Informational and Causal Architecture of Discrete-Time Renewal Processes

Sarah Marzen
James P. Crutchfield

SFI WORKING PAPER: 2014-08-032

SFI Working Papers contain accounts of scientific work of the author(s) and do not necessarily represent the views of the Santa Fe Institute. We accept papers intended for publication in peer-reviewed journals or proceedings volumes, but not papers that have already appeared in print. Except for papers by our external faculty, papers must be based on work done at SFI, inspired by an invited visit to or collaboration at SFI, or funded by an SFI grant.

©NOTICE: This working paper is included by permission of the contributing author(s) as a means to ensure timely distribution of the scholarly and technical work on a non-commercial basis. Copyright and all rights therein are maintained by the author(s). It is understood that all persons copying this information will adhere to the terms and constraints invoked by each author's copyright. These works may be reposted only with the explicit permission of the copyright holder.

www.santafe.edu
Informational and Causal Architecture of Discrete-Time Renewal Processes

Sarah Marzen\textsuperscript{1,*} and James P. Crutchfield\textsuperscript{2,†}

\textsuperscript{1}Redwood Center for Theoretical Neuroscience and Department of Physics, University of California at Berkeley, Berkeley, CA 94720-5800
\textsuperscript{2}Complexity Sciences Center and Department of Physics, University of California at Davis, One Shields Avenue, Davis, CA 95616

(Dated: August 28, 2014)

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 apply our new formulae for information measures to analyze the output of the parametrized simple nonunifilar source, a simple two-state machine with an infinite-state $\epsilon$-machine presentation. All in all, the results lay the groundwork for analyzing processes with divergent statistical complexity and divergent excess entropy.

\textbf{Keywords:} stationary renewal process, statistical complexity, predictable information, information anatomy, entropy rate

PACS numbers: 02.50.-r 89.70.+c 05.45.Tp 02.50.Ey 02.50.Ga

\section{I. INTRODUCTION}

Stationary renewal processes are widely used, analytically tractable, compact models of an important class of point processes \cite{1,2,3}. 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, memoryful behavior \cite{4,5}. Additionally, understanding more complicated processes \cite{6,7} requires fully analyzing renewal processes.

By studying them in-depth from a structural information viewpoint, we gain a new understanding of statistical signatures of apparent high complexity. To that end, we derive the causal-state minimal sufficient statistics—the $\epsilon$-machine—for renewal processes and then derive new formulae for their informational architecture in terms the interevent count distribution. The informational architecture here is a set of quantitative measures of a renewal process’ predictability, difficulty of prediction, inherent randomness, and Markovity, and the like, including:

\begin{itemize}
  \item the \textbf{statistical complexity} $C_\mu$, which quantifies the historical memory that must be stored in order to predict a process’s future as well as possible given its past;
  \item the \textbf{entropy rate} $h_\mu$, which quantifies a process’ inherent randomness as the uncertainty in the next symbol even given that we can predict as well as possible;
  \item the \textbf{excess entropy} $E$, which quantifies how much of a process’s future is predictable in terms of the mutual information between its past and future;
  \item the \textbf{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 symbol shared with the future, above and beyond that of the entire past; and
  \item the \textbf{enigmatic information} $\sigma_\mu$, which quantifies a process’s deviation from Markovity as the mutual information between the past and future conditioned on the present.
\end{itemize}

Analyzing a process’ information architecture in this way gives a detailed understanding of its structure and stochasticity. While it is certainly possible to estimate these quantities numerically, numerical methods are likely to fail precisely when the renewal process has large or divergent excess entropy; i.e., when the renewal process is complex. The closed-form formulae we develop for the information measures are particularly insightful in these cases, and allow us to examine the limit of infinitesimal time resolution in Ref. \cite{8} and conditions under which renewal processes can have divergent...
excess entropy in Ref. [9]. Beyond this, the informational and causal architectures are key to finding limits to a process’s optimal lossy predictive features [10][13], give insight into designing action policies for intelligent autonomous agents [14], and quantify whether or not a given process has one or another kind of infinite memory [15][17].

The development is organized as follows. Section II provides a quick introduction to computational mechanics and complexity measures for stationary time series. Section III identifies the causal states (in both forward and reverse time), the statistical complexity, and finally recall several information-theoretic measures designed to capture organization in process structure, and finally recall several information-theoretic measures designed to capture organization in structured processes.

A. 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 equation. The latter are important in applications, of course, as they are generators of behavior, as we will see in a later section.

As a result, our main object of study is a process \( P \): 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 joint probabilities \( \Pr(\ldots X_{-2}, X_{-1}, X_0, X_1, \ldots) \). 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 \( X_{-2} \cdots X_{-1} \) leading up the present, and the future is the chain following the present \( X_L = X_1 X_2 \cdots \). Finally, we assume a process is ergodic and stationary—\( \Pr(X_{0:L}) = \Pr(X_{L+1}) \) for all \( t \in \mathbb{Z} \)—and the measurement symbols \( x_t \) range over a finite alphabet: \( x \in \mathcal{A} \). We make no assumption that the symbols represent a process’s states—they are at best an indirect reflection of an internal Markov mechanism; these processes are hidden Markov processes [18].

Discrete-time stationary renewal processes have binary observation alphabets \( \mathcal{A} = \{0, 1\} \). Observation of the binary symbol 1 is called an event. The number of 0’s between successive 1’s 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 normalizable—\( \sum_{n=0}^{\infty} F(n) = 1 \). 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 \( \sum_{n=0}^{\infty} w(n) = \mu + 1 \).

B. Information

The information or uncertainty in a process is often defined as the Shannon entropy \( H[X_0] \) of a single symbol \( X_0 \) [19]:

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

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

\[
H(L) = H[X_{0:L}] = - \sum_{x \in \mathcal{A}} \Pr(X_{0:L} = x) \log_2 \Pr(X_{0:L} = x) .
\]

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. It measures 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:

\[
h_\mu = \lim_{L \to \infty} H[X_0|X_{-L:0}] .
\]
process's degree of unpredictability.

C. Causal architecture

Forward-time causal states $S^+$ are minimal sufficient statistics for predicting a process’s future [20, 21]. This follows from their definition as sets of pasts grouped by the equivalence relation $\sim^+$:

$$x_0 \sim^+ x'_0$$
$$\iff \Pr(X_0 | X_0 = x_0) = \Pr(X_0 | X_0 = x'_0).$$  (3)

As a shorthand, we denote a cluster of pasts so defined, a causal state, as $\sigma^+ \in S^+$. Each state $\sigma^+$ inherits a probability $\pi(\sigma^+)$ from the process’s probability over pasts $\Pr(X_0)$. The forward-time statistical complexity is defined as the Shannon entropy of the probability distribution over forward-time causal states [20]:

$$C^+ = H[S^+].$$  (4)

A generative model—the process’s $\epsilon$-machine—is built out of the causal states by endowing the state set with a transition probability matrix:

$$T_{\sigma \sigma'} = \Pr(S_{t+1}^+ = \sigma', X_t = x | S_t^+ = \sigma),$$

that gives the probability of generating the next symbol $x_t$ and ending in the next state $\sigma_{t+1}$, if starting in state $\sigma_t$. (Since output symbols are generated during transitions there is, in effect, a half time-step difference in index. We suppress notating this.) For a discrete-time, discrete-alphabet process, the $\epsilon$-machine is its minimal unifilar Hidden Markov Model (HMM) [20, 21]. (For general background on HMMs see [22–24].) Note that the causal state set can be finite, countable, or uncountable. Minimality can be defined by either the smallest number of states or the smallest statistical complexity $C^+$. Unifilarity is a constraint on the transition matrices such that the next state $\sigma_{t+1}$ is determined by knowing the current state $\sigma_t$ and the next symbol $x_t$.

A similar equivalence relation can be applied to find minimal sufficient statistics for retrodiction [25]. Futures are grouped together if they have equivalent conditional probability distributions over pasts:

$$x_0 \sim^+ x'_0$$
$$\iff \Pr(X_0 | X_0 = x_0) = \Pr(X_0 | X_0 = x'_0).$$  (5)

A cluster of futures—a reverse-time causal state—defined by $\sim^-$ is denoted $\sigma^- \in S^-$. Again, each $\sigma^-$ inherits a probability $\pi(\sigma^-)$ from the probability over futures $\Pr(X_0)$. And, the reverse-time statistical complexity is

the Shannon entropy of the probability distribution over reverse-time causal states:

$$C^- = H[S^-].$$  (6)

In general, the forward and reverse-time statistical complexities are not equal [25, 26]. That is, different amounts of information must be stored from the past (future) to predict (retrodict). Their difference $\Xi = C^+ - C^-$ is a process’s causal irreversibility and it reflects this statistical asymmetry.

D. Informational architecture

Shannon’s various information quantities—entropy, conditional entropy, mutual information, and the like—when applied to time series are functions of the joint distributions $\Pr(X_0, X_1)$. Importantly, they define an algebra of information measures for a given set of random variables [27]. Ref. [28] 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:

$$r_\mu = H[X_0; X_0, X_1],$$  (7)

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

$$b_\mu = I[X_0; X_1; X_0],$$  (8)

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

$$q_\mu = I[X_0; X_0; X_1],$$  (9)

which is the three-way mutual information between past, present, and future.

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

$$b_\mu = I[X_0; X_0 | X_1].$$  (10)

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

$$\sigma_\mu = I[X_0; X_1 | X_0].$$  (11)

This is called the elusive information [29]. It measures the amount of past-future correlation not contained in the present. It is nonzero if the process has hidden states and is therefore quite sensitive to how the state space
observed or coarse grained.

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

$$E = I[X_0; X_\infty].$$

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

$$E = b_\mu + \sigma_\mu + q_\mu. \quad (12)$$

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

$$h_\mu = H[X_0|X_{-\infty}] = r_\mu + b_\mu. \quad (13)$$

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 anatomy of a single measurement (Eq. (1)):

$$H[X_0] = q_\mu + 2b_\mu + r_\mu. \quad (13)$$

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

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

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

If $E$ is finite, then $E = \lim_{M,N \to \infty} E(M, N)$. When $E$ is infinite, then the way in which $E(M, N)$ diverges is one measure of a process’ complexity [15 30 31]. An analogous, finite past-future $(M, N)$-parametrized equivalence relation leads to finite-time causal states and statistical complexity $C_\mu(M, N)$.

III. CAUSAL ARCHITECTURE OF RENEWAL PROCESSES

We are now ready to develop the computational mechanics of discrete-time renewal processes. To aid readability we sequester most all of the detailed calculations and proofs in App. 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 $R$ is a function of the past such that:

$$H[X_0|X_0] = H[X_0|R].$$

Recall that causal states can be written as unions of prescient states [21]. 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. Knowing the past of a homogeneous Poisson process does not help one predict the future, because there is nothing to predict. We would therefore expect the prescient states in Lemma 1 to fail to be causal states precisely when the interevent distribution was reminiscent of a Poisson renewal process. This intuition is made precise by Def. 2.

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

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

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 $(\hat{n}, \Delta)$ eventually $\Delta$-Poisson has an interevent distribution that is $\Delta$-Poisson for all $n \geq \hat{n}$:

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

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 $\hat{n}$, then we choose the smallest possible $\Delta$ and the smallest possible $\hat{n}$.

A $\Delta$-Poisson process with $\Delta = 1$ is 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 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 last event up until $\hat{n}$ and pasts whose count $n$ since last event are in the same equivalence class as $\hat{n}$ modulo $\Delta$.

These statements are made more precisely in App. A. The main result is that causal states are sensitive to
two features: eventually Δ-Poisson structure in the interevent distribution and 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 Δ-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 \( \mu \), since the probability of observing at least \( n \) counts since last event is \( w(n) \). Hence, the probability of seeing \( n \) counts since the last event is simply the normalized survival function \( \frac{w(n)}{\mu + 1} \). Appendix A derives the statistical complexity using this and Theorem 1. The resulting formulae are given in Table I for the two cases.

As described in Sec. 1 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 to its boundedness. For instance, the probability of observing an event given that it has been \( n \) counts since the last event is \( \frac{F(n)}{w(n)} \). 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 App. A for proof details.

Figures 1-4 display the state-transition diagrams for the \( \epsilon \)-machines in the various cases delineated. Figure 1 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 2 is the \( \epsilon \)-machine of a renewal process whose interevent distribution never has exponential tails but which cannot have arbitrarily large interevent counts. The \( \epsilon \)-machine in Fig. 3 looks quite similar to the \( \epsilon \)-machine in Fig. 2 but it has an additional transition that connects the last state \( \tilde{n} \) to itself. This added transition changes our 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 Fig. 4 represents an eventually Δ-Poisson process with \( \Delta > 1 \) whose structure is con-
TABLE I. Structural measures and information anatomy of a stationary renewal process with interevent counts drawn from the distribution $F(n)$, $\mu \geq 0$, survival count distribution $w(n) = \sum_{n=0}^{\infty} F(n)$, and mean interevent count $\mu = \sum_{n=0}^{\infty} n F(n) < \infty$. The function $g(m, n)$ is defined by $g(m, n) = F(m + n + 1) - F(m)$. Cases are needed for $C_1$, but not other quantities, such as block entropy and information anatomy quantities, since the latter can be calculated just as well from prescient machines. The quantities $\chi$ and $E(M, N) = I[X_{-M:0};X_{0:N}]$ are no less interesting than the others given here, but their expressions are not compact; see App. [3]

rectional machine from these forward and reverse-time causal states, as described in Refs. [25, 26, 32]. Additional properties can then be deduced from the bidirectional machine, but we leave this for the future.

IV. INFORMATION ANATOMY OF RENEWAL PROCESSES

As Sec. II described, many quantities that capture a process’s predictability and randomness can be calculated from knowing the block entropy function $H(L)$. To calculate these various quantities, we do not necessarily need to find causal states. For example, we can use prescient states, as long as the corresponding HMM built from these prescient states is unifilar. Prescient statistics can be used to build prescient machines, which are non-minimal maximally-predictive HMMs of a process [21]. (Their non-minimality means they are not $\epsilon$-machines.) Prescient machines built from the prescient statistics of Lemma [1] happen to be unifilar Hidden Markov Models, corresponding to the unifilar Hidden Markov Model shown in Fig. [1]. The prescient machines make no distinction between eventually $\Delta$-Poisson renewal processes and one that is not, but they do contain information about the support of $F(n)$ through their transition dynamics. (See App. [1].)

Appendix [3] describes how a prescient machine can be used to calculate all information anatomy quantities—
\(r_\mu, b_\mu, \sigma_\mu, q_\mu,\) and the more familiar Shannon entropy rate \(h_\mu\) and excess entropy \(E\). A general strategy for calculating these quantities, as described in Sec. II and Refs. [13, 28], is to calculate \(b_\mu, h_\mu, E,\) and \(H[X_0]\), and then to derive the other quantities using the information-theoretic identities given in Sec. I.

The task of inferring an \(\varepsilon\)-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 underlying state given the previous underlying state 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.

To illustrate this point, we focus our attention on an incredibly simple nonunifilar HMM called (appropriately) the simple nonunifilar source (SNS) [35], it is shown in Fig. 5. 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 parametrized SNS is a discrete-time renewal process, in which the int...
FIG. 6. Contour plots of various information measures (in bits) as functions of SNS parameters $p$ and $q$. (Top left) $C_\mu$, increasing when $F(n)$ has slower decay. (Top right) $h_\mu$, higher when transition probabilities are maximally stochastic. (Bottom left) $E$, higher the closer the SNS comes to period-2. (Bottom right) $b_\mu$, highest between the maximally stochastic transition probabilities that maximize $h_\mu$ and maximally deterministic transition probabilities that maximize $E$.

terevent count distribution is:

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

(15)

Figure 5 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 Fig. 1 with a countable infinity of causal states.

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. And, that memory requires a countable infinity of states. In short,
the preferred complexity metric is a matter of implementation and use constraints, modulo important concerns regarding overfitting or ease of inference [36]. However, if we wish to calculate the information measures in Table I as accurately as possible, finding a maximally predictive model—a unifilar presentation, that is—is necessary.

Using the formulae of Table I, Fig. 6 shows how the single-measurement entropy \( H[X_0] \) is the solid red line, entropy rate \( h_\mu \) the solid green line, the bound information \( b_\mu \) the 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 \( \rho_\mu = H[X_0] - h_\mu \). (Bottom) The components of the predictable information—the excess entropy \( E = \sigma_\mu + b_\mu + q_\mu \) in bits—also as a function of \( p \) with \( p = q \). The blue line is \( q_\mu \): the green line is \( q_\mu + b_\mu \) so that the green area denotes \( b_\mu \)'s contribution to \( E \). The red line is \( E \) so that the red area denotes elusive information \( \sigma_\mu \) in \( E \). Note that for a range of \( p \) the co-information \( q_\mu \) is (slightly) negative.

![Graph](image)

**FIG. 7.** (Top) Information anatomy of the SNS as a function of \( p \) with parameters \( p = q \). The single-measurement entropy \( H[X_0] \) is the solid red line, entropy rate \( h_\mu \) the solid green line, the bound information \( b_\mu \) the 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 \( \rho_\mu = H[X_0] - h_\mu \). (Bottom) The components of the predictable information—the excess entropy \( E = \sigma_\mu + b_\mu + q_\mu \) in bits—also as a function of \( p \) with \( p = q \). The blue line is \( q_\mu \): the green line is \( q_\mu + b_\mu \) so that the green area denotes \( b_\mu \)'s contribution to \( E \). The red line is \( E \) so that the red area denotes elusive information \( \sigma_\mu \) in \( E \). Note that for a range of \( p \) the co-information \( q_\mu \) is (slightly) negative.

Using the formulae of Table I, Fig. 6 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 range of the intermittent 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 \( q \to 0 \), 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 \( h_\mu \) at the top right of Fig. 6 is maximized when transitions are uniformly stochastic and the bound information \( b_\mu \) at the bottom right is maximized somewhere between fully stochastic and fully deterministic regimes.

Figure 7 presents a more nuanced decomposition of the information measures as \( p = q \) vary from 0 to 1. The top most plot 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. The plot at the bottom decomposes the predictable information \( E \) 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 \( q_\mu = E - \sigma_\mu - b_\mu \) 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 a 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.

**VI. 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 surprisingly 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 some unanticipated subtleties. For instance, we had to define a new subclass of renewal process (“eventually \( \Delta \)-Poisson”) in order to completely classify \( \epsilon \)-machines of renewal processes. Eventually \( \Delta \)-Poisson renewal processes have a finite number of causal states despite the
unbounded support of their interevent count distribution. Additionally, the informational architecture formulae in Table I are surprisingly complicated, since exactly calculating these informational measures requires a unifilar presentation. In Sec. V we needed an infinite-state machine to study the informational architecture of a simplistic two-state machine.

Looking to the future, the new structural view of renewal processes should 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 do informational quantities scale, and the like. For example, Figs. 1-4 gave in the ideal setting—what are the effective states, what for infinite-state processes, as it tells us what to expect

Additionally, the informational architecture formulae in Table I are surprisingly complicated, since exactly calculating these informational measures requires a unifilar presentation. In Sec. V, we needed an infinite-state machine to study the informational architecture of a simplistic two-state machine.

The range of the results’ application is much larger than that explicitly considered here. The formulae in Table I will be most useful for understanding renewal processes with divergent statistical complexity. For instance, Ref. 8 applies the formulae to study the divergence of the statistical complexity of continuous-time processes as the observation time scale decreases. And in Ref. 9, we apply these formulae to renewal processes with divergent excess entropy. In particular, we are interested in understanding what in the architecture of infinite-state processes generates so-called critical phenomena—behavior with power-law temporal or spatial correlations 37. The analysis of critical systems often turns on having an appropriate “order parameter”. The statistical complexity and excess entropy are application-agnostic order parameters 38–40 and can allow us to better quantify when a “phase transition” in stochastic processes has or has not occurred, as seen in Ref. 9. Such critical behavior has even been implicated in early studies of human communication 41,42 and recently in neural dynamics 43 and in socially constructed, communal knowledge systems 44.

ACKNOWLEDGMENTS

The authors thank the Santa Fe Institute for its hospitality during visits. JPC is an SFI External Faculty member. This material is based upon work supported by, or in part by, the U.S. Army Research Laboratory and the U. S. Army Research Office under contracts W911NF-13-1-0390 and W911NF-12-1-0288. S.M. was funded by a National Science Foundation Graduate Student Research Fellowship and the U.C. Berkeley Chancellor’s Fellowship.

Appendix A: Causal Architecture

Notation. Rather than write pasts and futures as semi-infinite sequences, we note a past as a list of nonnegative integers 129. The semi-infinite past \( X_0 \) is equivalent to a list of interevent counts \( N_0 \) and the count \( N'_0 \) since last event. Similarly, the semi-infinite future \( X_\infty \) is equivalent to the count to next event \( N_0 - N'_0 \) and future interevent counts \( N_1 \).

Now, recall Lemma 4.

Lemma 4 The counts since last event are prescient statistics of a discrete-time stationary renewal process.

Proof. This follows almost immediately from the definition of stationary renewal process and the definition of causal states, since the random variables \( N_i \) are all i.i.d. Then:

\[
\Pr(X_0 | X_0) = \Pr(N_0 - N'_0 | N_0') \prod_{i=1}^{\infty} \Pr(N_i).
\]

And, 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 last event are prescient.

In light of Lemma 4 we introduce new notation to efficiently refer to groups of pasts with the same count since last event.

Notation. Let \( r_n^+ := \{x : x_{-n-1,0} = 10^n \} \) for \( n \in \mathbb{Z}_{\geq 0} \). Recall that \( 10^n = 100 \cdots 00 \) 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 as:

\[
\frac{\pi(r_n^+)}{\pi(r_m^+)} = \frac{\Pr(N_0' \geq n)}{\Pr(N_0' \geq m)} = \frac{w(n)}{w(m)},
\]

implying \( \pi(r_n^+) = w(n)/Z \). \( Z \) is a normalization constant that makes \( \sum_{n=0}^{\infty} w(n) = (T) \). And so:

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

In the main text, Thm. 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 $S^+ = R^+$, if $F$ has unbounded support. When the support is bounded such that $F(n) = 0$ for all $n \geq N$, $S^+ = \{r^+_n\}_{n=0}^N$. Finally, a discrete-time eventually $\Delta$-Poisson renewal process with characteristic $(\tilde{n}, \Delta)$ has forward-time causal states:

$$S^+ = \{r^+_n\}_{n=0}^{\tilde{n}-1} \cup \{\cup_{k=0}^{\infty} r^+_{\tilde{n}+k\Delta+m} \}_{m=0}^{\Delta-1}.$$

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:

$$\Pr(N_0 - N'_0 | N'_0 = n) = \Pr(N_0 - N'_0 | N'_0 = n').$$  \hspace{1cm} (A1)

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 $\Pr(N'_0 = n) = \Pr(N_0 \geq n) = w(n)$, we find that the equivalence class condition becomes:

$$\frac{F(m + n)}{w(n)} = \frac{F(m + n')}{w(n')} \hspace{1cm},$$  \hspace{1cm} (A2)

for all $m \geq 0$.

First, note that for these conditional probabilities to even be well defined, $w(n) > 0$ and $w(n') > 0$. Hence, if $F$ has bounded support—max$\text{supp} F(n) = N$—then the causal states do not include any $r^+_n$ for $n > N$. Furthermore, Eq. (A2) cannot be true for all $m \geq 0$, unless $n = n'$ for $n$ and $n'$ both. To see this, suppose that $n \neq n'$ but that Eq. (A2) holds. Then choose $m = N + 1 - \max(n, n')$ to give $0 = F(N + 1 - |n - n'|)/w(n')$, a contradiction unless $n = n'$.

So for all remaining cases, we can assume that $F$ in Eq. (A2) has unbounded support. A little rewriting makes the connection between Eq. (A2) 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 Eq. (A2) as:

$$\frac{F(m + n)}{F(n)} = \frac{F(m + n')}{F(n')}.$$  \hspace{1cm} (A3)

or more usefully:

$$F(n' + m) = F(n') F(n + m).$$

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:

$$F((n + m) + k\Delta') = \left(\frac{F(n')}{F(n)}\right)^k F(n + m).$$  \hspace{1cm} (A3)

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 Eq. (A2). But Eq. (A3) 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 Eq. (A2) 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 Eq. (A2) does hold for some $n \neq n'$, choose the minimal such $n$ and $n'$ both. The renewal process is eventually $\Delta$-Poisson with characterization $\Delta = n' - n$ and $\tilde{n}$. And, $F(\tilde{n} + m)/w(\tilde{n} + m) = F(\tilde{n} + m')/w(\tilde{n} + m')$ 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.

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

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

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

The probability distribution over these forward-time causal states is straightforward to derive from $\pi(r^+_n) = w(n)/(\mu + 1)$. So, 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 0 to $\tilde{n}$.) For an eventually $\Delta$-Poisson renewal process
\[ \pi(\sigma_n^+) = w(n)/\mu when n < 0 and:
\]
\[
\pi(\sigma_n^+) = \sum_{k=0}^{\infty} \pi(r_{n+k}^+ \\
= \sum_{k=0}^{\infty} w(n+k) \mu + 1 ,
\]
when \( \bar{n} < n < \bar{n} + \Delta \). And so, the statistical complexity given in Table I follows from \( C^+_\mu = H[S^+] \).

Recall Lemma 2 and Thm. 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 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 \( \bar{n} \), plus groupings of futures whose count since last event \( n \) are in the same equivalence class as \( \bar{n} \) 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 4 and Thm. 4.

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 \( r_{n}^+ \) since we just saw an event. Basic calculations show that these transition probabilities are:
\[
T^{(x)}_{r_{t} \rightarrow r_{t}^+ - r_{n}^+} = Pr(R_{t+1}^+ = r_{m}^+, X_{t+1} = x | R_{t}^+ = r_{n}^+) \\
= \frac{F(n)}{w(n)} \delta_{m,0} \delta_{x,1} + \frac{w(n+1)}{w(n)} \delta_{m,n+1} \times \delta_{x,0} .
\]
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:
\[
T^{(x)}_{\sigma \rightarrow \sigma'} = Pr(S_{t+1}^+ = \sigma', X_{t+1} = x | S_{t}^+ = \sigma) \\
= \sum_{r,r'} T^{(x)}_{r \rightarrow r'} Pr(S_{t+1}^+ = \sigma' | R_{t+1}^+ = r) \\
\times Pr(R_{t}^+ = r' | S_{t}^+ = \sigma) .
\]

Appendix B: Information Anatomy

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:
\[
h_\mu = \sum_{n=0}^{\infty} H[X_{t+1} | R_t^+] = r_{n}^+ \pi(r_{n}^+) \\
= - \sum_{n=0}^{\infty} w(n) \left( \frac{F(n)}{w(n)} \log_2 \frac{F(n)}{w(n)} \\
+ \frac{w(n+1)}{w(n)} \log_2 \frac{w(n+1)}{w(n)} \right) .
\]
And, 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 [25, 32]:
\[
E = I[R^+; R^-] \\
= H[R^+] - H[R^+ | R^-] .
\]
And so, to calculate, we note that:
\[
Pr(r_{n}^+ | r_{m}^-) = F(m+n) \mu + 1 and \\
Pr(r_{n}^+ | r_{m}^-) = F(n+m) w(m) .
\]
After some algebra, we find that:
\[
H[R^+] = - \sum_{n=0}^{\infty} \frac{w(n)}{\mu+1} \log_2 \frac{w(n)}{\mu+1} .
\]
and the forward crypticity is:

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

\[
= - \sum_{m=0}^{\infty} \frac{m+1}{\mu+1} \frac{F(m)}{\mu+1} + \sum_{m=0}^{\infty} \frac{w(m)}{\mu+1} \log_2 \frac{w(m)}{\mu+1}.
\]

These together imply:

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

And, 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: \(X_0 \rightarrow \mathcal{R}_0^+ \rightarrow \mathcal{R}_1^- \rightarrow X_1\). 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:

\[
\Pr(\mathcal{R}_1^- = n|\mathcal{R}_0^+ = m) = \sum_{r \in \mathcal{R}^+} \Pr(\mathcal{R}_1^- = n|\mathcal{R}_1^+ = r) \Pr(\mathcal{R}_1^+ = r|\mathcal{R}_0^+ = m) = F(m + n + 1) + F(m)F(m)
\]

\[
= \frac{w(m)}{w(m)}.
\]

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

\[
b_\mu = \sum_{n=0}^{\infty} (n+1) \frac{F(n) \log_2 F(n)}{\langle T \rangle} - \sum_{m,n=0}^{\infty} g(m,n) \frac{\log_2 g(m,n)}{\langle T \rangle},
\]

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 \(\sigma_\mu = E - q_\mu\). Thereby laying out information anatomy of stationary renewal processes.

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

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

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 [3] we know that counts from (to) the last (next) event are prescient forward-time (reverse-time) statistics. If our resolution on pasts (futures) is \(M\), then we cannot distinguish between counts since (to) the last (next) event which 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.

To calculate \(E(M,N)\), 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:

\[
u(m) = \sum_{n=m}^{\infty} w(n).
\]

Some algebra not shown here yields:

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

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

\[
- \sum_{m=0}^{M-1} w(m) \log_2 w(m)
\]

\[
+ \sum_{n=M}^{N+M-1} w(n) \log_2 w(n)
\]

\[
+ \sum_{n=N}^{N+M-1} w(n) \log_2 w(n)
\]

\[
- u(N) \log_2 u(N) + u(M) \log_2 u(M)
\]

\[
+ u(N + M) \log_2 u(N + M)
\]

\[
+ \sum_{m=0}^{M-1} \sum_{n=m}^{N+M-1} F(n) \log_2 F(n)
\]

\[
+ \sum_{m=0}^{M-1} \sum_{n=m}^{N+M-1} F(n) \log_2 F(n)
\]

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

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

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

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

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

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

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

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

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

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


[42] Though see [46, 47].


