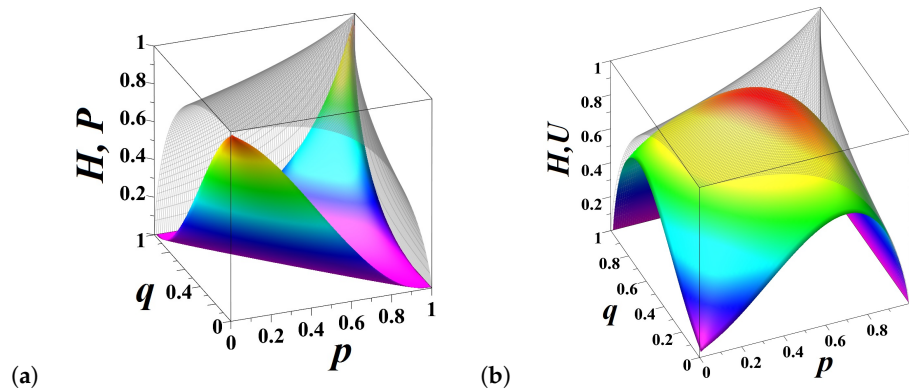
conditional on the present  $H(S_{t+1}|S_t)$ , the entropy of the present state conditional on the past  $H(S_t|S_{t-1})$ , and the entropy of the future state conditional on the past  $H(S_{t+1}|S_{t-1})$  as following:  $H(S_t|S_{t+1};S_{t-1}) = H(S_{t+1}|S_t) + H(S_t|S_{t-1}) - H(S_{t+1}|S_{t-1})$ . The accuracy of the obtained information decomposition of entropy,

$$H(p, q) = P(p, q) + U(p, q) = E(p, q) + G(p, q) + U(p, q), \tag{12}$$

is demonstrated immediately by the following computation involving the conditional entropies:

$$\begin{aligned} H(S_t) &= H(S_t) - H(S_{t+1}|S_t) + H(S_{t+1}|S_t) \\ &= (H(S_t) - H(S_{t+1}|S_t)) + H(S_{t+1}|S_t) + \{H(S_t|S_{t-1}) - H(S_t|S_{t-1})\} \\ &\quad + \{H(S_{t+1}|S_{t-1}) - H(S_{t+1}|S_{t-1})\} \\ &= \underbrace{(H(S_t) - H(S_{t+1}|S_t))}_{E(p,q)} + \underbrace{(H(S_{t+1}|S_{t-1}) - H(S_t|S_{t-1}))}_{G(p,q)} \\ &\quad + \underbrace{(H(S_{t+1}|S_t) + H(S_t|S_{t-1}) - H(S_{t+1}|S_{t-1}))}_{U(p,q)}. \end{aligned} \tag{13}$$

The predictable information component  $P(p, q)$  amounts to  $H(p, q)$  over the stationary sequences but disappears for  $q = 1 - p$  (Figure 3a). On the contrary, the share of unpredictable information  $U(p, q)$  attains the maximum value  $U(p, 1 - p) = H(p, 1 - p)$ , for  $q = 1 - p$  (Figure 3b).



**Figure 3.** (a) The entropy  $H(p, q)$  (transparent) and predictable information  $P(p, q)$  (hue colored) in the model of a biased coin for the different values of  $p$  and  $q$ ; (b) The entropy  $H(p, q)$  (transparent) and unpredictable information  $U(p, q)$  (hue colored) for the different values of  $p$  and  $q$ .

#### 4. The Model of Fractional Flipping a Biased Coin

In our work, we define the model of fractional flipping a biased coin using the fractional differencing of non-integer order [29,30] for the discrete time stochastic processes [31–33]. The Grunwald-Letnikov fractional difference  $\Delta_\tau^\alpha \equiv (1 - T)^\alpha$  of order  $\alpha$  with the unit step  $\tau$ , and the time lag operator  $T$  is defined [18,29,30,34–36] by

$$\Delta_\tau^\alpha x(t) \equiv (1 - T_\tau)^\alpha x(t) = \sum_{m=0}^{\infty} (-1)^m \cdot \binom{\alpha}{m} \cdot x(t - m \cdot \tau) \tag{14}$$

where  $T_\tau x(t) = x(t - \tau)$  is fixed  $\tau$ -delay, and  $\binom{\alpha}{m}$  is the binomial coefficient that can be written for integer or non-integer order  $\alpha$  using the Gamma function, viz.,

$$\binom{\alpha}{m} \equiv (-1)^{m-1} \frac{\alpha \cdot \Gamma(m - \alpha)}{\Gamma(1 - \alpha)\Gamma(m + 1)}. \tag{15}$$

It should be noted that for a Markov chain defined by Equation (1), the Grunwald-Letnikov fractional difference of a non-integer order  $1 - \varepsilon$  takes form of the following infinite series of binomial type, viz.,

$$(1 - T)^{1-\varepsilon} = \sum_{k=0}^{\infty} \frac{\Gamma(k - 1 + \varepsilon)}{\Gamma(k + 1)\Gamma(-1 + \varepsilon)} T^k = 1 + \sum_{k=1}^{\infty} \frac{\Gamma(k - 1 + \varepsilon)}{\Gamma(k + 1)\Gamma(-1 + \varepsilon)} T^k \equiv 1 - T_{1-\varepsilon} \tag{16}$$

that converges absolutely, for  $0 < \varepsilon < 1$ . In Equation (16), we have used a formal structural similarity between the fractional order difference operator and the power series of binomial type in order to introduce a *fractional backward-shift transition operator*  $T_{1-\varepsilon}$  for any fractional order  $0 < 1 - \varepsilon < 1$  as a convergent infinite power series of the transition matrix Equation (1), viz.,

$$\begin{aligned} T_{1-\varepsilon}(p, q) &\equiv - \sum_{k=1}^{\infty} \frac{\Gamma(k-1+\varepsilon)}{\Gamma(k+1)\Gamma(-1+\varepsilon)} T^k(p, q) \\ &= \begin{pmatrix} 1 - \frac{1-p}{(2-p-q)^\varepsilon} & \frac{1-p}{(2-p-q)^\varepsilon} \\ \frac{1-q}{(2-p-q)^\varepsilon} & 1 - \frac{1-q}{(2-p-q)^\varepsilon} \end{pmatrix} \equiv \begin{pmatrix} p_\varepsilon & 1 - p_\varepsilon \\ 1 - q_\varepsilon & q_\varepsilon \end{pmatrix}. \end{aligned} \tag{17}$$

The backward-shift fractional transition matrix defined by Equation (17) is a stochastic matrix preserving the structure of the initial Markov chain Equation (1), for any  $0 < 1 - \varepsilon < 1$ . Since the power series of binomial type in Equation (17) is convergent and summable for any value  $0 < 1 - \varepsilon < 1$ , we have also introduced in Equation (17) the *fractional probabilities*,  $p_\varepsilon$  and  $q_\varepsilon$ , as the corresponding elements of the fractional transition matrix. The fractional transition operator Equation (17) describes fractional flipping a biased coin for  $0 < \varepsilon < 1$  as a *moving average* over the probabilities of all future outcomes of the Markov chain Equation (1) described by integer powers  $T^k$ ,  $k = 1, \dots, \infty$ . The fractional Markov chain Equation (17) is also similar to the *fractional random walks* introduced recently in [8–12]. In these research efforts, the fractional Laplace operator describing anomalous transportation in connected networks and the fractional degree of a node are related to integer powers of the network adjacency matrix  $\mathbf{A}^m$  for  $m = 1, \dots, \infty$  for which the element  $(\mathbf{A}^m)_{ij}$  is the total number of all possible trajectories connecting nodes  $i$  and  $j$  by paths of length  $m$ . The fractional characteristics of the graph not only incorporate information related to the number of nearest neighbors of a node, but also include information of all far away neighbors of the node in the network, allowing for long-range transitions between the nodes and featuring anomalous diffusion [10].

In the proposed fractional Markov chain Equation (17), the kernel function (which can be called *memory function* following [19–21,37]) establishes strong coupling between the outcome of fractional coin flipping for the fractional order parameter  $\varepsilon$  and the probabilities of all future outcomes of the “integer”-order Markov chain Equation (1). It is worth mentioning that the fractional transition probabilities in Equation (17) equal those in the “integer”-order flipping model Equation (1) as  $\varepsilon \rightarrow 0$ , viz.,

$$\lim_{\varepsilon \rightarrow 0} p_\varepsilon = p, \quad \lim_{\varepsilon \rightarrow 0} q_\varepsilon = q, \tag{18}$$

but coincide with the densities Equation (2) of the ‘head’ and ‘tail’ states, as  $\varepsilon \rightarrow 1$ , viz.,

$$\lim_{\varepsilon \rightarrow 1} p_\varepsilon = \frac{1 - q}{2 - p - q} = \pi_1, \quad \text{and} \quad \lim_{\varepsilon \rightarrow 1} q_\varepsilon = \frac{1 - p}{2 - p - q} = \pi_2. \tag{19}$$

Thus, the minimal value of the fractional order parameter ( $\varepsilon = 0$ ) in the model Equation (17) may be attributed to the “integer”-order coin flipping when no information about the future flipping outcomes is available, i.e., the very moment of time when the present side of coin is revealed. Furthermore, the maximum value of the fractional order parameter ( $\varepsilon = 1$ ) corresponds to the maximum available information about all future coin-tossing outcomes. Averaging over all future states of the chain as  $\varepsilon = 1$  recovers the density of states Equation (19) of the Markov chain Equation (1) precisely as expected.

The transformation Equation (17) defines the  $(p_\varepsilon, q_\varepsilon)$ —flow of fractional probabilities over the fractional order parameter  $\varepsilon$  as shown in Figure 4a. In fractional flipping,  $0 < \varepsilon \leq 1$ , the state repetition probabilities  $p_\varepsilon$  and  $q_\varepsilon$  get entangled with one another due to the normalization factor  $(2 - p - q)^{-\varepsilon}$  in Equation (17). For the integer order coin flipping model  $\varepsilon = 0$ , the state repetition probabilities  $p_0$  and  $q_0$  are independent of each other (as shown by flow arrows on to top face of the cube in Figure 4a) but they are linearly dependent,  $\pi_1 = p_1 = 1 - q_1 = 1 - \pi_2$ , as  $\varepsilon = 1$  (see the bottom face of the cube in Figure 4a).

The degree of entanglement as a function of the fractional order parameter  $\varepsilon$  can be assessed by the expected divergence between the fractional model probabilities,  $p_\varepsilon$  and  $q_\varepsilon$ , in the models Equation (1) and Equation (17), viz.,

$$\begin{aligned} \text{Ent}(\varepsilon) &= \iint_0^1 dpdq \left( \pi_1 \log_2 \frac{p}{p_\varepsilon} + \pi_2 \log_2 \frac{q}{q_\varepsilon} \right) = 2 \iint_0^1 dpdq \left( \pi_1 \log_2 \frac{p}{p_\varepsilon} \right) \\ &= 2 \iint_0^1 dpdq \left( \pi_2 \log_2 \frac{q}{q_\varepsilon} \right). \end{aligned} \tag{20}$$

The integrand in Equation (20) turns to zero when the probabilities are independent of one another (as  $\varepsilon = 0$ ) but equals the doubled *Kullback–Leibler divergence* (relative entropy) [24] between  $p$  and  $\pi_1$  ( $q$  and  $\pi_2$ ) as  $\varepsilon = 1$  (due to the obvious  $p \leftrightarrow q$  symmetry of expressions). The degree of probability entanglement defined by Equation (20) attains the maximum value at  $\varepsilon = 0.855$  (Figure 4b).

Since the vector of ‘head’ and ‘tail’ densities Equation (2) is an eigenvector for all integer powers  $T^k$ , it is also an eigenvector for the fractional transition operator  $T_\varepsilon(p, q)$ , for any value of the fractional order parameter  $\varepsilon$ . Therefore, the fractional dynamics of transition probabilities does not change the densities of states in the Markov chain, so that the entropy function Equation (3) is an invariant of fractional dynamics in the model Equation (17) (Figure 4a). The Fisher information matrix Equation (4) is redefined for the probabilities  $p_\varepsilon, q_\varepsilon$ , viz.,

$$g_{p_\varepsilon, q_\varepsilon} = \sum_{k=1}^2 \pi_k \cdot \frac{\partial}{\partial p_\varepsilon} \log_2 \pi_k \cdot \frac{\partial}{\partial q_\varepsilon} \log_2 \pi_k, \tag{21}$$

which is also degenerate because the symmetry  $p \leftrightarrow q$  is preserved in all the expressions for all values  $0 < \varepsilon < 1$ . The nontrivial eigenvalue of the FIM Equation (21) turns to zero as well, for the stationary sequences with  $p = q = 1$ . The fractional flipping a biased coin model is singular, as well as the integer time flipping model Equation (1).



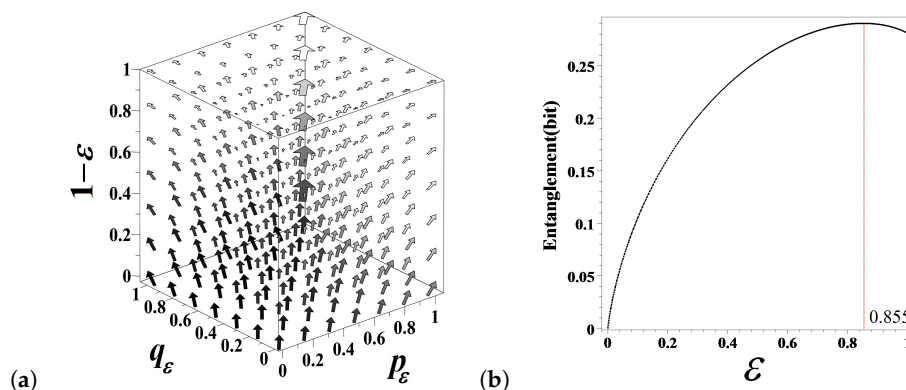


Figure 4. (a) The  $(p_\epsilon, q_\epsilon)$ —flow of the model Equation (17) over fractional order  $1 - \epsilon$ ; (b) The entanglement between the probabilities  $p_\epsilon$  and  $q_\epsilon$  defined by Equation (20) attains the maximum value at  $\epsilon = 0.855$ .

### 5. Evolution of Predictable and Unpredictable Information Components over the Fractional Order Parameter

The predictable and unpredictable information components defined by Equations ((9)–(11)) can be calculated for the fractional transition matrix Equation (17), for any value of the fractional order parameter  $0 < \epsilon \leq 1$ . In the present section, without loss of generality, we discuss the case of symmetric chain,  $q = p$ . For a symmetric chain, the densities of both states are equal,  $\pi = [\frac{1}{2}, \frac{1}{2}]$ , so that  $H(p, p) \equiv H(p) = -\log_2 \frac{1}{2} = 1$  bit, uniformly for all  $0 < p < 1$  (Figure 5a). The excess entropy Equation (9) quantifying predictable information encoded in the historical sequence of showed sides for a symmetric chain reads as follows [38]:

$$E(p, p) \equiv E(p) = 1 - H(S_t|S_{t-1}) = -p \cdot \log_2 p - (1 - p) \log_2 (1 - p). \tag{22}$$

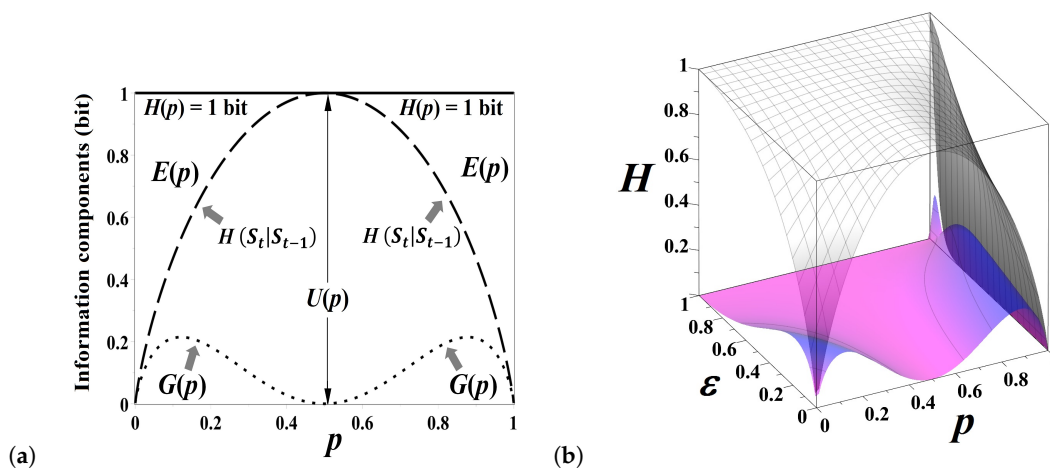
Forecasting the future state through discovering patterns in sequences of shown sides Equation (22) loses any predictive power when the coin is fair,  $p = \frac{1}{2}$ , but  $E(p) = 1$  bit when the series is stationary (i.e.,  $p = 0$ , or  $p = 1$ ). The mutual information Equation (10) measuring the reliability of the guess about the future state provided the present state is known [38],

$$G(p) = p \cdot \log_2 p + (1 - p) \cdot \log_2 (1 - p) - 2p(1 - p) \cdot \log_2 2p(1 - p) - (p^2 + (1 - p)^2) \cdot \log_2 (p^2 + (1 - p)^2), \tag{23}$$

increases as  $p \gtrsim 0$  ( $p \lesssim 1$ ) attaining maximum at  $p \approx 0.121$  ( $p \approx 0.879$ ). The effect of destructive interference between two incompatible hypotheses about alternating the current state ( $p \gtrsim 0$ ) and repeating the current state ( $p \lesssim 1$ ) culminates in fading this information component when the coin is fair,  $p = 1/2$  (Figure 5a). The difference between the entropy rate  $H(S_t|S_{t-1})$  and the mutual information  $G(p)$  may be viewed as the “degree of fairness” of the coin that attains maximum ( $U(p) = 1$  bit) for the fair coin  $p = 1/2$  (see Figure 5a).

The entropy decomposition presented in Figure 5a for “integer”-order flipping ( $\epsilon = 0$ ) evolves over the fractional order parameter,  $0 < \epsilon \leq 1$  as shown in Figure 5b: the decomposition of entropy shown in Figure 5a corresponds to the outer face of the three dimensional Figure 5b. When  $p = 1$ , the sequence of coin sides shown in integer flipping is stationary, so that there is no uncertainty about the coin tossing outcome. However, the amount of uncertainty for  $p = 1$  grows to 1 bit, for fractional flipping as  $\epsilon \rightarrow 1$ . When  $\epsilon = 1$ , the repetition probability of coin sides equals its relative frequency,

$p_1 = \pi_1 = 1/2$ , and therefore uncertainty about the future state of the chain cannot be reduced anyway,  $H(1/2) = U(1/2) = 1$  bit. Interestingly, there is some gain of predictable information component  $G(p)$  for  $p = 1$  as  $\varepsilon \lesssim 1$  (see Figure 5b). The information component  $G(p)$  quantifies the goodness of guess of the flipping outcome from the present state of the chain, so that the gain observed in Figure 5b might be interpreted as the reduction of uncertainty in a stationary sequence due to the choice of the present state, “0” or “1”. Despite the dramatic demise of unpredictable information for fractional flipping as  $\varepsilon \rightarrow 0$ , the fair coin ( $p = 1/2$ ) always stays fair.



**Figure 5.** (a) The decomposition of entropy into information components for a symmetric unfair coin,  $q = p$ , for “integer”-order coin flipping. The symmetric coin is fair when  $p = 1/2$ : the amount of uncertainty of the fair coin tossing cannot be reduced anyway, as the amount of unpredictable information equals  $U(1/2) = H(p) = 1$  bit; (b) The information components for fractional order flipping a biased coin. The decomposition of entropy at integer time shown in Figure 5a corresponds to the outer face of the three dimensional diagram in Figure 5b ( $\varepsilon = 0$ ).

## 6. Conclusions

A simple Markov chain generating binary sequences provides us with an analytically computable and telling example for studying conditional information quantities that quantify predictable and unpredictable information about the future states of the chain. The destructive interference between the mutually incompatible hypotheses about the forthcoming state of the chain results in damping of predictable information for a completely unpredictable, fair coin.

We have introduced and studied the fractional order generalization of the Markov chain defined as a convergent binomial series in the “integer”-order transition matrix. The proposed concept of fractional order Markov chain (fractional coin flipping) is similar to fractional random walks [8–11] and to the fractional order signal processing techniques generalizing the conventional integer order models—autoregressive integral moving average [14,15]. The backward-shift fractional order transition operator averages over all future states of the “integer”-order Markov chain exhibiting properties of long-time dependence, including the entanglement of state repetition probabilities assumed to be the independent parameters of the “integer”-order model.

**Funding:** This research received no external funding.

**Acknowledgments:** We are grateful to the reviewers for their efforts in carefully reading our manuscript and giving us very useful and concrete comments that helped us to largely improve the paper. We also thank Texas Tech University for the administrative and technical support.

**Conflicts of Interest:** The author declares no conflict of interest.

## References

1. Volchenkov, D. *Survival under Uncertainty. An Introduction to Probability Models of Social Structure and Evolution*; Springer Series “Understanding Complex Systems”; Springer International Publishing: Cham, Switzerland, 2016.
2. Keller, J.B. The probability of heads. *Am. Math. Mon.* **1986**, *93*, 191–197. [[CrossRef](#)]
3. Diaconis, P.; Holmes, S.; Montgomery, R. Dynamical bias in the coin toss. *SIAM Rev.* **2007**, *49*, 211. [[CrossRef](#)]
4. Clark, M.P.; Westerberg, B.D. How random is the toss of a coin? *CMAJ* **2009**, *181*, E306–E308. [[CrossRef](#)] [[PubMed](#)]
5. Stefan, R.C.; Cheche, T.O. Coin Toss Modeling. *arXiv* **2016**, arXiv:1612.06705.
6. Levitt, S.D. *Heads or Tails: The Impact of a Coin Toss on Major Life Decisions and Subsequent Happiness*; NBER Working Paper No. 22487; National Bureau of Economic Research: Cambridge, MA, USA, 2016.
7. Fabritius, F.; Hagemann, H.W. *The Leading Brain: Neuroscience Hacks to Work Smarter, Better, Happier*; Penguin: Westminister, UK, 2017.
8. Riascos, A.P.; Mateos, J.L. Long-range navigation on complex networks using Lévy random walks. *Phys. Rev. E* **2012**, *86*, 056110. [[CrossRef](#)] [[PubMed](#)]
9. Riascos, A.P.; Mateos, J.L. Fractional dynamics on networks: Emergence of anomalous diffusion and Lévy flights. *Phys. Rev. E* **2014**, *90*, 032809. [[CrossRef](#)]
10. Riascos, A.P.; Mateos, J.L. Fractional diffusion on circulant networks: Emergence of a dynamical small world. *J. Stat. Mech.* **2015**, P07015. [[CrossRef](#)]
11. Riascos, A.P.; Mateos, J.L. Fractional quantum mechanics on networks: Long-range dynamics and quantum transport. *Phys. Rev. E* **2015**, *92*, 052814. [[CrossRef](#)]
12. Michelitsch, T.M.; Collet, B.A.; Riascos, A.P.; Nowakowski, A.F.; Nicolleau, F.C.G.A. Fractional random walk lattice dynamics. *J. Phys. A Math. Theor.* **2017**, *50*, 055003. [[CrossRef](#)]
13. Box, G.E.; Hunter, W.G.; Hunter, J.S. *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*; John Wiley & Sons: Hoboken, NJ, USA, 1978; Volume 1.
14. Box, G.E.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*; John Wiley & Sons: Hoboken, NJ, USA, 2015.
15. Liu, K.; Chen, Y.; Zhang, X. An Evaluation of ARFIMA (Autoregressive Fractional Integral Moving Average) Programs. *Axioms* **2017**, *6*, 16. [[CrossRef](#)]
16. Sheng, H.; Chen, Y.; Qiu, T. *Fractional Processes and Fractional-Order Signal Processing: Techniques and Applications*; Springer Science & Business Media: New York, NY, USA, 2011.
17. Podlubny, I. Geometric and physical interpretation of fractional integration and fractional differentiation. *Fract. Calc. Appl. Anal.* **2002**, *5*, 367–386.
18. Podlubny, I. *Fractional Differential Equations*; Academic Press: San Diego, CA, USA, 1998; p. 340.
19. Lundstrom, B.N.; Higgs, M.H.; Spain, W.J.; Fairhall, A.L. Fractional differentiation by neocortical pyramidal neurons. *Nat. Neurosci.* **2008**, *11*, 1335–1342. [[CrossRef](#)]
20. Rossikhin, A.; Shitikova, M.V. Application of fractional calculus for dynamic problems of solid mechanics: Novel trends and recent results. *Appl. Mech. Rev.* **2010**, *63*, 010801. [[CrossRef](#)]
21. Du, M.; Wang, Z.; Hu, H. Measuring memory with the order of fractional derivative. *Sci. Rep.* **2013**, *3*, 3431. [[CrossRef](#)] [[PubMed](#)]
22. Fisher, R.A. Theory of Statistical Estimation. *Proc. Camb. Philos. Soc.* **1925**, *22*, 700. [[CrossRef](#)]
23. Amari, S. *Differential-Geometrical Methods in Statistics*; Lecture Notes in Statistics; Springer: Berlin/Heidelberg, Germany, 1985.
24. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; Wiley: New York, NY, USA, 1991.
25. Watanabe, S.; Accardi, L.; Freudenberg, W.; Ohya, M. (Eds.) *Algebraic Geometrical Method in Singular Statistical Estimation*; Series in Quantum Bio-Informatics; World Scientific: Singapore, 2008; pp. 325–336.
26. James, R.G.; Ellison, C.J.; Crutchfield, J.P. Anatomy of a bit: Information in a time series observation. *Chaos* **2011**, *21*, 037109. [[CrossRef](#)] [[PubMed](#)]

27. Travers, N.F.; Crutchfield, J.P. Infinite excess entropy processes with countable-state generators. *Entropy* **2014**, *16*, 1396–1413. [[CrossRef](#)]
28. Marzen, S.; Crutchfield, J.P. Information anatomy of stochastic equilibria. *Entropy* **2014**, *16*, 4713–4748. [[CrossRef](#)]
29. Granger, C.W.J.; Joyeux, R. An introduction to long memory time series models and fractional differencing. *J. Time Ser. Anal.* **1980**, *1*, 15–39. [[CrossRef](#)]
30. Hosking, J.R.M. Fractional differencing. *Biometrika* **1981**, *68*, 165–176. [[CrossRef](#)]
31. Ghysels, E.; Swanson, N.R.; Watson, M.W. *Essays in Econometrics, Collected Papers of Clive W.J. Granger, Volume II: Causality, Integration and Cointegration, and Long Memory*; Cambridge University Press: Cambridge, UK, 2001; p. 398.
32. Gil-Alana, L.A.; Hualde, J. Fractional Integration and Cointegration: An Overview and an Empirical Application. In *Palgrave Handbook of Econometrics: Volume 2: Applied Econometrics*; Mills, T.C., Patterson, K., Eds.; Springer: Berlin/Heidelberg, Germany, 2009; pp. 434–469.
33. Tarasov, V.; Tarasova, V. Long and Short Memory in Economics: Fractional-Order Difference and Differentiation. *IRA-Int. J. Manag. Soc. Sci.* **2016**, *5*, 327–334. [[CrossRef](#)]
34. Samko, S.G.; Kilbas, A.A.; Marichev, O.I. *Fractional Integrals and Derivatives Theory and Applications*; Gordon and Breach: New York, NY, USA, 1993; p. 1006.
35. Kilbas, A.A.; Srivastava, H.M.; Trujillo, J.J. *Theory and Applications of Fractional Differential Equations*; Elsevier: Amsterdam, The Netherlands, 2006; p. 540.
36. Tarasov, V.E. *Fractional Dynamics: Applications of Fractional Calculus to Dynamics of Particles, Fields and Media*; Springer: New York, NY, USA, 2010; p. 505. [[CrossRef](#)]
37. Mathai, M.; Haubold, H.J. *Special Functions for Applied Scientists*; Springer: New York, NY, USA, 2008.
38. Volchenkov, D. *Grammar of Complexity: From Mathematics to a Sustainable World*; World Scientific Series “Nonlinear Physical Science”; World Scientific International Publishing: Singapore, 2018.



© 2019 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).