*Article*

# Informational and Causal Architecture of Discrete-Time Renewal Processes

**Sarah E. Marzen** [1],* **and James P. Crutchfield** [2],*

[1] Department of Physics, University of California at Berkeley, Berkeley, CA 94720-5800, USA

[2] Complexity Sciences Center and Department of Physics, University of California at Davis, One Shields Avenue, Davis, CA 95616, USA

\* Authors to whom correspondence should be addressed; E-Mails: smarzen@berkeley.edu (S.E.M.), chaos@ucdavis.edu (J.P.C.).

**Abstract:** Renewal processes are broadly used to model stochastic behavior consisting of isolated events separated by periods of quiescence, whose durations are specified by a given probability law. Here, we identify the minimal sufficient statistic for their prediction (the set of causal states), calculate the historical memory capacity required to store those states (statistical complexity), delineate what information is predictable (excess entropy), and decompose the entropy of a single measurement into that shared with the past, future, or both. The causal state equivalence relation defines a new subclass of renewal processes with a finite number of causal states despite having an unbounded interevent count distribution. We use the resulting formulae to analyze the output of the parametrized Simple Nonunifilar Source, generated by a simple two-state hidden Markov model, but with an infinite-state $\epsilon$-machine presentation. All in all, the results lay the groundwork for analyzing more complex processes with infinite statistical complexity and infinite excess entropy.

## 1. Introduction

Stationary renewal processes are widely used, analytically tractable, compact models of an important class of point processes [1–4]. Realizations consist of sequences of events, e.g., neuronal spikes or earthquakes, separated by epochs of quiescence, the lengths of which are drawn independently from the same interevent distribution. Renewal processes on their own have a long history and, due to their offering a parsimonious mechanism, often are implicated in highly complex behavior [5–10]. Additionally, understanding more complicated processes [11–14] requires fully analyzing renewal processes and their generalizations.

As done here and elsewhere [15], analyzing them in-depth from a structural information viewpoint yields new statistical signatures of apparent high complexity: long-range statistical dependence, memory, and internal structure. To that end, we derive the causal-state minimal sufficient statistics, the $\epsilon$-machine, for renewal processes and then derive new formulae for their various information measures in terms of the interevent count distribution. The result is a thorough-going analysis of their information architecture, a shorthand referring to a collection of measures that together quantify key process properties: predictability, difficulty of prediction, inherent randomness, memory, Markovity, and the like. The measures include:

- the statistical complexity $C_\mu$, which quantifies the historical memory that must be stored in order to predict a process's future;
- the entropy rate $h_\mu$, which quantifies a process' inherent randomness as the uncertainty in the next observation, even given that we can predict as well as possible;
- the excess entropy $\mathbf{E}$, which quantifies how much of a process's future is predictable in terms of the mutual information between its past and future;
- the bound information $b_\mu$, which identifies the portion of the inherent randomness ($h_\mu$) that affects a process's future in terms of the information in the next observation shared with the future, above and beyond that of the entire past; and
- the elusive information $\sigma_\mu$, which quantifies a process's deviation from Markovity as the mutual information between the past and future conditioned on the present.

Analyzing a process in this way gives a more detailed understanding of its structure and stochasticity. Beyond this, these information measures are key to finding limits to a process's optimal lossy predictive features [16–19], designing action policies for intelligent autonomous agents [20], and quantifying whether or not a given process has one or another kind of infinite memory [21–23].

While it is certainly possible to numerically estimate information measures directly from trajectory data, statistical methods generally encounter a curse of dimensionality when a renewal process has long-range temporal correlations, since the number of typical trajectories grows exponentially (at entropy rate $h_\mu$). Alternatively, we gain substantial advantages by first building a maximally-predictive model of a process (e.g., using Bayesian inference [24]) and then using that model to calculate information measures (e.g., using recently-available closed-form expressions when the model is finite [25]). Mathematicians have known for over a half century [26] that alternative models that are not maximally predictive are inadequate for such calculations. Thus, maximally-predictive models are critical. Figure 1 depicts the overall procedure just outlined, highlighting its important role. Here, extending the benefits of this procedure, we determine formulae for the information measures mentioned above and the appropriate structures for a class of processes that require countably infinite models: the ubiquitous renewal processes.
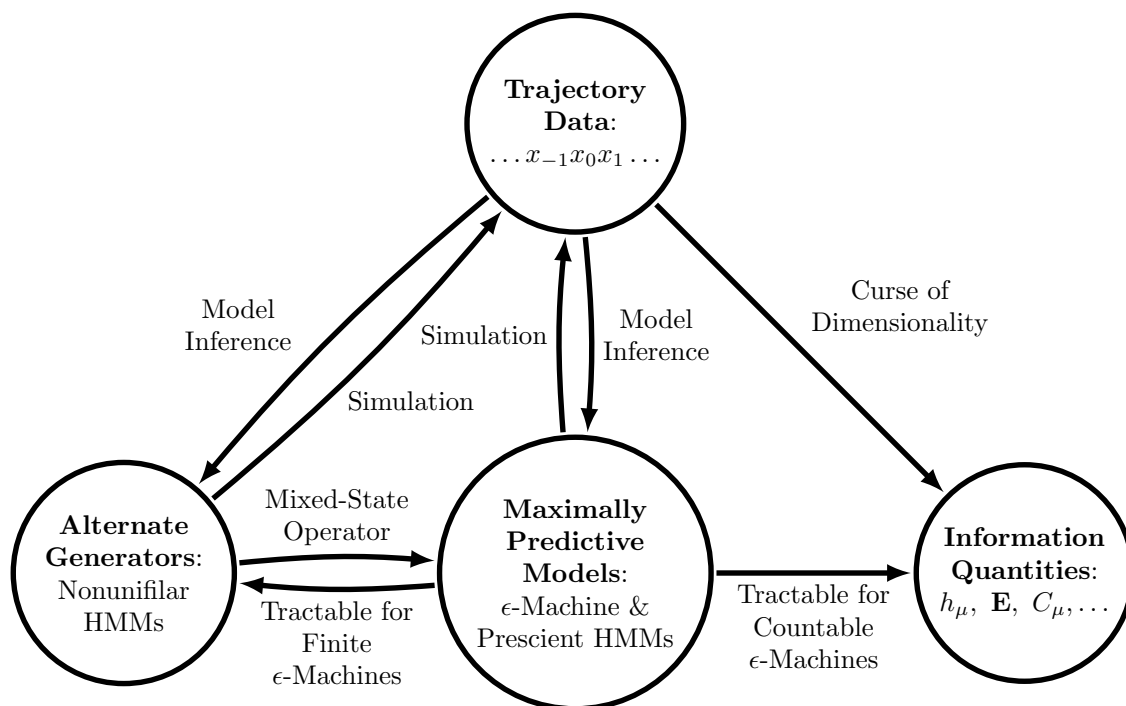
**Figure 1.** The role of maximally-predictive (prescient) models: Estimating information measures directly from trajectory data encounters a curse of dimensionality or, in other words, severe undersampling. Instead, one can calculate information measures in closed-form from (inferred) maximally-predictive models [25]. Alternate generative models that are not maximally predictive cannot be used directly, as Blackwell showed in the 1950s [26].

Our development requires familiarity with computational mechanics [27]. Those disinterested in its methods, but who wish to use the results, can skip to Figure 3a–d and Table 1. A pedagogical example is provided in Section 5. Two sequels will use the results to examine the limit of infinitesimal time resolution for information in neural spike trains [28] and the conditions under which renewal processes have infinite excess entropy [29].

The development is organized as follows. Section 2 provides a quick introduction to computational mechanics and prediction-related information measures of stationary time series. Section 3 identifies the causal states (in both forward and reverse time), the statistical complexity, and the $\epsilon$-machine of discrete-time stationary renewal processes. Section 4 calculates the information architecture and predictable information of a discrete-time stationary renewal process in terms of the interevent count distribution. Section 5 calculates these information-theoretic measures for the parametrized simple nonunifilar source, a simple two-state hidden Markov model with a countable infinity of causal states. Finally, Section 6 summarizes the results and lessons, giving a view toward future directions and mathematical and empirical challenges.

## 2. Background

We first describe renewal processes, then introduce a small piece of information theory, review the definition of process structure, and, finally, recall several information-theoretic measures designed to capture organization in structured processes.

## 2.1. Renewal Processes

We are interested in a system's immanent, possibly emergent, properties. To this end, we focus on behaviors and not, for example, particular equations of motion or particular forms of stochastic differential or difference equations. The latter are important in applications, because they are generators of behavior, as we will see shortly in Section 2.3. As Figure 1 explains, for a given process, some of its generators facilitate calculating key properties. Others lead to complicated calculations, and others still cannot be used at all.

As a result, our main object of study is a process $\mathcal{P}$: the list of all of a system's behaviors or realizations $\{\ldots, x_{-2}, x_{-1}, x_0, x_1, \ldots\}$ as specified by their measure $\mu(\ldots, X_{-2}, X_{-1}, X_0, X_1, \ldots)$. We denote a contiguous chain of random variables as $X_{0:L} = X_0 X_1 \cdots X_{L-1}$. Left indices are inclusive; right, exclusive. We suppress indices that are infinite. In this setting, the present $X_0$ is the random variable measured at $t = 0$; the past is the chain $X_{:0} = \cdots X_{-2} X_{-1}$ leading up the present; and the future is the chain following the present $X_{1:} = X_1 X_2 \cdots$. The joint probabilities $\Pr(X_{0:N})$ of sequences are determined by the measure of the corresponding cylinder sets: $\Pr(X_{0:N} = x_0 x_1 \ldots x_{N-1}) = \mu(\ldots, x_0, x_1, \ldots, x_{N-1}, \ldots)$. Finally, we assume that a process is ergodic and stationary ($\Pr(X_{0:L}) = \Pr(X_{t:L+t})$ for all $t, L \in \mathbb{Z}$) and that the observation values $x_t$ range over a finite alphabet: $x \in \mathcal{A}$. In short, we work with hidden Markov processes [30].

Discrete-time stationary renewal processes here have binary observation alphabets $\mathcal{A} = \{0, 1\}$. Observation of the binary symbol 1 is called an event. The event count is the number of 0's between successive 1's. Counts $n$ are i.i.d. random variables drawn from an interevent distribution $F(n)$, $n \geq 0$. We restrict ourselves to persistent renewal processes, such that the probability distribution function is normalized: $\sum_{n=0}^{\infty} F(n) = 1$. This translates into the processes being ergodic and stationary. We also define the survival function by $w(n) = \sum_{n'=n}^{\infty} F(n')$, and the expected interevent count is given by $\mu = \sum_{n=0}^{\infty} n F(n)$. We assume also that $\mu < \infty$. It is straightforward to check that $\sum_{n=0}^{\infty} w(n) = \mu + 1$.

Note the dual use of $\mu$. On the one hand, it denotes the measure over sequences and, since it determines probabilities, it appears in names for informational quantities. On the other hand, it is commonplace in renewal process theory and denotes mean interevent counts. Fortunately, context easily distinguishes the meaning through the very different uses.

## 2.2. Process Unpredictability

The information or uncertainty in a process is often defined as the Shannon entropy $H[X_0]$ of a single symbol $X_0$ [31]:

$$H[X_0] = -\sum_{x \in \mathcal{A}} \Pr(X_0 = x) \log_2 \Pr(X_0 = x) \,. \tag{1}$$

However, since we are interested in general complex processes, those with arbitrary dependence structure, we employ the block entropy to monitor information in long sequences:

$$\begin{aligned} H(L) &= H[X_{0:L}] \\ &= -\sum_{w \in \mathcal{A}^L} \Pr(X_{0:L} = w) \log_2 \Pr(X_{0:L} = w) \,. \end{aligned}$$

To measure a process's asymptotic per-symbol uncertainty, one then uses the Shannon entropy rate:

$$h_\mu = \lim_{L \to \infty} \frac{H(L)}{L} \,,$$

when the limit exists. (Here and elsewhere, μ reminds us that information quantities depend on the process's measure μ over sequences.) $h_\mu$ quantifies the rate at which a stochastic process generates information. Using standard informational identities, one sees that the entropy rate is also given by the conditional entropy:

$$h_\mu = \lim_{L \to \infty} H[X_0 | X_{-L:0}] \,. \tag{2}$$

This form makes transparent its interpretation as the residual uncertainty in a measurement given the infinite past. As such, it is often employed as a measure of a process's degree of unpredictability.

### 2.3. Maximally Predictive Models

Forward-time causal states $\boldsymbol{\mathcal{S}}^+$ are minimal sufficient statistics for predicting a process's future [32,33]. This follows from their definition: a causal state $\sigma^+ \in \boldsymbol{\mathcal{S}}^+$ is a set of pasts grouped by the equivalence relation $\sim^+$:

$$x_{:0} \sim^+ x'_{:0} \quad \Leftrightarrow \quad \Pr(X_{0:} | X_{:0} = x_{:0}) = \Pr(X_{0:} | X_{:0} = x'_{:0}) \,. \tag{3}$$

Therefore, $\boldsymbol{\mathcal{S}}^+$ is a set of classes, a coarse-graining of the uncountably infinite set of all pasts. At time $t$, we have the random variable $\mathcal{S}_t^+$ that takes values $\sigma^+ \in \boldsymbol{\mathcal{S}}^+$ and describes the causal-state process $\ldots, \mathcal{S}_{-1}^+, \mathcal{S}_0^+, \mathcal{S}_1^+, \ldots$. $\mathcal{S}_t^+$ is a partition of pasts $X_{:t}$ that, according to the indexing convention, does not include the present observation $X_t$. In addition to the set of pasts leading to it, a causal state $\sigma_t^+$ has an associated future morph: the conditional measure $\mu(X_{t:} | \sigma_t^+)$ of futures that can be generated from it. Moreover, each state $\sigma_t^+$ inherits a probability $\pi(\sigma_t^+)$ from the process's measure over pasts $\mu(X_{:t})$. The forward-time statistical complexity is defined as the Shannon entropy of the probability distribution over forward-time causal states [32]:

$$C_\mu^+ = H[\mathcal{S}_0^+] \,. \tag{4}$$

A generative model is constructed out of the causal states by endowing the causal-state process with transitions:

$$T_{\sigma\sigma'}^{(x)} = \Pr(\mathcal{S}_{t+1}^+ = \sigma', X_t = x | \mathcal{S}_t^+ = \sigma) \,,$$

that give the probability of generating the next symbol $x$ and ending in the next state $\sigma'$, if starting in state $\sigma$. (Residing in a state and generating a symbol do not occur simultaneously. Since symbols are generated during transitions there is, in effect, a half time step difference in the indexes of the random variables $X_t$ and $\mathcal{S}_t^+$. We suppress notating this.) To summarize, a process's forward-time $\epsilon$-machine is the tuple $\{\mathcal{A}, \boldsymbol{\mathcal{S}}^+, \{T^{(x)} : x \in \mathcal{A}\}\}$.

For a discrete-time, discrete-alphabet process, the $\epsilon$-machine is its minimal unifilar hidden Markov model (HMM) [32,33]. (For a general background on HMMs, see [34–36].) Note that the causal state set can be finite, countable, or uncountable; the latter two cases can occur even for processes generated by finite-state HMMs. Minimality can be defined by either the smallest number of states or the smallest

entropy over states [33]. Unifilarity is a constraint on the transition matrices $T^{(x)}$, such that the next state $\sigma'$ is determined by knowing the current state $\sigma$ and the next symbol $x$. That is, if the transition exists, then $\Pr(\mathcal{S}_{t+1}^{+}|X_t = x, \mathcal{S}_t^{+} = \sigma)$ has support on a single causal state.

While the $\epsilon$-machine is a process's minimal, maximally-predictive model, there can be alternative HMMs that are as predictive, but are not minimal. We refer to the maximally-predictive property by referring to the $\epsilon$-machine and these alternatives as prescient. The state and transition structure of a prescient model allow one to immediately calculate the entropy rate $h_\mu$, for example. More generally, any statistic that gives the same (optimal) level of predictability we call a prescient statistic.

A similar equivalence relation can be applied to find minimal sufficient statistics for retrodiction [37]. Futures are grouped together if they have equivalent conditional probability distributions over pasts:

$$x_{0:} \sim^{-} x'_{0:} \Leftrightarrow \Pr(X_{:0}|X_{0:} = x_{0:}) = \Pr(X_{:0}|X_{0:} = x'_{0:}) . \tag{5}$$

A cluster of futures, a reverse-time causal state, defined by $\sim^{-}$, is denoted $\sigma^{-} \in \boldsymbol{\mathcal{S}}^{-}$. Again, each $\sigma^{-}$ inherits a probability $\pi(\sigma^{-})$ from the measure over futures $\mu(X_{0:})$. Additionally, the reverse-time statistical complexity is the Shannon entropy of the probability distribution over reverse-time causal states:

$$C_\mu^{-} = H[\mathcal{S}_0^{-}] . \tag{6}$$

In general, the forward and reverse-time statistical complexities are not equal [37,38]. That is, different amounts of information must be stored from the past (future) to predict (retrodict). Their difference $\Xi = C_\mu^{+} - C_\mu^{-}$ is a process's causal irreversibility, and it reflects this statistical asymmetry.

Since we work with stationary processes in the following, the time origin is arbitrary and, so, we drop the time index $t$ when it is unnecessary.

## 2.4. Information Measures for Processes

Shannon's various information quantities—entropy, conditional entropy, mutual information, and the like—when applied to time series are functions of the joint distributions $\Pr(X_{0:L})$. Importantly, they define an algebra of information measures for a given set of random variables [39]. The work in [40] used this to show that the past and future partition the single-measurement entropy $H(X_0)$ into several distinct measure-theoretic atoms. These include the ephemeral information:

$$r_\mu = H[X_0|X_{:0}, X_{1:}] , \tag{7}$$

which measures the uncertainty of the present knowing the past and future; the bound information:

$$b_\mu = I[X_0; X_{1:}|X_{:0}] , \tag{8}$$

which is the mutual information shared between present and future conditioned on the past; and the enigmatic information:

$$q_\mu = I[X_0; X_{:0}; X_{1:}] , \tag{9}$$

which is the three-way mutual information between past, present and future. Multi-way mutual information is sometimes referred to as co-information [40,41] and, compared to Shannon entropies and two-way mutual information, can have counterintuitive properties, such as being negative.

For a stationary time series, the bound information is also the shared information between present and past conditioned on the future:

$$b_\mu = I[X_0; X_{:0}|X_{1:}].\tag{10}$$

One can also consider the amount of predictable information not captured by the present:

$$\sigma_\mu = I[X_{:0}; X_{1:}|X_0].\tag{11}$$

This is called the elusive information [42]. It measures the amount of past-future correlation not contained in the present. It is nonzero if the process necessarily has hidden states and is therefore quite sensitive to how the state space is observed or coarse grained.

The maximum amount of information in the future predictable from the past (or *vice versa*) is the excess entropy:

$$\mathbf{E} = I[X_{:0}; X_{0:}].$$

It is symmetric in time and a lower bound on the stored informations $C_\mu^+$ and $C_\mu^-$. It is directly given by the information atoms above:

$$\mathbf{E} = b_\mu + \sigma_\mu + q_\mu.\tag{12}$$

The process's Shannon entropy rate $h_\mu$ (recall the form of Equation (2)) can also be written as a sum of atoms:

$$\begin{aligned}h_\mu &= H[X_0|X_{:0}]\\ &= r_\mu + b_\mu.\end{aligned}$$

Thus, a portion of the information ($h_\mu$) a process spontaneously generates is thrown away ($r_\mu$), and a portion is actively stored ($b_\mu$). Putting these observations together gives the information architecture of a single measurement (Equation (1)):

$$H[X_0] = q_\mu + 2b_\mu + r_\mu.\tag{13}$$

These identities can be used to determine $r_\mu$, $q_\mu$, and $\mathbf{E}$ from $H[X_0]$, $b_\mu$, and $\sigma_\mu$, for example.

We have a particular interest in when $C_\mu$ and $\mathbf{E}$ are infinite and, so, will investigate finite-time variants of causal states and finite-time estimates of statistical complexity and $\mathbf{E}$. For example, the latter is given by:

$$\mathbf{E}(M, N) = I[X_{-M:0}; X_{0:N}].\tag{14}$$

If $\mathbf{E}$ is finite, then $\mathbf{E} = \lim_{M,N\to\infty} \mathbf{E}(M, N)$. When $\mathbf{E}$ is infinite, then the way in which $\mathbf{E}(M, N)$ diverges is one measure of a process' complexity [21,43,44]. Analogous, finite past-future $(M, N)$-parametrized equivalence relations lead to finite-time forward and reverse causal states and statistical complexities $C_\mu^+(M, N)$ and $C_\mu^-(M, N)$.

## 3. Causal Architecture of Renewal Processes

It will be helpful pedagogically to anchor our theory in the contrast between two different, but still simple, renewal processes. One is the familiar "memoryless" Poisson process with rate $\lambda$. Its HMM generator, a biased coin, is shown at the left of Figure 2. It has an interevent count distribution $F(n) = (1 - \lambda)\lambda^n$; a distribution with unbounded support. However, we notice in Figure 2 that it is a unifilar model with a minimal number of states. Therefore, in fact, this one-state machine is the $\epsilon$-machine of a Poisson process. The rate at which it generates information is given by the entropy rate: $h_\mu = H(\lambda)$ bits per output symbol. (Here, $H(p)$ is the binary entropy function.) It also has a vanishing statistical complexity $C_\mu^+ = 0$ and, so, stores no historical information.
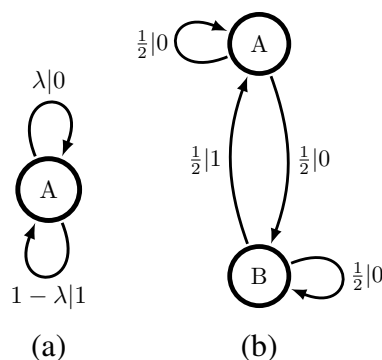


(a)  (b)

**Figure 2.** (**a**) Minimal generative model for the Poisson process with rate $\lambda$; (**b**) a generator for the Simple Nonunifilar Source (SNS). Both generate a stationary renewal process. Transition labels $p|s$ denote the probability $p$ of taking a transition and emitting symbol $s$.

The second example is the Simple Nonunifilar Source (SNS) [45]; its HMM generator is shown in Figure 2b. Transitions from state $B$ are unifilar, but transitions from state $A$ are not. In fact, a little reflection shows that the time series produced by the SNS is a discrete-time renewal process. Once we observe the "event" $x_t = 1$, we know the internal model state to be $\sigma_{t+1} = A$, so successive interevent counts are completely uncorrelated.

This SNS generator is not an $\epsilon$-machine and, moreover, cannot be used to calculate the process's information per output symbol (entropy rate). If we can only see $0$'s and $1$'s, we will usually be uncertain as to whether we are in state $A$ or state $B$, so this generative model is not maximally predictive. How can we calculate this basic quantity? Additionally, if we cannot use the two-state generator, how many states are required, and what is their transition dynamic? The following uses computational mechanics to answer these and a number of related questions. To aid readability, though, we sequester most of the detailed calculations and proofs to Appendix A.

We start with a simple Lemma that follows directly from the definitions of a renewal process and the causal states. It allows us to introduce notation that simplifies the development.

**Lemma 1.** *The count since last event is a prescient statistic of a discrete-time stationary renewal process.*

That is, if we remember only the number of counts since the last event and nothing prior, we can predict the future as well as if we had memorized the entire past. Specifically, a prescient state $\mathcal{R}$ is a function of the past such that:

$$H[X_{0:}|X_{:0}] = H[X_{0:}|\mathcal{R}] \ .$$

Causal states can be written as unions of prescient states [33]. We start with a definition that helps to characterize the converse; *i.e.*, when the prescient states of Lemma 1 are also causal states.

To ground our intuition, recall that Poisson processes are "memoryless". This may seem counterintuitive, if viewed from a parameter estimation point of view. After all, if observing longer pasts, one makes better and better estimates of the Poisson rate. However, finite data fluctuations in estimating model parameters are irrelevant to the present mathematical setting unless the parameters are themselves random variables, as in the nonergodic processes of [44]. This is not our setting here: the parameters are fixed. In fact, we restrict ourselves to studying ergodic processes, in which the conditional probability distributions of futures given pasts of a Poisson process are independent of the past.
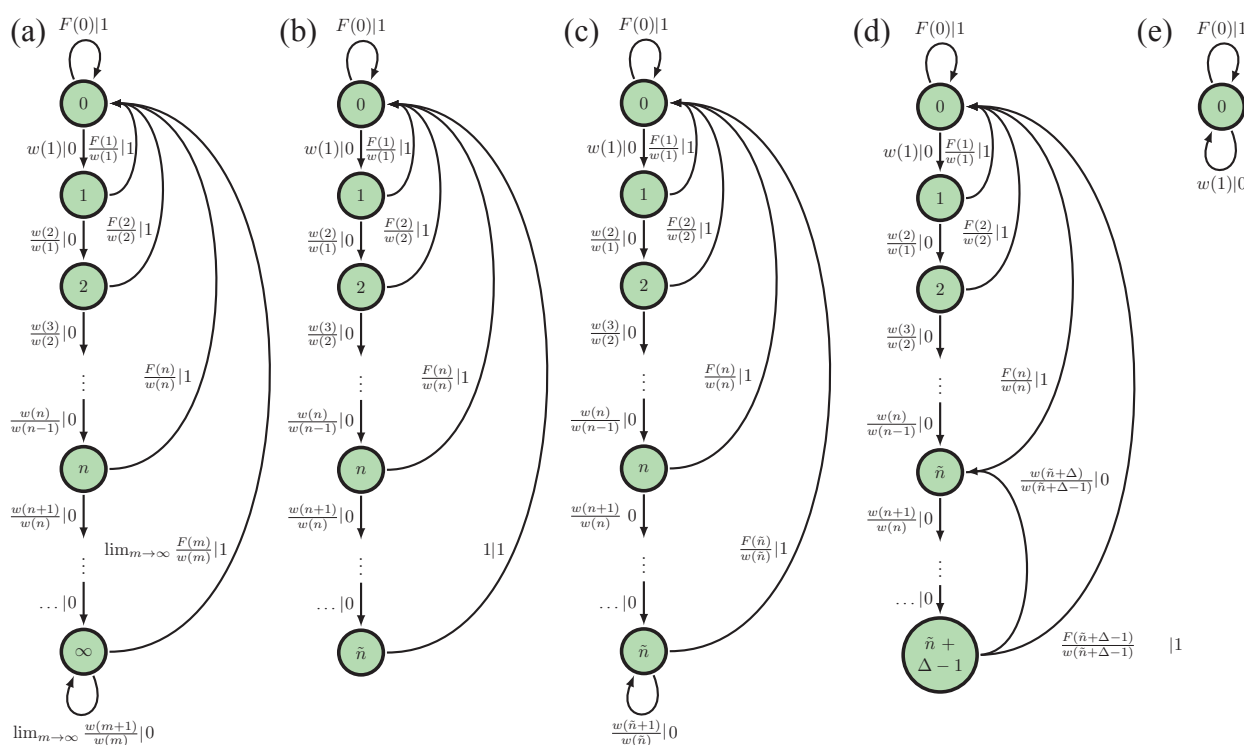


**Figure 3.** $\epsilon$-Machine architectures for discrete-time stationary renewal processes: (**a**) not eventually $\Delta$-Poisson with unbounded-support interevent distribution; (**b**) not eventually $\Delta$-Poisson with bounded-support interevent distribution; (**c**) eventually $\Delta$-Poisson with characteristic $(\tilde{n}, \Delta = 1)$ in Definition 2; (**d**) eventually $\Delta$-Poisson with characteristic $(\tilde{n}, \Delta > 1)$ in Definition 2; (**e**) for comparison, a Poisson process $\epsilon$-machine.

We therefore expect the prescient states in Lemma 1 to fail to be causal states precisely when the interevent distribution is similar to that of a Poisson renewal process. This intuition is made precise by Definition 2.

**Definition 1.** *A $\Delta$-Poisson process has an interevent distribution:*

$$F(n) = F(n \bmod \Delta)\, \lambda^{\lfloor n/\Delta \rfloor} \ ,$$

*for all $n$ and some $\lambda > 0$. If this statement holds for multiple $\Delta \geq 1$, then we choose the smallest possible $\Delta$.*

**Definition 2.** *A $(\tilde{n}, \Delta)$ eventually $\Delta$-Poisson process has an interevent distribution that is $\Delta$-Poisson for all $n \geq \tilde{n}$:*

$$F(\tilde{n} + k\Delta + m) = \lambda^k F(\tilde{n} + m) \, ,$$

*for all $0 \leq m < \Delta$, for all $k \geq 0$, and for some $\lambda > 0$. If this statement holds for multiple $\Delta \geq 1$ and multiple $\tilde{n}$, then we choose the smallest possible $\Delta$ and the smallest possible $\tilde{n}$.*

If this statement holds for multiple $\Delta \geq 1$ and multiple $n$, first, we choose the smallest possible $\Delta$ and, then, select the smallest possible $n$ for that $\Delta$. Order matters here.

Thus, a Poisson process is a $\Delta$-Poisson process with $\Delta = 1$ and an eventually $\Delta$-Poisson process with $\Delta = 1$ and $\tilde{n} = 0$. Moreover, we will now show that at some finite $\tilde{n}$, any renewal process is either: (i) Poisson, if $\Delta = 1$; or (ii) a combination of several Poisson processes, if $\Delta > 1$.

Why identify new classes of renewal process? In short, renewal processes that are similar to, but not the same as, the Poisson process do not have an infinite number of causal states. The particular condition for when they do not is given by the eventually $\Delta$-Poisson definition. Notably, this new class is what emerged, rather unexpectedly, by applying the causal-state equivalence relation $\sim^+$ to renewal processes. The resulting insight is that general renewal processes, after some number of counts (the "eventually" part) and after some coarse-graining of counts (the $\Delta$ part), behave like a Poisson process.

With these definitions in hand, we can proceed to identify the causal architecture of discrete-time stationary renewal processes.

**Theorem 1.** *(a) The forward-time causal states of a discrete-time stationary renewal process that is not eventually $\Delta$-Poisson are groupings of pasts with the same count since the last event; (b) the forward-time causal states of a discrete-time eventually $\Delta$-Poisson stationary renewal process are groupings of pasts with the same count since the last event up until $\tilde{n}$ and pasts whose count $n$ since the last event are in the same equivalence class as $\tilde{n}$ modulo $\Delta$.*

The Poisson process, as an eventually $\Delta$-Poisson with $\tilde{n} = 0$ and $\Delta = 1$, is represented by the one-state $\epsilon$-machine despite the unbounded support of its interevent count distribution. Unlike most processes, the Poisson process' $\epsilon$-machine is the same as its generative model shown in Figure 2a.

The SNS, on the other hand, has an interevent count distribution that is not eventually $\Delta$-Poisson. According to Theorem 1, then, the SNS has a countable infinity of causal states despite its simple two-state generative model in Figure 2b. Compare Figure 3a. Each causal state corresponds to a different probability distribution over the internal states $A$ and $B$. These internal state distributions are the mixed states of [46]. Observing more 0's, one becomes increasingly convinced that the internal state is $B$. For maximal predictive power, however, we must track the probability that the process is still in state $A$. Both Figure 3a and Figure 2b are "minimally complex" models of the same process, but with different definitions of model complexity. We return to this point in Section 5.

Appendix A makes the statements in Theorem 1 precisely. The main result is that causal states are sensitive to two features: (i) eventually $\Delta$-Poisson structure in the interevent distribution and (ii) the boundedness of $F(n)$'s support. If the support is bounded, then there are a finite number of causal states rather than a countable infinity of causal states. Similarly, if $F(n)$ has $\Delta$-Poisson tails, then there are a finite number of causal states despite the support of $F(n)$ having no bound. Nonetheless, one can say that the generic discrete-time stationary renewal process has a countable infinity of causal states.

Finding the probability distribution over these causal states is straightforwardly related to the survival-time distribution $w(n)$ and the mean interevent interval μ, since the probability of observing at least $n$ counts since the last event is $w(n)$. Hence, the probability of seeing $n$ counts since the last event is simply the normalized survival function $w(n)/(\mu + 1)$. Appendix A derives the statistical complexity using this and Theorem 1. The resulting formulae are given in Table 1 for the various cases.

As described in Section 2, we can also endow the causal state space with a transition dynamic in order to construct the renewal process $\epsilon$-machine: the process's minimal unifilar hidden Markov model. The transition dynamic is sensitive to $F(n)$'s support and not only its boundedness. For instance, the probability of observing an event given that it has been $n$ counts since the last event is $F(n)/w(n)$. For the generic discrete-time renewal process, this is exactly the transition probability from causal state $n$ to causal state 0. If $F(n) = 0$, then there is no probability of transition from $\sigma = n$ to $\sigma = 0$. See Appendix A for details.

**Table 1.** Structural measures and information architecture of a stationary renewal process with interevent counts drawn from the distribution $F(n)$, $n \geq 0$, survival count distribution $w(n) = \sum_{m=n}^{\infty} F(m)$, and mean interevent count $\mu = \sum_{n=0}^{\infty} nF(n) < \infty$. The function $g(m, n)$ is defined by $g(m, n) = F(m + n + 1) + F(m)F(n)$. Cases are needed for $C_\mu$, but not other quantities, such as block entropy and information architecture quantities, since the latter can be calculated just as well from prescient machines. The quantities $\chi$ (crypticity) and $\mathbf{E}(M, N) = I[X_{-M:0}; X_{0:N}]$ are no less interesting than the others given here, but their expressions are not compact; see Appendix B.

| Quantity | Expression | |
|---|---|---|
| $C_\mu^+ = H[\mathcal{S}^+]$ | $-\sum_{n=0}^{\infty} \frac{w(n)}{\mu+1} \log_2 \frac{w(n)}{\mu+1}$ | Not eventually $\Delta$-Poisson |
| | $-\sum_{n=0}^{\tilde{n}-1} \frac{w(n)}{\mu+1} \log_2 \frac{w(n)}{\mu+1} - \sum_{m=0}^{\Delta-1} \frac{\sum_{k=0}^{\infty} w(\tilde{n}+k\Delta+m)}{\mu+1} \log_2 \frac{\sum_{k=0}^{\infty} w(\tilde{n}+k\Delta+m)}{\mu+1}$ | Eventually $\Delta$-Poisson |
| $\mathbf{E} = I[X_{:0}; X_{0:}]$ | $-2\sum_{n=0}^{\infty} \frac{w(n)}{\mu+1} \log_2 \frac{w(n)}{\mu+1} + \sum_{n=0}^{\infty} (n+1) \frac{F(n)}{\mu+1} \log_2 \frac{F(n)}{\mu+1}$ | |
| $h_\mu = H[X_0|X_{:0}]$ | $-\frac{1}{\mu+1} \sum_{n=0}^{\infty} F(n) \log_2 F(n)$ | |
| $b_\mu = I[X_{1:}; X_0|X_{:0}]$ | $\frac{1}{\mu+1} \left\{ \sum_{n=0}^{\infty} (n+1)F(n) \log_2 F(n) - \sum_{m,n=0}^{\infty} g(m,n) \log_2 g(m,n) \right\}$ | |
| $\sigma_\mu = I[X_{1:}; X_{:0}|X_0]$ | $\frac{1}{\mu+1} \left\{ \mu \log_2 \mu + \sum_{n=0}^{\infty} nF(n) \log_2 F(n) - 2\sum_{n=0}^{\infty} w(n) \log_2 w(n) \right\}$ | |
| $q_\mu = I[X_{1:}; X_0; X_{:0}]$ | $\frac{1}{\mu+1} \left\{ \sum_{m,n=0}^{\infty} g(m,n) \log_2 g(m,n) - \sum_{n=0}^{\infty} w(n) \log_2 w(n) + (\mu+1)\log_2(\mu+1) - \mu \log_2 \mu \right\}$ | |
| $r_\mu = H[X_0|X_{1:}, X_{:0}]$ | $\frac{1}{\mu+1} \left\{ \sum_{m,n=0}^{\infty} g(m,n) \log_2 g(m,n) - \sum_{n=0}^{\infty} (n+2)F(n) \log_2 F(n) \right\}$ | |
| $H_0 = H[X_0]$ | $-\frac{1}{\mu+1} \log_2 \frac{1}{\mu+1} - \left(1 - \frac{1}{\mu+1}\right) \log_2 \left(1 - \frac{1}{\mu+1}\right)$ | |

Figure 3a–d displays the causal state architectures, depicted as state-transition diagrams, for the $\epsilon$-machines in the various cases delineated. Figure 3a is the $\epsilon$-machine of a generic renewal process whose interevent interval can be arbitrarily large and whose interevent distribution never has exponential tails. Figure 3b is the $\epsilon$-machine of a renewal process whose interevent distribution never has exponential tails, but cannot have arbitrarily large interevent counts. The $\epsilon$-machine in Figure 3c looks quite similar to the $\epsilon$-machine in Figure 3b, but it has an additional transition that connects the last state $\tilde{n}$ to itself. This added transition changes our structural interpretation of the process. Interevent counts can be arbitrarily large for this $\epsilon$-machine, but past an interevent count of $\tilde{n}$, the interevent distribution is exponential. Finally, the $\epsilon$-machine in Figure 3d represents an eventually $\Delta$-Poisson process with $\Delta > 1$ whose

structure is conceptually most similar to that of the $\epsilon$-machine in Figure 3c. (See Definition 2 for the precise version of that statement.) If our renewal process disallows seeing interevent counts of a particular length $L$, then this will be apparent from the $\epsilon$-machine, since there will be no transition between the causal state corresponding to an interevent count of $L$ and causal state 0.

As described in Section 2, we can analytically characterize a process' information architecture far better once we characterize its statistical structure in reverse time.

**Lemma 2.** *Groupings of futures with the same counts to the next event are reverse-time prescient statistics for discrete-time stationary renewal processes.*

**Theorem 2.** *(a) The reverse-time causal states of a discrete-time stationary renewal process that is not eventually $\Delta$-Poisson are groupings of futures with the same count to the next event; (b) the reverse-time causal states of a discrete-time eventually $\Delta$-Poisson stationary renewal process are groupings of futures with the same count to the next event up until $\tilde{n}$ plus groupings of futures whose count since the last event $n$ are in the same equivalence class as $\tilde{n}$ modulo $\Delta$.*

As a result, in reverse time, a stationary renewal process is effectively the same stationary renewal process: counts between events are still independently drawn from $F(n)$. Thus, the causal irreversibility vanishes: $\Xi = 0$.

Moreover, these results taken together indicate that we can straightforwardly build a renewal process's bidirectional machine from these forward and reverse-time causal states, as described in [37,38,46]. Additional properties can then be deduced from the bidirectional machine, but we leave this for the future.

In closing, we note a parallel in [47]: their $A_t$ is the forward-time prescient statistic and $E_t$ is the reverse-time prescient statistic. They call these the backward and forward recurrence times, respectively. They address continuous-time processes, however, and so, the setting differs somewhat.

## 4. Information Architecture of Renewal Processes

As Section 2 described, many quantities that capture a process's predictability and randomness can be calculated from knowing the block entropy function $H(L)$. Often, the block entropy is estimated by generating samples of a process and estimating the entropy of a trajectory distribution. This method has the obvious disadvantage that at large $L$, there are $|\mathcal{A}|^L$ possible trajectories and $|\mathcal{A}|^{h_\mu L}$ typical trajectories. Therefore, one easily runs into the problem of severe undersampling, previously referred to as the curse of dimensionality. This matters most when the underlying process has long-range temporal correlations.

Nor can one calculate the block entropy and other such information measures exactly from generative models that are not maximally predictive (prescient). Then, the model states do not shield the past from the future. For instance, as noted above, one cannot calculate the SNS's entropy rate from its simple two-state generative HMM. The entropy of the next symbol given the generative model's current state ($A$ or $B$) actually underestimates the true entropy rate by assuming that we can almost always precisely determine the underlying model state from the past. For a sense of the fundamental challenge, see [26,48].

However, we can calculate the block entropy and various other information measures in closed-form from a maximally-predictive model. In other words, finding an $\epsilon$-machine allows one to avoid the curse of dimensionality inherently involved in calculating the entropy rate, excess entropy, or the other information measures discussed here.

Figure 1 summarizes the above points. This section makes good on the procedure outlined there by providing analytic formulae for various information measures of renewal processes. The formulae for the entropy rate of a renewal process is already well known, but all others are new.

Prescient HMMs built from the prescient statistics of Lemma 1 are maximally predictive models and correspond to the unifilar hidden Markov model shown in Figure 3a. The prescient machines make no distinction between eventually $\Delta$-Poisson renewal processes and ones that are not, but they do contain information about the support of $F(n)$ through their transition dynamics (see Appendix A). Appendix B describes how a prescient machine can be used to calculate all information architecture quantities: $r_\mu$, $b_\mu$, $\sigma_\mu$, $q_\mu$, and the more familiar Shannon entropy rate $h_\mu$ and excess entropy $\mathbf{E}$. A general strategy for calculating these quantities, as described in Section 2 and [19,40], is to calculate $b_\mu$, $h_\mu$, $\mathbf{E}$ and $H[X_0]$ and then to derive the other quantities using the information-theoretic identities given in Section 2.

Table 1 gives the results of these calculations. It helps one's interpretation to consider two base cases. For a Poisson process, we gain no predictive power by remembering specific pasts, and we would expect the statistical complexity, excess entropy, and bound information rate to vanish. The entropy rate and ephemeral information, though, are nonzero. One can check that this is, indeed, the case. For a periodic process with period $T$, in contrast, one can check that $\mu + 1 = T$, since the period is the length of the string of 0's (mean interevent interval $\mu$) concatenated with the subsequent event $x = 1$. The statistical complexity and excess entropy of this process are $\log_2 T$, and the entropy rate is $h_\mu = 0$, as expected.

Calculating the predictable information $\mathbf{E}(M, N)$ requires identifying finite-time prescient statistics, since the predictable information is the mutual information between forward-time causal states over pasts of length $M$ and reverse-time causal states over futures of length $N$. Such finite-time prescient statistics are identified in Corollary 1, below, and the predictable information is derived in Appendix B. The final expression is not included in Table 1 due to its length.

**Corollary 1.** *Forward-time (and reverse-time) finite-time $M$ prescient states of a discrete-time stationary renewal process are the counts from (and to) the next event up until and including $M$.*

All of these quantities can be calculated using a mixed-state presentation, as described in [46], though the formulae developed there are as yet unable to describe processes with a countably infinite set of mixed states. Calculations of finite-time entropy rate estimates using a mixed-state presentation are consistent with all other results here, though. Purely for simplicity, we avoid discussing mixed-state presentations, though interested readers can find relevant definitions in [46].

## 5. Nonunifilar HMMs and Renewal Processes

The task of inferring an $\epsilon$-machine for discrete-time, discrete-alphabet processes is essentially that of inferring minimal unifilar HMMs; what are sometimes also called "probabilistic deterministic" finite automata. In unifilar HMMs, the transition to the next hidden state given the previous one and next emitted symbol is determined. Nonunifilar HMMs are a more general class of time series models in which the transitions between underlying states given the next emitted symbol can be stochastic.

This simple difference in HMM structure has important consequences for calculating the predictable information, information architecture, and statistical complexity of time series generated by nonunifilar HMMs. First, note that for processes with a finite number of transient and recurrent causal states, these quantities can be calculated in closed form [25]. Second, the autocorrelation function and power spectrum can also be calculated in closed form for nonunifilar presentations [49]. Unlike these cases, though, most of Table 1's quantities defy current calculational techniques. As a result, exact

calculations of these prediction-related information measures for even the simplest nonunifilar HMMs can be surprisingly difficult.

To illustrate this point, we focus our attention on a parametrized version of the SNS shown in Figure 4a. As for the original SNS in Figure 2, transitions from state $B$ are unifilar, but transitions from state $A$ are not. As noted before, the time series generated by the parametrized SNS is a discrete-time renewal process with interevent count distribution:

$$F(n) = \begin{cases} (1-p)(1-q)(p^n - q^n)/(p-q) & p \neq q \,, \\ (1-p)^2 np^{n-1} & p = q \,. \end{cases} \quad (15)$$

Figure 4b also shows $F(n)$ at various parameter choices. The nonunifilar HMM there should be contrasted with the unifilar HMM presentation of the parametrized SNS, which is the $\epsilon$-machine in Figure 3a, with a countable infinity of causal states.
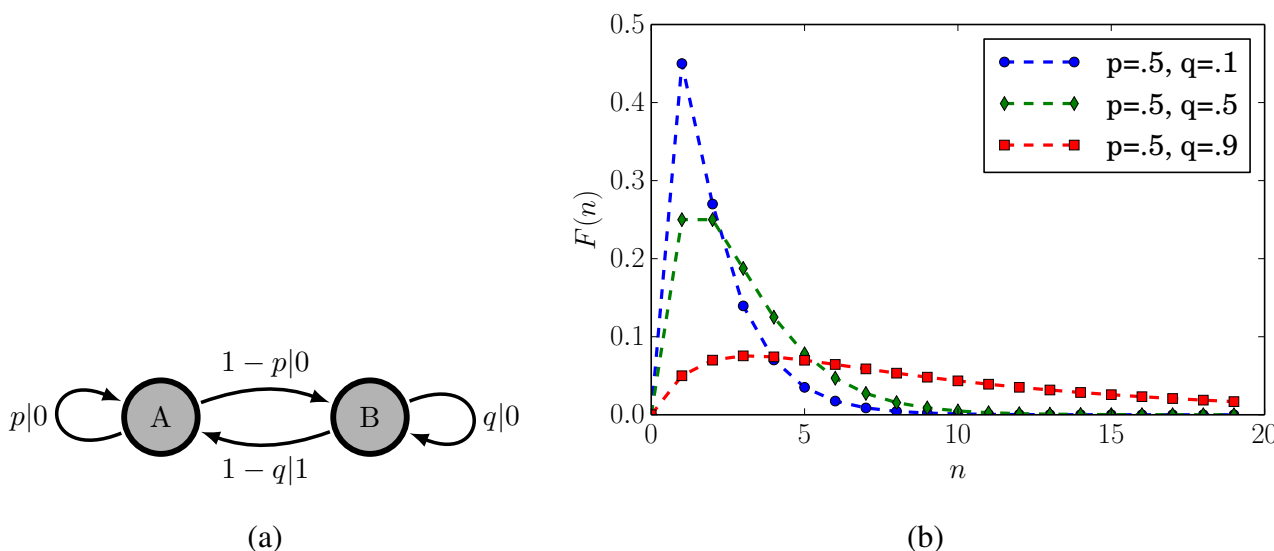


**Figure 4.** (**a**) A (nonunifilar) hidden Markov model for the $(p, q)$ parametrized SNS; (**b**) example interevent distributions $F(n)$ from Equation (15) for three parameter settings of $(p, q)$.

Both parametrized SNS presentations are "minimally complex", but according to different metrics. On the one hand, the nonunifilar presentation is a minimal generative model: No one-state HMM (*i.e.*, biased coin) can produce a time series with the same statistics. On the other, the unifilar HMM is the minimal maximally-predictive model: In order to predict the future as well as possible given the entire past, we must at least remember how many 0's have been seen since the last 1. That memory requires a countable infinity of prescient states. The preferred complexity metric is a matter of taste and desired implementation, modulo important concerns regarding overfitting or ease of inference [24]. However, if we wish to calculate the information measures in Table 1 as accurately as possible, finding a maximally predictive model, that is, a unifilar presentation, is necessary.

The SNS is not eventually $\Delta$-Poisson with an unbounded-support interevent distribution. It can only be approximated by an eventually $\Delta$-Poisson process. Therefore, one estimates its information anatomy via infinite sums. Using the formulae of Table 1, Figure 5 shows how the statistical complexity $C_\mu$,

excess entropy **E**, entropy rate $h_\mu$, and bound information $b_\mu$ vary with the transition probabilities $p$ and $q$. $C_\mu$ often reveals detailed information about a process' underlying structure, but for the parametrized SNS and other renewal processes, the statistical complexity merely reflects the spread of the interevent distribution. Thus, it increases with increasing $p$ and $q$. **E**, a measure of how much can be predicted rather than historical memory required for prediction, increases as $p$ and $q$ decrease. The intuition for this is that as $p$ and $q$ vanish, the process arrives at a perfectly predictable period-2 sequence. We see that the SNS constitutes a simple example of a class of processes over which information transmission between the past and future (**E**) and information storage ($C_\mu$) are anticorrelated. The entropy rate $h_\mu$ in Figure 5b is maximized when transitions are uniformly stochastic and the bound information $b_\mu$ in Figure 5d is maximized somewhere between fully stochastic and fully deterministic regimes.
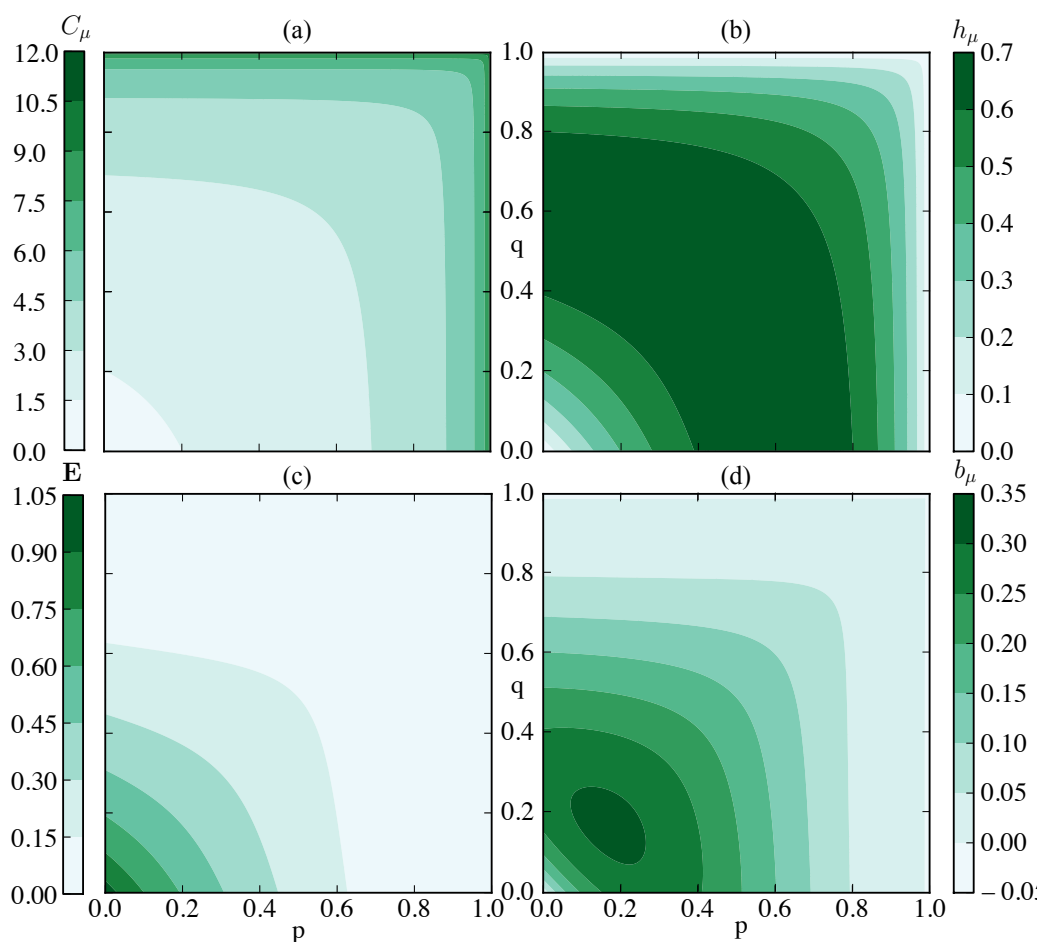


**Figure 5.** Contour plots of various information measures (in bits) as functions of SNS parameters $p$ and $q$. (**a**) $C_\mu$, increasing when $F(n)$ has slower decay; (**b**) $h_\mu$, higher when transition probabilities are maximally stochastic; (**c**) **E**, higher the closer the SNS comes to Period 2; (**d**) $b_\mu$, highest between the maximally-stochastic transition probabilities that maximize $h_\mu$ and maximally-deterministic transition probabilities that maximize **E**.

Figure 6 presents a more nuanced decomposition of the information measures as $p = q$ vary from zero to one. Figure 6a breaks down the single-measurement entropy $H[X_0]$ into redundant information $\rho_\mu$ in a single measurement, predictively useless generated information $r_\mu$, and predictively useful generated entropy $b_\mu$. As $p$ increases, the SNS moves from mostly predictable (close to Period-2) to mostly unpredictable, shown by the relative height of the (green) line denoting $h_\mu$ to the (red)

line denoting $H[X_0]$. The portion $b_\mu$ of $h_\mu$ predictive of the future is maximized at lower $p$ when the single-measurement entropy is close to a less noisy Period-2 process. Figure 6b decomposes the predictable information $\mathbf{E}$ into the predictable information hidden from the present $\sigma_\mu$, the predictable generated entropy in the present $b_\mu$, and the co-information $q_\mu$ shared between past, present, and future. Recall that the co-information $q_\mu = \mathbf{E} - \sigma_\mu - b_\mu$ can be negative, and for a large range of values, it is. Most of the predictable information passes through the present as indicated by $\sigma_\mu$ being small for most parameters $p$. Hence, even though the parametrized SNS is technically an infinite-order Markov process, it can be well approximated by a finite-order Markov process without much predictable information loss, as noted previously with rate distortion theory [50].
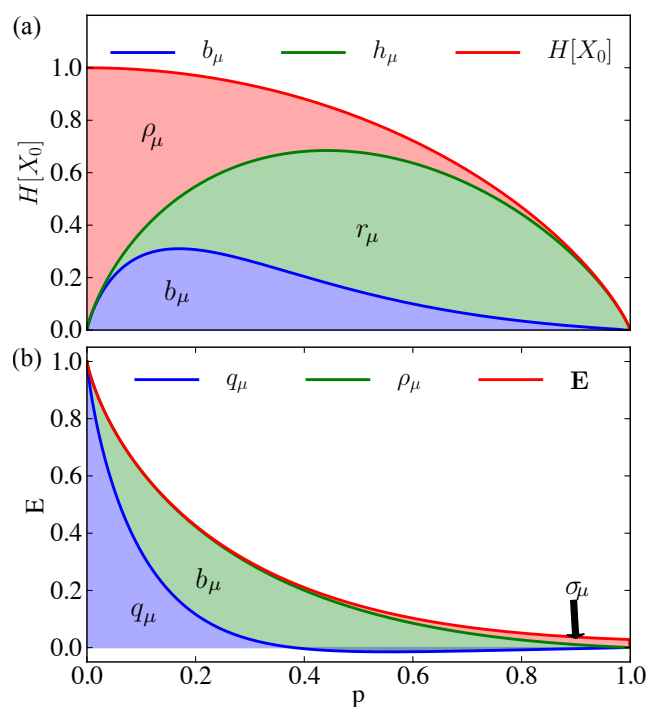


**Figure 6.** (**a**) Simple nonunifilar source information anatomy as a function of $p$ with parameters $q = p$. The single-measurement entropy $H[X_0]$ is the upper solid (red) line, entropy rate $h_\mu$ the middle solid (green) line and the bound information $b_\mu$ the lower solid (blue) line. Thus, the blue area corresponds to $b_\mu$, the green area to the ephemeral information $r_\mu = h_\mu - b_\mu$, and the red area to the single-symbol redundancy $\rho_\mu = H[X_0] - h_\mu$. (**b**) The components of the predictable information (the excess entropy $\mathbf{E} = \sigma_\mu + b_\mu + q_\mu$ in bits) also as a function of $p$ with $q = p$. The lowest (blue) line is $q_\mu$; the middle (green) line is $q_\mu + b_\mu$, so that the green area denotes $b_\mu$'s contribution to $\mathbf{E}$. The upper (red) line is $\mathbf{E}$, so that the red area denotes elusive information $\sigma_\mu$ in $\mathbf{E}$. Note that for a large range of $p$, the co-information $q_\mu$ is (slightly) negative.

## 6. Conclusions

Stationary renewal processes are well studied, easy to define, and, in many ways, temporally simple. Given this simplicity and their long history, it is somewhat surprising that one is still able to discover new properties; in our case, by viewing them through an information-theoretic lens. Indeed, their simplicity becomes apparent in the informational and structural analyses. For instance, renewal processes are causally reversible with isomorphic $\epsilon$-machines in forward and reverse-time, *i.e.*,

temporally reversible. Applying the causal-state equivalence relation to renewal processes, however, also revealed several unanticipated subtleties. For instance, we had to delineate new subclasses of renewal process ("eventually $\Delta$-Poisson") in order to completely classify $\epsilon$-machines of renewal processes. Additionally, the informational architecture formulae in Table 1 are surprisingly complicated, since exactly calculating these informational measures requires a unifilar presentation. In Section 5, we needed an infinite-state machine to study the informational architecture of a process generated by simple two-state HMM.

Looking to the future, the new structural view of renewal processes will help improve inference methods for infinite-state processes, as it tells us what to expect in the ideal setting: what are the effective states, what are appropriate null models, how informational quantities scale, and the like. For example, Figure 3a–e gave all possible causal architectures for discrete-time stationary renewal processes. Such a classification will allow for more efficient Bayesian inference of $\epsilon$-machines of point processes, as developed in [24]. That is, we can leverage "expert" knowledge that one is seeing a renewal process to delineate the appropriate subset of model architectures and thereby avoid searching over the superexponentially large set of all HMM topologies.

The range of the results' application is much larger than that explicitly considered here. The formulae in Table 1 will be most useful for understanding renewal processes with infinite statistical complexity. For instance, [28] applies the formulae to study the divergence of the statistical complexity of continuous-time processes as the observation time scale decreases. Additionally, [29] applies these formulae to renewal processes with infinite excess entropy. In particular, there, we investigate the causal architectures of infinite-state processes that generate so-called critical phenomena: behavior with power-law temporal or spatial correlations [51]. The analysis of such critical systems often turns on having an appropriate order parameter. The statistical complexity and excess entropy are application-agnostic order parameters [52–54] that allow one to better quantify when a phase transition in stochastic processes has or has not occurred, as seen in [29]. Such critical behavior has even been implicated in early studies of human communication [55] (though, see [56,57]) and recently in neural dynamics [58] and in socially-constructed, communal knowledge systems [59].

## Author Contributions

Both authors contributed equally to conception and writing. Sarah E. Marzen performed the bulk of the calculations and numerical computation. Both authors have read and approved the final manuscript.

## Conflicts of Interest

The authors declare no conflict of interest.

## A. Causal Architecture

**Notation.** *Rather than write pasts and futures as semi-infinite sequences, we notate a past as a list of nonnegative integers [60]. The semi-infinite past $X_{:0}$ is equivalent to a list of interevent counts $N_{:0}$ and the count $N'_0$ since the last event. Similarly, the semi-infinite future $X_{0:}$ is equivalent to the count to next event $N_0 - N'_0$ and future interevent counts $N_{1:}$.*

Now, recall Lemma 1.

**Lemma 1.** The counts since the last event are prescient statistics of a discrete-time stationary renewal process.

**Proof.** This follows almost immediately from the definition of stationary renewal processes and the definition of causal states, since the random variables $N_i$ are all i.i.d.. Then:

$$\Pr(X_{0:}|X_{:0}) = \Pr(N_0 - N'_0|N'_0) \prod_{i=1}^{\infty} \Pr(N_i) \; .$$

Therefore, $\Pr(X_{0:}|X_{:0} = x_{:0}) = \Pr(X_{0:}|X_{:0} = x'_{:0})$ is equivalent to $\Pr(N_0 - N'_0|N'_0 = n_0) = \Pr(N_0 - N'_0|N'_0 = n'_0)$. Hence, the counts since the last event are prescient. □

In light of Lemma 1, we introduce new notation to efficiently refer to groups of pasts with the same count since the last event.

**Notation.** *Let $r_n^+ := \{\overleftarrow{x} : x_{-n-1:0} = 10^n\}$ for $n \in \mathbb{Z}_{\geq 0}$. Recall that $10^n = 100\cdots 00$, the sequence with $n$ 0's following a 1.*

**Remark.** *Note that $\mathcal{R}^+ = \{r_n^+\}_{n=0}^{\infty}$ is always at least a forward-time prescient rival, if not the forward-time causal states $\mathcal{S}^+$. The probability distribution over $r_n^+$ is straightforward to derive. Saying that $N'_0 = n$ means that there were $n$ 0s since the last event, so that the symbol at $X_{-n-1}$ must have been a 1. That is:*

$$\begin{aligned} \pi(r_n^+) &= \Pr(N'_0 = n) \\ &= \sum_{x \in \mathcal{A}} \Pr(N'_0 = n, X_{-n-1} = x) \\ &= \Pr(N'_0 = n, X_{-n-1} = 1) \\ &= \Pr(X_{-n-1} = 1) \Pr(X_{-n:0} = 0^n | X_{-n-1} = 1) \; . \end{aligned}$$

*Since this is a stationary process, $\Pr(X_{-n-1} = 1)$ is independent of $n$, implying:*

$$\begin{aligned} \pi(r_n^+) &\propto \Pr(X_{-n:0} = 0^n | X_{-n-1} = 1) \\ &= \sum_{m=0}^{\infty} \Pr(X_{-n:m+1} = 0^{n+m}1 | X_{-n-1} = 1) \\ &= \sum_{m=n}^{\infty} F(m) \\ &= w(n) \; . \end{aligned}$$

*We see that $\pi(r_n^+) = w(n)/Z$, with $Z$ a normalization constant that makes $\sum_{n=0}^{\infty} w(n) = \mu + 1$. Therefore:*

$$\pi(r_n^+) = \frac{w(n)}{\mu + 1} .$$

In the main text, Theorem 1 was stated with less precision so as to be comprehensible. Here, we state it with more precision, even though the meaning is obfuscated somewhat by doing so. In the proof, we still err somewhat on the side of comprehensibility, and so, one might view this proof as more of a proof sketch.

**Theorem 1** The forward-time causal states of a discrete-time stationary renewal process that is not eventually $\Delta$-Poisson are exactly $\boldsymbol{\mathcal{S}}^+ = \boldsymbol{\mathcal{R}}^+$, if $F$ has unbounded support. When the support is bounded, such that $F(n) = 0$ for all $n \geq N$, $\boldsymbol{\mathcal{S}}^+ = \{r_n^+\}_{n=0}^N$. Finally, a discrete-time eventually $\Delta$-Poisson renewal process with characteristic $(\tilde{n}, \Delta)$ has forward-time causal states:

$$\boldsymbol{\mathcal{S}}^+ = \{r_n^+\}_{n=0}^{\tilde{n}-1} \cup \{\cup_{k=0}^{\infty} r_{\tilde{n}+k\Delta+m}^+\}_{m=0}^{\Delta-1} .$$

This is a complete classification of the causal states of any persistent renewal process.

**Proof.** From the proof of Lemma 1 in this Appendix, we know that two prescient states $r_n^+$ and $r_{n'}^+$ are minimal only when:

$$\Pr(N_0 - N_0'|N_0' = n) = \Pr(N_0 - N_0'|N_0' = n') . \tag{16}$$

Since $\Pr(N_0 - N_0' = m|N_0' = n) = \Pr(N_0 = m + n)/\Pr(N_0' = n)$, $\Pr(N_0 = m + n) = F(m + n)$ and $\Pr(N_0' = n) = w(n)/(\mu + 1)$ from earlier, we find that the equivalence class condition becomes:

$$\frac{F(m + n)}{w(n)} = \frac{F(m + n')}{w(n')} , \tag{17}$$

for all $m \geq 0$.

First, note that for these conditional probabilities even to be well defined, $w(n) > 0$ and $w(n') > 0$. Hence, if $F$ has bounded support $(\max \operatorname{supp} F(n) = N)$ then the causal states do not include any $r_n^+$ for $n > N$. Furthermore, Equation (17) cannot be true for all $m \geq 0$, unless $n = n'$ for $n$ and $n' \leq N$. To see this, suppose that $n \neq n'$, but that Equation (17) holds. Then, choose $m = N + 1 - \max(n, n')$ to give $0 = F(N + 1 - |n - n'|)/w(n')$, a contradiction, unless $n = n'$.

Therefore, for all remaining cases, we can assume that $F$ in Equation (17) has unbounded support.

A little rewriting makes the connection between Equation (17) and an eventually $\Delta$-Poisson process clearer. First, we choose $m = 0$ to find:

$$\frac{F(n)}{w(n)} = \frac{F(n')}{w(n')},$$

which we can use to rewrite Equation (17) as:

$$\frac{F(m + n)}{F(n)} = \frac{F(m + n')}{F(n')},$$

or more usefully:

$$F(n' + m) = \frac{F(n')}{F(n)} F(n + m).$$

A particularly compact way of rewriting this is to define $\Delta' := n' - n$, which gives $F(n' + m) = F((n + m) + \Delta')$. In this form, it is clear that the above equation is a recurrence relation on $F$ in steps of $\Delta'$, so that we can write:

$$F((n + m) + k\Delta') = \left( \frac{F(n')}{F(n)} \right)^k F(n + m) . \tag{18}$$

This must be true for every $m \geq 0$. Importantly, since $w(n) = \sum_{m=n}^{\infty} F(m)$, satisfying this recurrence relation is equivalent to satisfying Equation (17). However, Equation (18) is just the definition of an eventually $\Delta$-Poisson process in disguise; relabel with $\lambda := F(n')/F(n)$, $\tilde{n} := n$ and $\Delta = \Delta'$.

Therefore, if Equation (17) does not hold for any pair $n \neq n'$, the process is not eventually $\Delta$-Poisson, and the prescient states identified in Lemma 1 are minimal; *i.e.*, they are the causal states.

If Equation (17) does hold for some $n \neq n'$, choose the minimal such $n$ and $n'$. The renewal process is eventually $\Delta$-Poisson with characterization $\Delta = n' - n$ and $\tilde{n}$. Additionally, $F(\tilde{n} + m)/w(\tilde{n} + m) = F(\tilde{n} + m')/w(\tilde{n} + m')$ implies that $m \equiv m' \mod \Delta$, since otherwise, the $n$ and $n'$ chosen would not be minimal. Hence, the causal states are exactly those given in the theorem's statement. $\square$

**Remark.** *For the resulting $F(n)$ to be a valid interevent distribution, $\lambda = F(\tilde{n} + \Delta)/F(\tilde{n}) < 1$, as normalization implies:*

$$\sum_{n=0}^{\tilde{n}-1} F(n) + \sum_{n=\tilde{n}}^{\tilde{n}+\Delta-1} \frac{F(n)}{1 - \lambda} = 1 .$$

**Notation.** *Let us denote $\mathcal{S}^+ = \{\sigma_n^+ := r_n^+\}_{n=0}^{\infty}$ for a renewal process that is not eventually $\Delta$-Poisson, $\mathcal{S}^+ = \{\sigma_n^+ := r_n^+\}_{n=0}^{\tilde{n}}$ for an eventually $\Delta$-Poisson renewal process with bounded support, and $\mathcal{S}^+ = \{\sigma_n^+ := r_n^+\}_{n=0}^{\tilde{n}-1} \cup \{\sigma_{\tilde{n}+m}^+ := \cup_{k=0}^{\infty} r_{\tilde{n}+k\Delta+m}^+\}_{m=0}^{\Delta-1}$ for an eventually $\Delta$-Poisson process.*

The probability distribution over these forward-time causal states is straightforward to derive from $\pi(r_n^+) = w(n)/(\mu + 1)$. Therefore, for a renewal process that is not eventually $\Delta$-Poisson or one that is with bounded support, $\pi(\sigma_n^+) = w(n)/(\mu + 1)$. (For the latter, $n$ only runs from zero to $\tilde{n}$.) For an eventually $\Delta$-Poisson renewal process $\pi(\sigma_n^+) = w(n)/(\mu + 1)$ when $n < \tilde{n}$ and:

$$\pi(\sigma_n^+) = \sum_{k=0}^{\infty} \pi(r_{n+k\Delta}^+)$$
$$= \frac{\sum_{k=0}^{\infty} w(n + k\Delta)}{\mu + 1} ,$$

when $\tilde{n} \leq n < \tilde{n} + \Delta$. Therefore, the statistical complexity given in Table 1 follows from $C_\mu^+ = H[\mathcal{S}^+]$.

Recall Lemma 2 and Theorem 2.

**Lemma 2.** Groupings of futures with the same counts to next event are reverse-time prescient statistics for discrete-time stationary renewal processes.

**Theorem 2.** (a) The reverse-time causal states of a discrete-time stationary renewal process that is not eventually $\Delta$-Poisson are groupings of futures with the same count to the next event up until and

including $N$, if $N$ is finite. (b) The reverse-time causal states of a discrete-time eventually $\Delta$-Poisson stationary renewal process are groupings of futures with the same count to next event up until $\tilde{n}$, plus groupings of futures whose counts since the last event $n$ are in the same equivalence class as $\tilde{n}$ modulo $\Delta$.

**Proof.** The proof for both claims relies on a single fact: in reverse-time, a stationary renewal process is still a stationary renewal process with the same interevent count distribution. The lemma and theorem therefore follow from Lemma 1 and Theorem 1. $\square$

Since the forward and reverse-time causal states are the same with the same future conditional probability distribution, we have $C_\mu^+ = C_\mu^-$, and the causal irreversibility vanishes: $\Xi = 0$.

Transition probabilities can be derived for both the renewal process's prescient states and its $\epsilon$-machine as follows. For the prescient machine, if a $0$ is observed when in $r_n^+$, we transition to $r_{n+1}^+$; else, we transition to $r_0^+$, since we just saw an event. Basic calculations show that these transition probabilities are:

$$
\begin{aligned}
T_{r_n^+ r_m^+}^{(x)} &= \Pr(\mathcal{R}_{t+1}^+ = r_m^+, X_{t+1} = x | \mathcal{R}_t^+ = r_n^+) \\
&= \frac{F(n)}{w(n)} \delta_{m,0} \delta_{x,1} + \frac{w(n+1)}{w(n)} \delta_{m,n+1} \delta_{x,0} \, .
\end{aligned}
$$

Not only do these specify the prescient machine transition dynamic, but due to the close correspondence between prescient and causal states, they also automatically give the $\epsilon$-machine transition dynamic:

$$
\begin{aligned}
T_{\sigma\sigma'}^{(x)} &= \Pr(\mathcal{S}_{t+1}^+ = \sigma', X_{t+1} = x | \mathcal{S}_t^+ = \sigma) \\
&= \sum_{r,r' \in \mathcal{R}^+} T_{r' \to r}^{(x)} \Pr(\mathcal{S}_{t+1}^+ = \sigma' | \mathcal{R}_{t+1}^+ = r) \times \Pr(\mathcal{R}_t^+ = r' | \mathcal{S}_t^+ = \sigma) \, .
\end{aligned}
$$

## B. Information Architecture

It is straightforward to show that $\Pr(X_0 = 0) = \frac{1}{\mu+1}$, and thus:

$$
H[X_0] = -\frac{1}{\mu+1} \log_2 \frac{1}{\mu+1} - \left(1 - \frac{1}{\mu+1}\right) \log_2 \left(1 - \frac{1}{\mu+1}\right) \, .
$$

The entropy rate is readily calculated from the prescient machine:

$$
\begin{aligned}
h_\mu &= \sum_{n=0}^{\infty} H[X_{t+1} | \mathcal{R}_t^+ = r_n^+] \pi(r_n^+) \\
&= -\sum_{n=0}^{\infty} \frac{w(n)}{\mu+1} \left(\frac{F(n)}{w(n)} \log_2 \frac{F(n)}{w(n)} + \frac{w(n+1)}{w(n)} \log_2 \frac{w(n+1)}{w(n)}\right) \, .
\end{aligned}
$$

Additionally, after some algebra, this simplifies to:

$$
h_\mu = -\frac{1}{\mu+1} \sum_{n=0}^{\infty} F(n) \log_2 F(n) \, ,
$$

once we recognize that $w(0) = 1$, and so $w(0) \log_2 w(0) = 0$, and we recall that $w(n+1)+F(n) = w(n)$. The excess entropy, being the mutual information between forward and reverse-time prescient states, is [37,46]:

$$\mathbf{E} = I[\mathcal{R}^+; \mathcal{R}^-]$$
$$= H[\mathcal{R}^+] - H[\mathcal{R}^+|\mathcal{R}^-] .$$

Therefore, to calculate, we note that:

$$\Pr(r_n^+, r_m^-) = \frac{F(m+n)}{\mu+1}$$
$$\Pr(r_n^+|r_m^-) = \frac{F(n+m)}{w(m)} .$$

After some algebra, we find that:

$$H[\mathcal{R}^+] = -\sum_{n=0}^{\infty} \frac{w(n)}{\mu+1} \log_2 \frac{w(n)}{\mu+1}$$

and that:

$$H[\mathcal{R}^+|\mathcal{R}^-] = -\sum_{m,n=0}^{\infty} \frac{F(n+m)}{\mu+1} \log_2 \frac{F(n+m)}{w(m)}$$
$$= -\sum_{m=0}^{\infty} \frac{m+1}{\mu+1} F(m) \log_2 \frac{F(m)}{\mu+1} + \sum_{m=0}^{\infty} \frac{w(m)}{\mu+1} \log_2 \frac{w(m)}{\mu+1} .$$

The above quantity is the forward crypticity $\chi^+$ [37] when the renewal process is not eventually $\Delta$-Poisson. These together imply:

$$\mathbf{E} = -2\sum_{n=0}^{\infty} \frac{w(n)}{\mu+1} \log_2 \frac{w(n)}{\mu+1} + \sum_{m=0}^{\infty} (m+1) \frac{F(m)}{\mu+1} \log_2 \frac{F(m)}{\mu+1} .$$

Finally, the bound information $b_\mu$ is:

$$b_\mu = I[X_{1:}; X_0|X_{:0}]$$
$$= I[\mathcal{R}_1^-; X_0|\mathcal{R}_0^+]$$
$$= H[\mathcal{R}_1^-|\mathcal{R}_0^+] - H[\mathcal{R}_1^-|\mathcal{R}_1^+] ,$$

where we used the causal shielding properties of prescient states, $X_{:0} \to \mathcal{R}_0^+ \to \mathcal{R}_1^- \to X_{1:}$, and the unifilarity of the prescient machines, as shown in Figure 3a–d. While we already calculated $H[\mathcal{R}_1^-|\mathcal{R}_1^+]$,

we still need to calculate $H[\mathcal{R}_1^-|\mathcal{R}_0^+]$. We do so using the prescient machine's transition dynamic. In particular:

$$
\begin{aligned}
\Pr(\mathcal{R}_1^- = n|\mathcal{R}_0^+ = m) &= \sum_{r \in \mathcal{R}^+} \Pr(\mathcal{R}_1^- = n|\mathcal{R}_1^+ = r)\Pr(\mathcal{R}_1^+ = r|\mathcal{R}_0^+ = m) \\
&= \frac{F(m+n+1) + F(n)F(m)}{w(m)} \,.
\end{aligned}
$$

Where we omit details getting to the last line. Eventually, the calculation yields:

$$
b_\mu = \frac{\sum_{n=0}^\infty (n+1)F(n)\log_2 F(n)}{\mu+1} - \frac{\sum_{m,n=0}^\infty g(m,n)\log_2 g(m,n)}{\mu+1} \,,
$$

where:

$$
g(m,n) = F(m+n+1) + F(n)F(m) \,.
$$

From the expressions above, we immediately solve for $r_\mu = h_\mu - b_\mu$, $q_\mu = H[X_0] - h_\mu - b_\mu$, and $\sigma_\mu = \mathbf{E} - q_\mu$; thereby laying out the information architecture of stationary renewal processes.

Finally, we calculate the finite-time predictable information $\mathbf{E}(M, N)$ as the mutual information between finite-time forward and reverse-time prescient states:

$$
\mathbf{E}(M, N) = H[\mathcal{R}^{-N}] - H[\mathcal{R}^{-N}|\mathcal{R}^{+M}] \,. \tag{19}
$$

Recall Corollary 1.

**Corollary 1.** Forward-time (and reverse-time) finite-time $M$ prescient states of a discrete-time stationary renewal process are the counts from (and to) the next event up until and including $M$.

**Proof.** From Lemmas 1 and 2, we know that counts from (to) the last (next) event are prescient forward-time (reverse-time) statistics. If our window on pasts (futures) is $M$, then we cannot distinguish between counts since (to) the last (next) event that are $M$ and larger. Hence, the finite-time $M$ prescient statistics are the counts from (and to) the next event up until and including $M$, where a finite-time $M$ prescient state includes all pasts with $M$ or more counts from (to) the last (next) event. $\square$

To calculate $\mathbf{E}(M, N)$, we find $\Pr(\mathcal{R}^{+M}, \mathcal{R}^{-N})$ by marginalizing $\Pr(\mathcal{R}^+, \mathcal{R}^-)$. For ease of notation, we first define a function:

$$
u(m) = \sum_{n=m}^\infty w(n) \,.
$$

Algebra not shown here yields:

$$\mathbf{E}(M, N) = H[\mathcal{S}^{-N}] - H[\mathcal{S}^{-N}|\mathcal{S}^{+M}]$$

$$= \log_2(\mu + 1) - \frac{\sum_{n=0}^{N-1} w(n) \log_2 w(n)}{\mu + 1}$$

$$- \frac{\sum_{m=0}^{M-1} w(m) \log_2 w(m)}{\mu + 1}$$

$$+ \frac{\sum_{n=M}^{N+M-1} w(n) \log_2 w(n)}{\mu + 1}$$

$$+ \frac{\sum_{n=N}^{N+M-1} w(n) \log_2 w(n)}{\mu + 1}$$

$$- \frac{u(N) \log_2 u(N) + u(M) \log_2 u(M)}{\mu + 1}$$

$$+ \frac{u(N + M) \log_2 u(N + M)}{\mu + 1}$$

$$+ \frac{\sum_{m=0}^{M-1} \sum_{n=m}^{N+m-1} F(n) \log_2 F(n)}{\mu + 1} .$$

Two cases of interest are equal windows ($N = M$) and semi-infinite pasts ($M \to \infty$). In the former, we find:

$$\mathbf{E}(M, M) = \log_2(\mu + 1) - \frac{2 \sum_{m=0}^{M-1} w(m) \log_2 w(m)}{\mu + 1}$$

$$+ \frac{2 \sum_{m=M}^{2M-1} w(m) \log_2 w(m)}{\mu + 1}$$

$$- \frac{2u(M) \log_2 u(M)}{\mu + 1} + \frac{u(2M) \log_2 u(2M)}{\mu + 1}$$

$$+ \frac{\sum_{m=0}^{M-1} \sum_{n=m}^{M+m-1} F(n) \log_2 F(n)}{\mu + 1} .$$

In the latter case of semi-infinite pasts, several terms vanish, and we have:

$$\mathbf{E}(N) = \log_2(\mu + 1) - \frac{2 \sum_{n=0}^{N-1} w(n) \log_2 w(n)}{\mu + 1}$$

$$- \frac{u(N) \log_2 u(N)}{\mu + 1}$$

$$+ \frac{N \sum_{n=N}^{\infty} F(n) \log_2 F(n)}{\mu + 1}$$

$$+ \frac{\sum_{n=0}^{N-1} (n + 1) F(n) \log_2 F(n)}{\mu + 1} .$$

## References

1. Smith, W.L. Renewal Theory and Its Ramifications. *J. R. Stat. Soc. B* **1958**, *20*, 243–302.

2. Gerstner, W.; Kistler, W.M. Statistics of spike trains. In *Spiking Neuron Models: Single Neurons, Populations, Plasticity*; Cambridge University Press: Cambridge, UK, 2002; Chapter 5.2.

3. Beichelt, F. *Stochastic Processes in Science, Engineering and Finance*; Chapman and Hall: New York, NY, USA, 2006.

4. Barbu, V.S.; Limnios, N. *Semi-Markov Chains and Hidden Semi-Markov Models toward Applications: Their Use in Reliability and DNA Analysis*; Springer: Berlin/Heidelberg, Germany, 2008; Volume 191.

5. Lowen, S.B.; Teich, M.C. Fractal Renewal Processes Generate 1/f Noise. *Phys. Rev. E* **1993**, *47*, 992–1001.

6. Lowen, S.B.; Teich, M.C. Fractal Renewal Processes. *IEEE Trans. Inf. Theory* **1993**, *39*, 1669–1671.

7. Cakir, R.; Grigolini, P.; Krokhin, A.A. Dynamical Origin of Memory and Renewal. *Phys. Rev. E* **2006**, *74*, 021108.

8. Akimoto, T.; Hasumi, T.; Aizawa, Y. Characterization of intermittency in renewal processes: Application to earthquakes. *Phys. Rev. E* **2010**, *81*, 031133.

9. Montero, M.; Villarroel, J. Monotonic Continuous-time Random Walks with Drift and Stochastic Reset Events. *Phys. Rev. E* **2013**, *87*, 012116.

10. Bologna, M.; West, B.J.; Grigolini, P. Renewal and Memory Origin of Anomalous Diffusion: A discussion of their joint action. *Phys. Rev. E* **2013**, *88*, 062106.

11. Bianco, S.; Ignaccolo, M.; Rider, M.S.; Ross, M.J.; Winsor, P.; Grigolini, P. Brain, Music, and Non-Poisson Renewal Processes. *Phys. Rev. E* **2007**, *75*, 061911.

12. Li, C.B.; Yang, H.; Komatsuzaki, T. Multiscale Complex Network of Protein Conformational Fluctuations in Single-Molecule Time Series. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 536–541.

13. Kelly, D.; Dillingham, M.; Hudson, A.; Wiesner, K. A New Method for Inferring Hidden Markov Models from Noisy Time Sequences. *PLoS ONE* **2012**, *7*, e29703.

14. Onaga, T.; Shinomoto, S. Bursting Transition in a Linear Self-exciting Point Process. *Phys. Rev. E* **2014**, *89*, 042817.

15. Valenza, G.; Citi, L.; Scilingo, E.P.; Barbieri, R. Inhomogeneous Point-Process Entropy: An Instantaneous measure of complexity in discrete systems. *Phys. Rev. E* **2014**, *89*, 052803.

16. Shalizi, C.R.; Crutchfield, J.P. Information Bottlenecks, Causal States, and Statistical Relevance Bases: How to Represent Relevant Information in Memoryless Transduction. *Adv. Complex Syst.* **2002**, *5*, 91–95.

17. Still, S.; Crutchfield, J.P. Structure or Noise? **2007**, arXiv:0708.0654.

18. Still, S.; Crutchfield, J.P.; Ellison, C.J. Optimal Causal Inference: Estimating Stored Information and Approximating Causal Architecture. *Chaos* **2010**, *20*, 037111.

19. Marzen, S.; Crutchfield, J.P. Information Anatomy of Stochastic Equilibria. *Entropy* **2014**, *16*, 4713–4748.

20. Martius, G.; Der, R.; Ay, N. Information driven self-organization of complex robotics behaviors. *PLoS ONE* **2013**, *8*, e63400.

21. Crutchfield, J.P.; Feldman, D.P. Regularities Unseen, Randomness Observed: Levels of Entropy Convergence. *Chaos* **2003**, *13*, 25–54.

22. Debowski, L. On Hidden Markov Processes with Infinite Excess Entropy. *J. Theor. Prob.* **2012**, *27*, 539–551.

23. Travers, N.; Crutchfield, J.P. Infinite Excess Entropy Processes with Countable-State Generators. *Entropy* **2014**, *16*, 1396–1413.

24. Strelioff, C.C.; Crutchfield, J.P. Bayesian Structural Inference for Hidden Processes. *Phys. Rev. E* **2014**, *89*, 042119.

25. Crutchfield, J.P.; Riechers, P.; Ellison, C.J. Exact Complexity: Spectral Decomposition of Intrinsic Computation. **2013**, arXiv:1309.3792 [cond-mat.stat-mech].

26. Blackwell, D. The Entropy of Functions of Finite-state Markov Chains. In *Transactions of the first Prague Conference on Information Theory, Statistical Decision Functions, Random Processes: Held at Liblice near Prague, from November 28 to 30, 1956*; Publishing House of the Czechoslovak Academy of Sciences: Prague, Czech Republic, 1957; pp. 13–20.

27. Crutchfield, J.P. Between Order and Chaos. *Nat. Phys.* **2012**, *8*, 17–24.

28. Marzen, S.E.; DeWeese, M.R.; Crutchfield, J.P. Time Resolution Dependence of Information Measures for Spiking Neurons: Atoms, Scaling and Universality. **2015**, arXiv:1504.04756.

29. Marzen, S.E.; Crutchfield, J.P. Long-Range Memory in Stationary Renewal Processes. **2015**, in preparation.

30. Ephraim, Y.; Merhav, N. Hidden Markov Processes. *IEEE Trans. Inf. Theory* **2002**, *48*, 1518–1569.

31. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; Wiley: New York, NY, USA, 2006.

32. Crutchfield, J.P.; Young, K. Inferring Statistical Complexity. *Phys. Rev. Let.* **1989**, *63*, 105–108.

33. Shalizi, C.R.; Crutchfield, J.P. Computational Mechanics: Pattern and Prediction, Structure and Simplicity. *J. Stat. Phys.* **2001**, *104*, 817–879.

34. Paz, A. *Introduction to Probabilistic Automata*; Academic Press: New York, NY, USA, 1971.

35. Rabiner, L.R.; Juang, B.H. An Introduction to Hidden Markov Models. *IEEE ASSP Mag.* **1986**, *3*, 4–16.

36. Rabiner, L.R. A Tutorial on Hidden Markov Models and Selected Applications. *IEEE Proc.* **1989**, *77*, 257.

37. Crutchfield, J.P.; Ellison, C.J.; Mahoney, J.R. Time's Barbed Arrow: Irreversibility, Crypticity, and Stored Information. *Phys. Rev. Lett.* **2009**, *103*, 094101.

38. Ellison, C.J.; Mahoney, J.R.; James, R.G.; Crutchfield, J.P.; Reichardt, J. Information Symmetries in Irreversible Processes. *Chaos* **2011**, *21*, 037107.

39. Yeung, R.W. *Information Theory and Network Coding*; Springer: New York, NY, USA, 2008.

40. James, R.G.; Ellison, C.J.; Crutchfield, J.P. Anatomy of a Bit: Information in a Time Series Observation. *Chaos* **2011**, *21*, 037109.

41. Bell, A.J. The Co-information Lattice. In Proceedings of the Fourth International Symposium on Independent Component Analysis and Blind Signal Separation, Nara, Japan, 1–4 April 2003; Amari, S.-i., Cichocki, A., Makino, S., Murata, N., Eds.; Springer: New York, NY, USA, 2003; pp. 921–926.

42. Ara, P.M.; James, R.G.; Crutchfield, J.P. The Elusive Present: Hidden Past and Future Correlation and Why We Build Models. **2015**, arXiv:1507.00672.

43. Crutchfield, J.P.; Young, K. Computation at the Onset of Chaos. In *Entropy, Complexity, and the Physics of Information*; *SFI Studies in the Sciences of Complexity*; Zurek, W., Ed.; Addison-Wesley: Reading, MA, USA, 1990; Volume VIII, pp. 223–269.

44. Bialek, W.; Nemenman, I.; Tishby, N. Predictability, Complexity, and Learning. *Neural Comp.* **2001**, *13*, 2409–2463.

45. Crutchfield, J.P. The Calculi of Emergence: Computation, Dynamics, and Induction. *Physica D* **1994**, *75*, 11–54.

46. Ellison, C.J.; Mahoney, J.R.; Crutchfield, J.P. Prediction, Retrodiction, and the Amount of Information Stored in the Present. *J. Stat. Phys.* **2009**, *136*, 1005–1034.

47. Godreche, C.; Majumdar, S.N.; Schehr, G. Statistics of the Longest Interval in Renewal Processes. *J. Stat. Mech.* **2015**, P03014, doi:10.1088/1742-5468/2015/03/P03014.

48. Birch, J.J. Approximations for the Entropy for Functions of Markov Chains. *Ann. Math. Stat.* **1962**, *33*, 930–938.

49. Riechers, P.M.; Varn, D.P.; Crutchfield, J.P. Pairwise Correlations in Layered Close-Packed Structures. *Acta Cryst. A* **2015**, *71*, 423–443.

50. Still, S.; Ellison, C.J.; Crutchfield, J.P. Optimal Causal Inference: Estimating stored information and approximating causal architecture. *Chaos* **2010**, *20*, 037111.

51. Crutchfield, J.P. Critical Computation, Phase Transitions, and Hierarchical Learning. In *Towards the Harnessing of Chaos*; Yamaguti, M., Ed.; Elsevier Science: Amsterdam, The Netherlands, 1994; pp. 29–46.

52. Crutchfield, J.P.; Feldman, D.P. Statistical Complexity of Simple One-Dimensional Spin Systems. *Phys. Rev. E* **1997**, *55*, R1239–R1243.

53. Feldman, D.P.; Crutchfield, J.P. Structural Information in Two-Dimensional Patterns: Entropy Convergence and Excess Entropy. *Phys. Rev. E* **2003**, *67*, 051103.

54. Tchernookov, M.; Nemenman, I. Predictive Information in a Nonequilibrium Critical Model. *J. Stat. Phys.* **2013**, *153*, 442–459.

55. Zipf, G.K. *The Psycho-Biology of Language: An Introduction to Dynamic Philology*; Houghton Mifflin Company: Boston, MA, USA, 1935.

56. Mandelbrot, B. An informational theory of the statistical structure of languages. In *Communication Theory*; Jackson, W., Ed.; Butterworths: London, UK, 1953; pp. 486–502.

57. Miller, G.A. Some effects of intermittent silence. *Am. J. Psychiatry* **1957**, *70*, 311–314.

58. Beggs, J.M.; Plenz, D. Neuronal Avalanches in Neocortical Circuits. *J. Nuerosci.* **2003**, *23*, 11167–11177.

59. Dedeo, S. Collective Phenomena and Non-Finite State Computation in a Human Social System. *PLoS ONE* **2013**, *8*, e75808.

60. Cessac, B.; Cofre, R. Spike Train statistics and Gibbs Distributions. *J. Physiol. Paris* **2013**, *107*, 360–368