Information Accessibility and Cryptic Processes

John R. Mahoney
Christopher J. Ellison
James P. Crutchfield

SFI WORKING PAPER: 2009-05-018

SANTA FE INSTITUTE

Santa Fe Institute Working Paper 09-05-018
arxiv.org:0905.4787 [physics.cond-mat]

Information Accessibility and Cryptic Processes

John R. Mahoney,[1,*] Christopher J. Ellison,[1,†] and James P. Crutchfield[1,2,‡]

[1]*Complexity Sciences Center and Physics Department, University of California at Davis, One Shields Avenue, Davis, CA 95616*
[2]*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501*

(Dated: May 29, 2009)

We give a systematic expansion of the crypticity—a recently introduced measure of the inaccessibility of a stationary process's internal state information. This leads to a hierarchy of k-cryptic processes and allows us to identify finite-state processes that have infinite crypticity—the internal state information is present across arbitrarily long, observed sequences. The crypticity expansion is exact in both the finite- and infinite-order cases. It turns out that k-crypticity is complementary to the Markovian finite-order property that describes state information in processes. One application of these results is an efficient expansion of the excess entropy—the mutual information between a process's infinite past and infinite future—that is finite and exact for finite-order cryptic processes.

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

INTRODUCTION

The data of phenomena come to us through observation. A large fraction of the theoretical activity of model building, though, focuses on internal mechanism. How are observation and modeling related? A first step is to frame the problem in terms of hidden processes—internal mechanisms probed via instruments that, in particular, need not accurately report a process's internal state. A practical second step is to measure the difference between internal structure and the information in observations.

We recently established that the amount of observed information a process communicates from the past to the future—the *excess entropy*—is the mutual information between its forward- and reverse-time minimal causal representations [1, 2]. This closed-form expression gives a concrete connection between the observed information and a process's internal structure.

Excess entropy, and related mutual information quantities, are widely used diagnostics for complex systems. They have been applied to detect the presence of organization in dynamical systems [3–6], in spin systems [7–9], in neurobiological systems [10, 11], and even in language [12, 13], to mention only a very few uses. Thus, understanding how much internal state structure is reflected in the excess entropy is critical to whether or not these and other studies of complex systems can draw structural inferences about the internal mechanisms that produce observed behavior.

Unfortunately, there is a fundamental problem. The excess entropy is *not* the internal state information the process stores—rather, the latter is the process's *statistical complexity* [1, 2]. On the positive side, there is a diagnostic. The difference between, if you will, experiment and theory (between observed information and internal structure) is controlled by the difference between a process's excess entropy and its statistical complexity. This difference is called the *crypticity*—how much internal state information is inaccessible [1, 2]. Here we introduce a classification of processes using a systematic expansion of crypticity.

The starting point is *computational mechanics*'s minimal causal representation of a stochastic process $\mathcal{P}$—the ϵ-*machine* [14, 15]. There, a process is viewed as a channel that communicates information from the past, $\overleftarrow{X} = \ldots X_{-3}X_{-2}X_{-1}$, to the future, $\overrightarrow{X} = X_0X_1X_2\ldots$. ($X_t$ takes values in a finite measurement alphabet $\mathcal{A}$.) The excess entropy is the shared (or mutual) information between the past and the future: $\mathbf{E} = I[\overleftarrow{X};\overrightarrow{X}]$. The amount of historical information that a process stores in the present is different. It is given by the Shannon information $C_\mu = H[\mathcal{S}]$ of the distribution over the ϵ-machine's *causal states* $\boldsymbol{\mathcal{S}}$. C_μ is called the *statistical complexity* and the causal states are sets of pasts $\overleftarrow{x}$ that are equivalent for prediction [14]:

$$\epsilon(\overleftarrow{x}) = \{\overleftarrow{x}' : \Pr(\overrightarrow{X}|\overleftarrow{x}) = \Pr(\overrightarrow{X}|\overleftarrow{x}')\} \ . \qquad (1)$$

Causal states have a Markovian property that they render the past and future statistically independent; they *shield* the future from the past [15]:

$$\Pr(\overleftarrow{X},\overrightarrow{X}|\mathcal{S}) = \Pr(\overleftarrow{X}|\mathcal{S})\Pr(\overrightarrow{X}|\mathcal{S}) \ . \qquad (2)$$

ϵ-Machines are also *unifilar* [14, 16]: From the start state, each observed sequence $\ldots x_{-3}x_{-2}x_{-1}\ldots$ corresponds to one and only one sequence of causal states. The signature of unifilarity is that on knowing the current state and measurement, the uncertainty in the next state vanishes: $H[\mathcal{S}_{t+1}|\mathcal{S}_t, X_t] = 0$.

Although they are not the same, the basic relationship between these quantities is clear: $\mathbf{E}$ is the process's effective channel utilization and C_μ is the sophistication of that channel. Their difference, one of our main concerns in the following, indicates how a process stores, manipulates, and hides internal state information.

Until recently, $\mathbf{E}$ could not be as directly calculated from the ϵ-machine as the process's entropy rate h_μ and its statistical complexity. Ref. [1] and Ref. [2] solved this problem, giving a closed-form expression for the excess entropy:

$$\mathbf{E} = I[\mathcal{S}^+;\mathcal{S}^-] \ , \quad (3)$$

where $\mathcal{S}^+$ are the causal states of the process scanned in the "forward" direction and $\mathcal{S}^-$ are the causal states of the process scanned in the "reverse" time direction.

This result comes in a historical context. Some time ago, an explicit expression for the excess entropy had been developed from the Hamiltonian for one-dimensional spin chains with range-R interactions [8]:

$$\mathbf{E} = C_\mu - R\, h_\mu \ . \quad (4)$$

A similar, but slightly less compact form is known for order-R Markov processes:

$$\mathbf{E} = H[X_0^R] - R\, h_\mu \ , \quad (5)$$

where $X_0^R = X_0, \ldots, X_{R-1}$. It has also been known for some time that the statistical complexity is an upper bound on the excess entropy [16]:

$$\mathbf{E} \leq C_\mu \ ,$$

which follows from the equality derived there:

$$\mathbf{E} = C_\mu - H[\mathcal{S}^+|\overrightarrow{X}] \ .$$

Using forward and reverse ϵ-machines, Ref. [1] extended this, deriving the closed-form expression for $\mathbf{E}$ in Eq. (3) and two new bounds on $\mathbf{E}$: $\mathbf{E} \leq C_\mu^-$ and $\mathbf{E} \leq C_\mu^+$. It also showed that:

$$H[\mathcal{S}^+|\overrightarrow{X}] = H[\mathcal{S}^+|\mathcal{S}^-] \quad (6)$$

and identified this quantity as controlling how a process hides its internal state information. For this reason, it is called the process's *crypticity*:

$$\chi^+ = H[\mathcal{S}^+|\overrightarrow{X}] \ . \quad (7)$$

In the context of forward and reverse ϵ-machines, one must distinguish two crypticities; depending on the scan direction one has:

$$\chi^+ = H[\mathcal{S}^+|\mathcal{S}^-] \text{ or}$$
$$\chi^- = H[\mathcal{S}^+|\mathcal{S}^-] \ .$$

In the following we will not concern ourselves with reverse representations and so can simplify the notation, using C_μ for C_μ^+ and χ for χ^+.

Here we show that, for a restricted class of processes, the crypticity in Eq. (6) can be systematically expanded to give an alternative closed-form to the excess entropy in Eq. (3). One ancillary benefit is a new and, we argue, natural hierarchy of processes in terms of information accessibility.

K-CRYPTICITY

The process classifications based on spin-block length and order-R Markov are useful. They give some insight into the nature of the kinds of process we can encounter and, concretely, they allow for closed-form expressions for the excess entropy (and other system properties). In a similar vein, we wish to carve the space of processes with a new blade. We define the class of *k-cryptic* processes and develop their properties and closed-form expressions for their excess entropies.

For convenience, we need to introduce several shorthands. First, to denote a symbol sequence that begins at time t and is L symbols long, we write X_t^L. Note that X_t^L includes X_{t+L-1}, but not X_{t+L}. Second, to denote a symbol sequence that begins at time t and continues on to infinity, we write $\overrightarrow{X}_t$.

Definition. *The* k-crypticity criterion *is satisfied when*

$$H[\mathcal{S}_k|\overrightarrow{X}_0] = 0 \ . \quad (8)$$

Definition. *A* k-cryptic *process is one for which the process's ϵ-machine satisfies the k-crypticity criterion.*

Definition. *An* ∞-cryptic *process is one for which the process's ϵ-machine does not satisfy the k-crypticity criterion for any finite k.*

Lemma 1. *$H[\mathcal{S}_k|\overrightarrow{X}_0]$ is a nonincreasing function of k.*

Proof. This follows directly from stationarity and the fact that conditioning on more random variables cannot increase entropy:

$$H[\mathcal{S}_{k+1}|\overrightarrow{X}_0] = [\mathcal{S}_k|\overrightarrow{X}_{-1}] \leq H[\mathcal{S}_k|\overrightarrow{X}_0] \ .$$

□

Lemma 2. *If $\mathcal{P}$ is k-cryptic, then $\mathcal{P}$ is also j-cryptic for all $j > k$.*

Proof. Being k-cryptic implies $H[\mathcal{S}_k|\overrightarrow{X}_0] = 0$. Applying Lem. 1, $H[\mathcal{S}_j|\overrightarrow{X}_0] \leq H[\mathcal{S}_k|\overrightarrow{X}_0] = 0$. By positivity of entropy, we conclude that $\mathcal{P}$ is also j-cryptic. □

This provides us with a new way of partitioning the space of processes. We create a parametrized class of sets $\{\chi_k : k = 0, 1, 2, \ldots\}$, where $\chi_k = \{\mathcal{P} : k\text{-cryptic and not } (k-1)\text{-cryptic}\}$.

The following result provides a connection to a very familiar class of processes.

Proposition 1. *If a process $\mathcal{P}$ is order-k Markov, then it is k-cryptic.*

Proof. If $\mathcal{P}$ is order-k Markov, then $H[\mathcal{S}_k|X_0^k] = 0$. Conditioning on more variables does not increase uncertainty, so:

$$H[\mathcal{S}_k|X_0^k, \overrightarrow{X}_k] = 0 \ .$$

But the lefthand side is $H[\mathcal{S}_k|\overrightarrow{X}_0]$. Therefore, $\mathcal{P}$ is k-cryptic. $\square$

Note that the converse of Prop. 1 is not true. For example, the Even Process (EP), the Random Noisy Copy Process (RnC), and the Random Insertion Process (RIP) (see Ref. [1] and Ref. [2]), are all 1-cryptic, but are not order-R Markov for any finite R.

Note also that Prop. 1 does not preclude an order-k Markov process from being j-cryptic, where $j < k$. Later we will show an example demonstrating this.

Given a process, in general one will not know its crypticity order. One way to investigate this is to study the sequence of estimates of χ at different orders. To this end, we define the k-cryptic approximation.

Definition. *The* k-cryptic approximation *is defined as*

$$\chi(k) = H[\mathcal{S}_0|X_0^k, \mathcal{S}_k] \ .$$

The k-Cryptic Expansion

We will now develop a systematic expansion of χ to order k in which $\chi(k)$ appears directly and the k-crypticity criterion plays the role of an error term.

Theorem 1. *The process crypticity is given by*

$$\chi = \chi(k) + H[\mathcal{S}_k|\overrightarrow{X}_0] \ . \qquad (9)$$

Proof. We calculate directly, starting from the definition, adding and subtracting the k-crypticity criterion term from χ's definition, Eq. (7):

$$\chi = H[\mathcal{S}_0|\overrightarrow{X}_0] - H[\mathcal{S}_k|\overrightarrow{X}_0] + H[\mathcal{S}_k|\overrightarrow{X}_0] \ .$$

We claim that the first two terms are $\chi(k)$. Expanding the conditionals in the purported $\chi(k)$ terms and then canceling, we get joint distributions:

$$H[\mathcal{S}_0|\overrightarrow{X}_0] - H[\mathcal{S}_k|\overrightarrow{X}_0] = H[\mathcal{S}_0, \overrightarrow{X}_0] - H[\mathcal{S}_k, \overrightarrow{X}_0] \ .$$

Now, splitting the future into two pieces and using this to write conditionals, the righthand side becomes:

$$H[\overrightarrow{X}_k|\mathcal{S}_0, X_0^k] + H[\mathcal{S}_0, X_0^k] - H[\overrightarrow{X}_k|\mathcal{S}_k, X_0^k] - H[\mathcal{S}_k, X_0^k] \ .$$

Appealing to the ϵ-machine's unifilarity, we then have:

$$H[\overrightarrow{X}_k|\mathcal{S}_k] + H[\mathcal{S}_0, X_0^k] - H[\overrightarrow{X}_k|\mathcal{S}_k, X_0^k] - H[\mathcal{S}_k, X_0^k] \ .$$

Now, applying causal shielding gives:

$$H[\overrightarrow{X}_k|\mathcal{S}_k] + H[\mathcal{S}_0, X_0^k] - H[\overrightarrow{X}_k|\mathcal{S}_k] - H[\mathcal{S}_k, X_0^k] \ .$$

Canceling terms, this simplifies to:

$$H[\mathcal{S}_0, X_0^k] - H[\mathcal{S}_k, X_0^k] \ .$$

We now re-expand, using unifilarity to give:

$$H[\mathcal{S}_0, X_0^k, \mathcal{S}_k] - H[\mathcal{S}_k, X_0^k] \ .$$

Finally, we combine these, using the definition of conditional entropy, to simplify again:

$$H[\mathcal{S}_0|X_0^k, \mathcal{S}_k] \ .$$

Note that this is our definition of $\chi(k)$.

This establishes our original claim:

$$\chi = \chi(k) + H[\mathcal{S}_k|\overrightarrow{X}_0] \ ,$$

with the k-crypticity criterion playing the role of an approximation error.

$\square$

Corollary 1. *A process $\mathcal{P}$ is k-cryptic if and only if*

$$\chi = \chi(k) \ .$$

Proof. Given the order-k expansion of χ just developed, we now assume the k-crypticity criterion is satisfied; viz., $H[\mathcal{S}_k|\overrightarrow{X}_0] = 0$. Thus, we have from Eq. (9):

$$\chi = \chi(k) \ .$$

Likewise, assuming $\chi = \chi(k)$ requires, by Eq. (9) that $H[\mathcal{S}_k|\overrightarrow{X}_0] = 0$ and thus the process is k-cryptic. $\square$

Corollary 2. *For any process, $\chi(0) = 0$.*

Proof.

$$\begin{aligned} \chi(0) &= H[\mathcal{S}_0|X_0^0, \mathcal{S}_0] \\ &= H[\mathcal{S}_0|\mathcal{S}_0] = 0 \ . \end{aligned}$$

$\square$

Convergence

Proposition 2. *The approximation $\chi(k)$ is a nondecreasing function of k.*

Proof. Lem. 1 showed that $H[\mathcal{S}_k|\overrightarrow{X}_0]$ is a nonincreasing function of k. By Thm. 1, $\chi(k)$ must be a nondecreasing function of k. $\square$

Corollary 3. *Once $\chi(k)$ reaches the value χ, $\chi(j) = \chi$ for all $j > k$.*

Proof. If there exists such a k, then by Thm. 1 the process is k-cryptic. By Lem. 2, the process is j-cryptic for all $j > k$. Again, by Thm. 1, $\chi(j) = \chi$. $\square$

Corollary 4. *If there is a $k \geq 1$ for which $\chi(k) = 0$, then $\chi(1) = 0$.*

Proof. By positivity of the conditional entropy $H[\mathcal{S}_0|X_0, \mathcal{S}_1]$, $\chi(1) \geq 0$. By the nondecreasing property of $\chi(k)$ from Prop. 2, $\chi(1) \leq \chi(k) = 0$. Therefore, $\chi(1) = 0$. □

Corollary 5. *If $\chi(1) = 0$, then $\chi(k) = 0$ for all k.*

Proof. Applying stationarity, $\chi(1) = H[\mathcal{S}_0|X_0, \mathcal{S}_1] = H[\mathcal{S}_k|X_k, \mathcal{S}_{k+1}]$. We are given $\chi(1) = 0$ and so $H[\mathcal{S}_k|X_k, \mathcal{S}_{k+1}] = 0$. We use this below. Expanding $\chi(k+1)$,

$$\begin{aligned}
\chi(k+1) &= H[\mathcal{S}_0|X_0^{k+1}, \mathcal{S}_{k+1}] \\
&= H[\mathcal{S}_0|X_0^k, X_k, \mathcal{S}_{k+1}] \\
&= H[\mathcal{S}_0|X_0^k, \mathcal{S}_k, X_k, \mathcal{S}_{k+1}] \\
&\leq H[\mathcal{S}_0|X_0^k, \mathcal{S}_k] \\
&= \chi(k) \ .
\end{aligned}$$

The third line follows from $\chi(1) = 0$. By Prop. 2, $\chi(k+1) \geq \chi(k)$. Therefore, $\chi(k+1) = \chi(k)$. Finally, using $\chi(1) = 0$, we have by induction that $\chi(k) = 0$ for all k. □

Corollary 6. *If there is a $k \geq 1$ for which $\chi(k) = 0$, then $\chi(j) = 0$ for all $j \geq 1$.*

Proof. This follows by composing Cor. 4 with Cor. 5. □

Together, the proposition and its corollaries show that $\chi(k)$ is a nondecreasing function of k which, if it reaches χ at a finite k, remains at that value for all larger k.

Proposition 3. *The cryptic approximation $\chi(k)$ converges to χ as $k \to \infty$.*

Proof. Note that $\chi = \lim_{k\to\infty} H[\mathcal{S}_0|X_0^k]$ and recall that $\chi(k) = H[\mathcal{S}_0|X_0^k, \mathcal{S}_k]$. We show that the difference approaches zero:

$$\begin{aligned}
H[\mathcal{S}_0&|X_0^k] - H[\mathcal{S}_0|X_0^k, \mathcal{S}_k] \\
&= H[\mathcal{S}_0, X_0^k] - H[X_0^k] \\
&\quad - H[\mathcal{S}_0, X_0^k, \mathcal{S}_k] + H[X_0^k, \mathcal{S}_k] \\
&= H[\mathcal{S}_0, X_0^k] - H[X_0^k] \\
&\quad - H[\mathcal{S}_0, X_0^k] + H[X_0^k, \mathcal{S}_k] \\
&= H[X_0^k, \mathcal{S}_k] - H[X_0^k] \\
&= H[\mathcal{S}_k|X_0^k] \ .
\end{aligned}$$

Moreover, $\lim_{k\to\infty} H[\mathcal{S}_k|X_0^k] = 0$ by the ϵ map from pasts to causal states of Eq. (1). Therefore, as $k \to \infty$, $\chi(k) \to \chi$. □

Excess Entropy for k-Cryptic Processes

Given a k-cryptic process, we can calculate its excess entropy in a form that involves a sum of $\propto |\mathcal{A}^k|$ terms, where each term involves products of k matrices. Specifically, we have the following.

Corollary 7. *A process $\mathcal{P}$ is k-cryptic if and only if $\mathbf{E} = C_\mu - \chi(k)$.*

Proof. From Ref. [1], we have $\mathbf{E} = C_\mu - \chi$, and by Cor. 1, $\chi = \chi(k)$. Together, these complete the proof. □

The following proposition is a simple and useful consequence of the class of k-cryptic processes.

Corollary 8. *A process $\mathcal{P}$ is 0-cryptic* if *and* only if $\mathbf{E} = C_\mu$.

Proof. If $\mathcal{P}$ is 0-cryptic, our general expression then reads

$$\begin{aligned}
\mathbf{E} &= C_\mu - H[\mathcal{S}_0|X_0^0, \mathcal{S}_0] \\
&= C_\mu \ .
\end{aligned}$$

To establish the opposite direction, $\mathbf{E} = C_\mu$ and Cor. 7 imply that $\chi(k) = 0$ for all k. In particular, $\chi(0)$ and the process is 0-cryptic. □

Crypticity versus Markovity

Equation (4) and Equation (5) give expressions for $\mathbf{E}$ in the cases when the process is order-R Markov and when it is an order-R spin chain. These results hinge on whether or not $H[X_0^R] = C_\mu$.

Reference [8] stated a condition under which equality holds in terms of transfer matrices. Here we state a simpler condition by equating two chain rule expansions of $H[X_0^R, \mathcal{S}_R]$:

$$H[X_0^R|\mathcal{S}_R] + H[\mathcal{S}_R] = H[\mathcal{S}_R|X_0^R] + H[X_0^R] \ .$$

$H[\mathcal{S}_R|X_0^R] = 0$ by virtue of the fact that each such (history) word maps to exactly one causal state by Eq. (1). Thus, we conclude that for order-R Markov processes:

$$H[X_0^R] = H[\mathcal{S}_R] \quad \iff \quad H[X_0^R|\mathcal{S}_R] = 0 \ .$$

So, an order-R Markov process is also a spin chain *if* and *only if* $H[X_0^R|\mathcal{S}_R] = 0$. This means that there is a $1-1$ correspondence between the R-blocks and causal states, confirming the interpretation specified in Ref. [8].

We can also extend the condition for $H[X_0^R] = C_\mu$ to the results presented here in the following way.

Proposition 4.

$$H[X_0^R|\mathcal{S}_R] = 0 \quad \iff \quad \chi(R) = R\, h_\mu \ , \tag{10}$$

where h_μ is the process's entropy rate.

Proof. The proof is a direct calculation:

$$\begin{aligned}
\chi(R) &= H[\mathcal{S}_0|X_0^R, \mathcal{S}_R] \\
&= H[\mathcal{S}_0, X_0^R] - H[X_0^R, \mathcal{S}_R] \\
&= H[\mathcal{S}_0, X_0^R] - H[X_0^R|\mathcal{S}_R] - H[\mathcal{S}_R] \\
&= H[\mathcal{S}_0, X_0^R] - H[X_0^R|\mathcal{S}_R] - H[\mathcal{S}_0] \\
&= H[X_0^R|\mathcal{S}_0] - H[X_0^R|\mathcal{S}_R] \\
&= Rh_\mu - H[X_0^R|\mathcal{S}_R] \ .
\end{aligned}$$

□

Proposition 5. *Periodic processes can be arbitrary order-R Markov, but are all* 0*-cryptic.*

Proof. According to Ref. [17], we have $\mathbf{E} = C_\mu$. By Cor. 8 the process is 0-cryptic. □

Proposition 6. *A positive entropy-rate process that is an order-R Markov spin chain is not* $(R-1)$*-cryptic.*

Proof. Assume that the order-R Markov spin chain *is* $(R-1)$-cryptic.

For $R \geq 1$, If the process is $(R-1)$-cryptic, then by Cor. 1 $\chi(R-1) = \chi$. Combining this with the above Prop. 4, we have $\chi(R-1) = (R-1)h_\mu - H[X_0^{R-1}|\mathcal{S}_{R-1}]$. If it is an order-$R$ Markov spin chain, then we also have from Eq. (4) that $\chi = Rh_\mu$. Combining this with the previous equation, we find that $H[X_0^{R-1}|\mathcal{S}_{R-1}] = -h_\mu$. By positivity of conditional entropies, we have reached a contradiction. Therefore an order-R Markov spin chain must not be $(R-1)$-cryptic.

For $R = 0$, the proof also holds since negative cryptic orders are not defined. □

Proposition 7. *A positive entropy-rate process that is an order-R Markov spin chain is not* $(R-n)$*-cryptic for any* $1 \geq n \geq R$.

Proof. For $R \geq 1$, By Lem. 2, if the process were $(R-n)$-cryptic for some $1 \geq n \geq R$, then it would be $(R-1)$-cryptic. By Prop. 6, this is not true. Therefore, the primitive orders of Markovity and crypticity are the same. Similarly, for $R = 0$, the proof also holds since negative cryptic orders are not defined. □

EXAMPLES

It is helpful to see crypticity in action. We now turn to a number of examples to illustrate how various orders of crypticity manifest themselves in ϵ-machine structure and what kinds of processes are cryptic and so hide internal state information from an observer. For details (transition matrices, notation, and the like) not included in the following and for complementary discussions and analyses of them, see Refs. [1, 2, 17].

We start at the bottom of the crypticity hierarchy with a 0-cryptic process and then show examples of 1-cryptic and 2-cryptic processes. Continuing up the hierarchy, we generalize and give a parametrized family of processes that are k-cryptic. Finally, we demonstrate an example that is ∞-cryptic.

Even Process: 0-Cryptic

Figure 1 gives the ϵ-machine for the Even Process. The Even Process produces binary sequences in which all blocks of uninterrupted 1s are even in length, bounded by zeros. Further, after each even length is reached, there is a probability p of breaking the block of 1s by inserting one or more 0s.

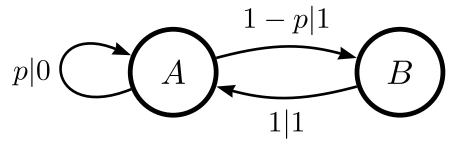

FIG. 1: A 0-cryptic process: Even Process. The transitions denote the probability p of generating symbol x as $p|x$.

Reference [2] showed that the Even Process is 0-cryptic with a statistical complexity of $C_\mu = H(1/(2-p))$, an entropy rate of $h_\mu = H(p)/(2-p)$, and crypticity of $\chi = 0$. If $p = \frac{1}{2}$, then $C_\mu = \log_2(3) - \frac{2}{3}$ bits and $\mathbf{E} = \log_2(3) - \frac{2}{3}$ bits. (As Ref. [2] notes, these closed-form expressions for C_μ and $\mathbf{E}$ have been known for some time.)

To see why the Even Process is 0-cryptic, note that if $X = 0$, then $\mathcal{S}_0 = A$; and if $X = 1$, then $\mathcal{S}_0 = B$. Therefore, the 0-crypticity criterion of Eq. (8) is satisfied.

It is important to note that this process is *not* order-R Markov for any finite R [17]. Nonetheless, our new expression for $\mathbf{E}$ is valid. This shows the broadening of our ability to calculate $\mathbf{E}$ even for low complexity processes that are, in effect, infinite-order Markov.

Golden Mean Process: 1-Cryptic

Figure 2 shows the ϵ-machine for the Golden Mean Process [17]. The Golden Mean Process is one in which no two 0s occur consecutively. After each 1, there is a probability p of generating a 0. As sequence length grows, the ratio of the number of allowed words of length L to the number of allowed words at length $L-1$ approaches the golden ratio; hence, its name. The Golden Mean Process ϵ-machine looks remarkably similar to that for the Even Process. The informational analysis, however, shows that they have markedly different properties.

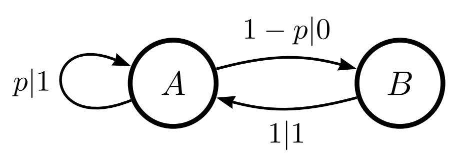

FIG. 2: A 1-cryptic process: Golden Mean Process.

Reference [2] showed that the Golden Mean Process has the same statistical complexity and entropy rate as the Even Process: $C_\mu = H(1/(2-p))$ and $h_\mu = H(p)/(2-p)$. However, the crypticity is not zero (for

$0 < p < 1$). From Cor. 1 we calculate:

$$\begin{aligned}\chi &= \chi(1)\\ &= H[\mathcal{S}_0|X_0^1, \mathcal{S}_1]\\ &= H[\mathcal{S}_0|X_0^1]\\ &= Pr(0)H[\mathcal{S}_0|X_0 = 0] + Pr(1)H[\mathcal{S}_0|X_0 = 1]\\ &= H(p)/(2-p)\ .\end{aligned}$$

If $p = \frac{1}{2}$, $C_\mu = \log_2(3) - \frac{2}{3}$ bits, an excess entropy of $\mathbf{E} = \log_2(3) - \frac{4}{3}$ bits, and a crypticity of $\chi = \frac{2}{3}$. Thus, the excess entropy differs from that of the Even Process. (As with the Even Process, these closed-form expressions for C_μ and $\mathbf{E}$ have been known for some time.)

The Golden Mean Process is 1-cryptic. To see why, it is enough to note that it is order-1 Markov. By Prop. 1, it is 1-cryptic. We know it is not 0-cryptic since any future beginning with 1 could have originated in either state A or B. In addition, the spin-block expression for excess entropy of Ref. [17], Eq. (4) here, applies for an $R = 1$ Markov chain.

Butterfly Process: 2-Cryptic

The next example, the Butterfly Process of Fig. 3, illustrates in a more explicit way than possible with the previous processes the role that crypticity plays and how it can be understood in terms of an ϵ-machine's structure. Much of the explanation does not require calculating much, if anything.

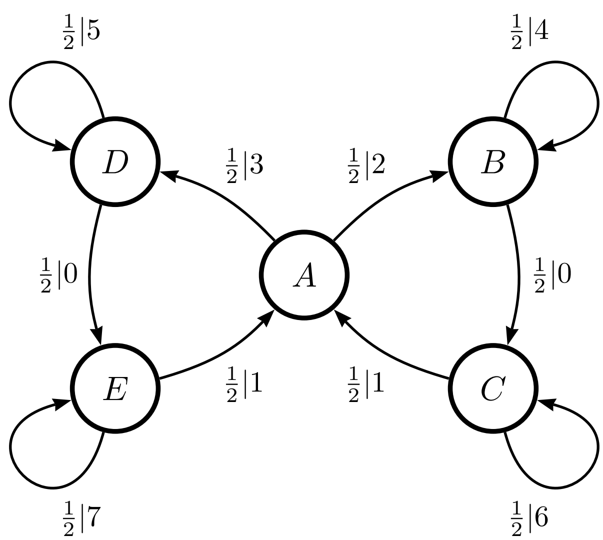

FIG. 3: A 2-cryptic process: Butterfly Process over a 6-symbol alphabet.

It is first instructive to see why the Butterfly Process is *not* 1-cryptic.

If we can find a family $\{\overrightarrow{x}_0\}$ such that $H[\mathcal{S}_1|\overrightarrow{X}_0 = \overrightarrow{x}_0] \neq 0$, then the total conditional entropy will be positive and, thus, the machine will not be 1-cryptic. To show that this can happen, consider the future $\overrightarrow{x}_0 = (0, 1, 2, 4, 4, 4, \ldots)$. It is clear that the state following 1 must be A. Thus, in order to generate 0 or 1 before arriving at A, the state pair $(\mathcal{S}_0, \mathcal{S}_1)$ can be either (B, C) or (D, E). This uncertainty in $\mathcal{S}_1$ is enough to break the criterion. And this occurs for the family $\{\overrightarrow{x}_0\} = \{0, 1, \ldots\}$.

To see that the process is 2-cryptic, notice that the two paths (B, C) and (D, E) converge on A. Therefore, there is no uncertainty in $\mathcal{S}_2$ given this future. It is reasonably straightforward to see that indeed *any* (X_0, X_1) will lead to a unique causal state. This is because the Butterfly Process is a very limited version of an 8-symbol order-2 Markov process.

Note that the transition matrix is doubly-stochastic and so the stationary distribution is uniform. The statistical complexity is rather direct in this case: $C_\mu = \log_2(5)$. We now can calculate χ using Cor. 1:

$$\begin{aligned}\chi &= \chi(2)\\ &= H[\mathcal{S}_0|X_0^2, \mathcal{S}_2]\\ &= H[\mathcal{S}_0|X_0^2]\\ &= \mathrm{Pr}(01)H[\mathcal{S}_0|X_0^2 = 01] + \mathrm{Pr}(12)H[\mathcal{S}_0|X_0^2 = 12]\\ &\quad + \mathrm{Pr}(13)H[\mathcal{S}_0|X_0^2 = 13]\\ &= 2\frac{1}{4}\frac{1}{5}1 + 2\frac{1}{4}\frac{1}{5}1 + 2\frac{1}{4}\frac{1}{5}1\\ &= \frac{3}{10}\text{ bits.}\end{aligned}$$

From Cor. 7, we get an excess entropy of

$$\begin{aligned}\mathbf{E} &= C_\mu - \chi(2)\\ &= \log 2(5) - \frac{3}{10}\\ &\approx 2.0219\text{ bits.}\end{aligned}$$

For comparison, if we had assumed the Butterfly Process was 1-cryptic, then we would have:

$$\begin{aligned}\mathbf{E} &= C_\mu - \chi(1)\\ &= C_\mu - (H[\mathcal{S}_0, X_0] - H[\mathcal{S}_1, X_0])\\ &\approx \log 2(5) - (3.3219 - 2.5062)\\ &= \log 2(5) - 0.8156 \approx 1.5063\text{ bits.}\end{aligned}$$

We can see that this is substantially below the true value: a 25% error.

Restricted Golden Mean: k-Cryptic

Now we turn to illustrate a crypticity-parametrized family of processes, giving examples of k-cryptic processes for any k. We call this family the Restricted Golden Mean as its support is a restriction of the Golden Mean support. (See Fig. 4 for its ϵ-machines.) The $k = 1$ member of the family is exactly the Golden Mean.

It is straightforward to see that this process is order-k Markov. Proposition 1 then implies it is (at most) k-cryptic. In order to show that it is not $(k-1)$-cryptic, consider the case $\overrightarrow{x}_0 = 1^k, 0, \ldots$. The first $(k-1)$ 1s will induce a mixture over states k and 0. The following future $\overrightarrow{x}_k = 1, 0, \ldots$ is consistent with both states k and 0. Therefore, the $(k-1)$-crypticity criterion is not satisfied. Therefore, it is k-cryptic.

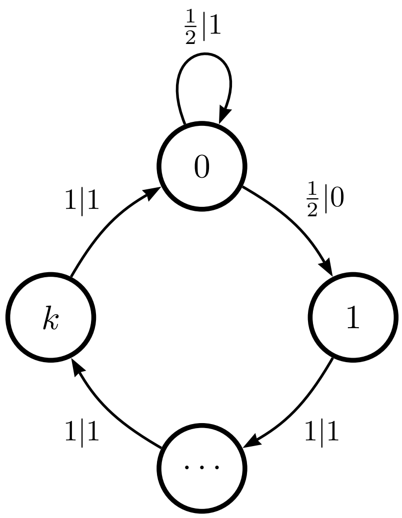

FIG. 4: k-cryptic processes: Restricted Golden Mean Family.

For arbitrary k, there are $k+1$ causal states and the stationary distribution is:

$$\pi = \left(\frac{2}{k+2}, \frac{1}{k+2}, \frac{1}{k+2}, \ldots, \frac{1}{k+2} \right) .$$

The statistical complexity is

$$C_\mu = \log_2(k+2) - \frac{2}{k+2} .$$

For the k-th member of the family, we have for the crypticity:

$$\chi = \chi(k) = \frac{2k}{k+2} .$$

And the excess entropy follows directly from Cor. 7:

$$\begin{aligned} \mathbf{E} &= C_\mu - \chi \\ &= \log_2(k+2) - \frac{2(k+1)}{k+2} , \end{aligned}$$

which diverges with k. (Calculational details will be provided elsewhere.)

Stretched Golden Mean

The Stretched Golden Mean is a family of processes that does not occupy the same support as the Golden Mean. Instead of requiring that blocks of 0s are of length 1, we require that they are of length k. Here, the Markov order (k) grows, but the cryptic order remains 1 for all k.

Again, it is straightforward to see that this process is order-k Markov. To see that it is 1-cryptic, first note that if $X_0 = 1$, then $\mathcal{S}_1 = 0$. Next consider the case when $X_0 = 0$. If the future $\overrightarrow{x}_1 = 1, \ldots$, then $\mathcal{S}_1 = k$. Similarly, if the future $\overrightarrow{x}_1 = 0^n, 1, \ldots$, then $\mathcal{S}_1 = k - n$. This family exhibits arbitrary separation between its Markov order and its cryptic order and so demonstrates that these properties are not redundant.

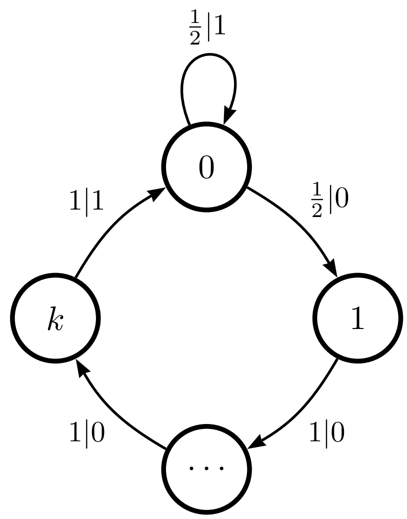

FIG. 5: k-cryptic processes: Stretched Golden Mean Family.

The stationary distribution is the same as for the Restricted Golden Mean and so, then, is the statistical complexity. In addition, we have:

$$\begin{aligned} \chi = \chi(1) &= H[\mathcal{S}_0 | X_0, \mathcal{S}_1] \\ &= h_\mu . \end{aligned}$$

Consequently,

$$\mathbf{E} = C_\mu - \chi = C_\mu - h_\mu .$$

The Nemo Process: ∞-Cryptic

We close our cryptic process bestiary with a (very) finite-state process that has infinite crypticity: The three-state Nemo Process. Over no finite-length sequence will all of the internal state information be present in the observations. The Nemo Process ϵ-machine is shown in Fig. 6.

Its stationary state distribution is

$$\Pr(\mathcal{S}) \equiv \pi = \frac{1}{3-2p} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{pmatrix} 1 & 1-p & 1-p \end{pmatrix} \end{array},$$

from which one calculates the statistical complexity:

$$C_\mu = \log_2(3-2p) - \frac{2(1-p)}{3-2p} \log_2(1-p) .$$

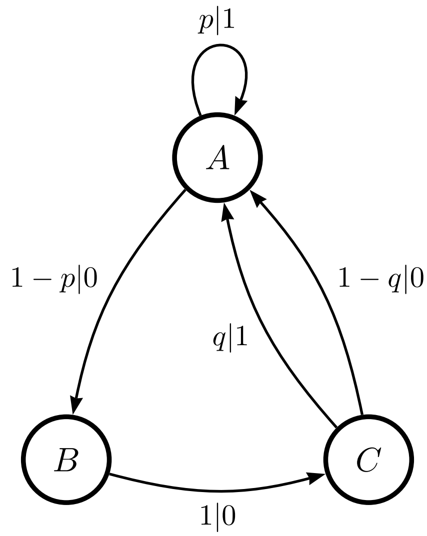

FIG. 6: The ∞-cryptic Nemo Process.

The Nemo Process is not a finite-cryptic process. That is, there exists no finite k for which $H[\mathcal{S}_k|\overrightarrow{X}_0] = 0$. To show this, we must demonstrate that there exists a family of futures such that for each future $H[\mathcal{S}_k|\overrightarrow{X}_0 = \overrightarrow{x}] > 0$. The family of futures we use begins with all 0s and then has a 1. Intuitively, the 1 is chosen because it is a synchronizing word for the process—after observing a 1, the ϵ-machine is always in state A. Then, causal shielding will decouple the infinite future from the first few symbols, thereby allowing us to compute the conditional entropies for the entire family of futures.

First, recall the shorthand:

$$\Pr(\mathcal{S}_k|\overrightarrow{X}_0) = \lim_{L\to\infty} \Pr(\mathcal{S}_k|X_0^L) \ .$$

Without loss of generality, assume $k < L$. Then,

$$\begin{aligned}\Pr(\mathcal{S}_k|X_0^L) &= \frac{\Pr(X_0^k, \mathcal{S}_k, X_k^L)}{\Pr(X_0^L)} \\ &= \frac{\Pr(X_k^L|X_0^k, \mathcal{S}_k)\Pr(X_0^k, \mathcal{S}_k)}{\Pr(X_0^L)} \\ &= \frac{\Pr(X_k^L|\mathcal{S}_k)\Pr(X_0^k, \mathcal{S}_k)}{\Pr(X_0^L)} \ ,\end{aligned}$$

where the last step is possible since the causal states are Markovian [15], shielding the past from the future. Each of these quantities is given by:

$$\begin{aligned}\Pr(X_k^L = w|\mathcal{S}_k = \sigma) &= [T^{(w)}\mathbf{1}]_\sigma \\ \Pr(X_0^k = w, \mathcal{S}_k = \sigma) &= [\pi T^{(w)}]_\sigma \\ \Pr(X_0^L = w) &= \pi T^{(w)}\mathbf{1} \ .\end{aligned}$$

where $T^{(w)} \equiv T^{(x_0)}T^{(x_1)}\cdots T^{(x_{L-1})}$, $\mathbf{1}$ is a column vector of 1s, and $T^{(x)}_{\sigma\sigma'} = \Pr(\mathcal{S}' = \sigma', X = x|\mathcal{S} = \sigma)$. To establish $H[\mathcal{S}_k|\overrightarrow{X}_0] > 0$ for any k, we rely on using values of k that are multiples of three. So, we concentrate on the following for $n = 0, 1, 2, \ldots$:

$$H[\mathcal{S}_{3n}|X_0^{3n+1} = 0^{3n}1, \overrightarrow{X}_{3n+1}] > 0 \ .$$

Since 1 is a synchronizing word, we can greatly simplify the conditional probability distribution. First, we freely include the synchronized causal state A and rewrite the conditional distribution as fraction:

$$\begin{aligned}&\Pr(\mathcal{S}_{3n}|X_0^{3n+1} = 0^{3n}1, \overrightarrow{X}_{3n+1}) \\ &= \Pr(\mathcal{S}_{3n}|X_0^{3n+1} = 0^{3n}1, \mathcal{S}_{3n+1} = A, \overrightarrow{X}_{3n+1}) \\ &= \frac{\Pr(\mathcal{S}_{3n}, X_0^{3n+1} = 0^{3n}1, \mathcal{S}_{3n+1} = A, \overrightarrow{X}_{3n+1})}{\Pr(X_0^{3n+1} = 0^{3n}1, \mathcal{S}_{3n+1} = A, \overrightarrow{X}_{3n+1})} \ .\end{aligned}$$

Then, we factor everything except $\overrightarrow{X}_{3n+1}$ out of the numerator and make use of causal shielding to simplify the conditional. For example, the numerator becomes:

$$\begin{aligned}&\Pr(\mathcal{S}_{3n}, X_0^{3n+1} = 0^{3n}1, \mathcal{S}_{3n+1} = A, \overrightarrow{X}_{3n+1}) \\ &= \Pr(\overrightarrow{X}_{3n+1}|\mathcal{S}_{3n}, X_0^{3n+1} = 0^{3n}1, \mathcal{S}_{3n+1} = A) \\ &\quad\times \Pr(\mathcal{S}_{3n}, X_0^{3n+1} = 0^{3n}1, \mathcal{S}_{3n+1} = A) \\ &= \Pr(\overrightarrow{X}_{3n+1}|\mathcal{S}_{3n+1} = A) \\ &\quad\times \Pr(\mathcal{S}_{3n}, X_0^{3n+1} = 0^{3n}1, \mathcal{S}_{3n+1} = A) \\ &= \Pr(\overrightarrow{X}_{3n+1}|\mathcal{S}_{3n+1} = A)\Pr(\mathcal{S}_{3n}, X_0^{3n+1} = 0^{3n}1) \ .\end{aligned}$$

Similarly, the denominator becomes:

$$\begin{aligned}&\Pr(X_0^{3n+1} = 0^{3n}1, \mathcal{S}_{3n+1} = A, \overrightarrow{X}_{3n+1}) \\ &= \Pr(\overrightarrow{X}_{3n+1}|\mathcal{S}_{3n+1} = A)\Pr(X_0^{3n+1} = 0^{3n}1) \ .\end{aligned}$$

Combining these results, we obtain a finite form for the entropy of S_{3n} conditioned on a family of infinite futures, first noting:

$$\Pr(\mathcal{S}_{3n}|X_0^{3n+1} = 0^{3n}1, \overrightarrow{X}_{3n+1}) = \Pr(\mathcal{S}_{3n}|X_0^{3n+1} = 0^{3n}1) \ .$$

Thus, for all $\overrightarrow{x}_{3n+1}$, we have:

$$\begin{aligned}&H[\mathcal{S}_{3n}|X_0^{3n+1} = 0^{3n}1, \overrightarrow{X}_{3n+1} = \overrightarrow{x}_{3n+1}] \\ &= H[\mathcal{S}_{3n}|X_0^{3n+1} = 0^{3n}1] \ .\end{aligned}$$

Now, we are ready to compute the conditional entropy for the entire family. First, note that $T^{(0)}$ raised to the third power is a diagonal matrix with each element equal to $(1-p)(1-q)$. Thus, for $j = 1, 2, 3\ldots$:

$$\left[T^{(0)}\right]^{3j}_{\sigma\sigma} = (1-p)^j(1-q)^j \ .$$

Using all of the above relations, we can easily calculate:

$$\Pr(\mathcal{S}_{3n}|X_0^{3n+1} = 0^{3n+1}1) = \frac{1}{3-2p}\ \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{pmatrix} p & 0 & q(1-p) \end{pmatrix} \end{array} \ .$$

Thus, for $p, q \in (0,1)$, we have:

$$\begin{aligned}
&H[\mathcal{S}_{3n}|\overrightarrow{X}_0] \\
&\geq H[\mathcal{S}_{3n}|X_0^{3n+1} = 0^{3n}1, \overrightarrow{X}_{3n+1}] \\
&= \sum_{\overrightarrow{x}_{3n+1}} \Pr\left(X_0^{3n+1} = 0^{3n}1, \overrightarrow{X}_{3n+1} = \overrightarrow{x}_{3n+1}\right) \\
&\quad \times H[\mathcal{S}_{3n}|X_0^{3n+1} = 0^{3n}1, \overrightarrow{X}_{3n+1} = \overrightarrow{x}_{3n+1}] \\
&= H[\mathcal{S}_{3n}|X_0^{3n+1} = 0^{3n}1] \\
&\quad \times \sum_{\overrightarrow{x}_{3n+1}} \Pr\left(X_0^{3n+1} = 0^{3n}1, \overrightarrow{X}_{3n+1} = \overrightarrow{x}_{3n+1}\right) \\
&= H[\mathcal{S}_{3n}|X_0^{3n+1} = 0^{3n}1] \Pr(X_0^{3n+1} = 0^{3n}1) \\
&= \left(\frac{p}{3-2}\log_2\frac{3-2p}{p} + \frac{q(1-p)}{3-2p}\log_2\frac{q(1-p)}{3-2p}\right) \\
&\quad \times [(1-p)(1-q)]^{3n} \\
&> 0 \ .
\end{aligned}$$

So, any time k is a multiple of three, $H[S_k|\overrightarrow{X}_0] > 0$. Finally, suppose $(k \mod 3) = i$, where $i \neq 0$. That is, suppose k is not a multiple of three. By Lem. 1, $H[\mathcal{S}_k|\overrightarrow{X}_0] \geq H[\mathcal{S}_{k+i}|\overrightarrow{X}_0]$ and, since we just showed that the latter quantity is always strictly greater than zero, we conclude that $H[\mathcal{S}_k|\overrightarrow{X}_0] > 0$ for every value of k.

The above establishes that the Nemo Process does not satisfy the k-crypticity criterion for any finite k. Thus, the Nemo process is ∞-cryptic. This means that we cannot make use of the k-cryptic approximation to calculate χ or $\mathbf{E}$.

Fortunately, the techniques introduced in Ref. [1] and Ref. [2] do not rely on an approximation method. To avoid ambiguity denote the statistical complexity we just computed as C_μ^+. When the techniques are applied to the Nemo Process, we find that the process is causally reversible ($C_\mu^+ = C_\mu^-$) and has the following forward-reverse causal-state conditional distribution:

$$\Pr(\mathcal{S}^+|\mathcal{S}^-) = \frac{1}{p+q-pq} \begin{array}{c} \\ D \\ E \\ F \end{array} \!\!\begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{pmatrix} p & 0 & q(1-p) \\ 0 & q & p(1-q) \\ q & p(1-q) & 0 \end{pmatrix} \end{array} \ .$$

With this, one can calculate $\mathbf{E}$, in closed-form, via:

$$\mathbf{E} = C_\mu^+ - H[\mathcal{S}^+|\mathcal{S}^-] \ .$$

(Again, calculational details will be provided elsewhere.)

CONCLUSION

Calculating the excess entropy $I[\overleftarrow{X};\overrightarrow{X}]$ is, at first blush, a daunting task. We are asking for a mutual information between two infinite sets of random variables. Appealing to $\mathbf{E} = I[\mathcal{S};\overrightarrow{X}]$, we use the compact representation of the ϵ-machine to reduce one infinite set (the past) to a (usually) finite set. A process's k-crypticity captures something similar about the infinite set of future variables and allows us to further compact our form for excess entropy, reducing an infinite variable set to a finite one. The resulting stratification of process space is a novel way of thinking about its *structure* and, as long as we know which stratum we lie in, we can rapidly calculate many quantities of interest.

Unfortunately, in the general case, one will not know a priori a process's crypticity order. Worse, as far as we are aware, there is no known finite method for calculating the crypticity order. This strikes us as an interesting open problem and challenge.

If, by construction or by some other means, one does know it, then, as we showed, crypticity and $\mathbf{E}$ can be calculated using the crypticity expansion. Failing this, though, one might consider using the expansion to search for the order. There is no known stopping criterion, so this search may not find k in finite time. Moreover, the expansion is a calculation that grows exponentially in computational complexity with crypticity order, as we noted. Devising a stopping criterion would be very useful to such a search.

Even without knowing the k-crypticity, the expansion is often still useful. For use in estimating $\mathbf{E}$, it provides us with a bound from above. This is complementary to the bound below one finds using the typical expansion $\mathbf{E}(L) = H[X_0^L] - h_\mu L$ [17]. Using these upper and lower bounds, one may determine that for a given purpose, the estimate of χ or $\mathbf{E}$ is within an acceptable tolerance.

The crypticity hierarchy is a revealing way to carve the space of processes in that it concerns how they hide internal state information from an observer. The examples were chosen to illustrate several features of this new view. The Even Process, a canonical example of order-∞ Markov, resides instead at the very bottom of this ladder. The two example families show us how k-cryptic is neither a parallel nor independent concept to order-R Markov. Finally, we see in the last example an apparently simple process with ∞-crypticity.

The general lesson is that internal state information need not be immediately available in measurement values, but instead may be spread over long measurement sequences. If a process is k-cryptic and k is finite, then internal state information is accessible over sequences of length k. The existence, as we demonstrated, of processes that are ∞-cryptic is rather sobering. (The Appendix comments on what happens when one fails to appreciate this.) Interpreted as a statement of the impossibility of extracting state information, it reminds us of earlier work on hidden spatial dynamical systems that exhibit a similar encrypting of internal structure in observed spacetime patterns [18].

Due to the exponentially growing computational effort to search for the crypticity order and, concretely, the existence of ∞-cryptic processes, the general theory introduced in Ref. [1] and Ref. [2] is seen to be necessary. It allows one to directly calculate $\mathbf{E}$ and crypticity and to do so efficiently.

Acknowledgments

Chris Ellison was partially supported on a GAANN fellowship. The Network Dynamics Program funded by Intel Corporation also partially supported this work.

Appendix: Crypticity Untamed

Recently, Ref. [19] asserted that a process's $\mathbf{E}$ can be obtained from its ϵ-machine using the following expression:

$$\mathbf{E} = C_\mu - I_{erased} \ ,$$

where $I_{erased} = H[\mathcal{S}_0, X_0] - H[\mathcal{S}_1, X_0]$. Though renamed, I_{erased} is the crypticity of Ref. [1]. However, as we showed in the main development, it is $\chi^+(1)$ and so the above expression is valid only for 0-cryptic and 1-cryptic processes.

Ref. [19] considered only the Even and Golden Mean Processes. These, as we saw, are 0-cryptic and 1-cryptic and so it is no surprise that the expression worked. Indeed, their low-order crypticity is why closed-form expressions for their excess entropies have been known for quite some time, prior to the recent developments.

In short, the claims in Ref. [19] are incorrect. The implication there that all ϵ-machines are 1-cryptic is also. The examples we gave show how wrong such an approximation can be. We showed how large the errors can grow. The full theory of Ref. [1] and Ref. [2] is required. The richness of the space of processes leads us to conjecture that it will suffer no shortcuts.

* Electronic address: jrmahoney@@ucdavis.edu
† Electronic address: cellison@@cse.ucdavis.edu
‡ Electronic address: chaos@@cse.ucdavis.edu

[1] J. P. Crutchfield, C. J. Ellison, and J. Mahoney. Time's barbed arrow: Irreversibility, crypticity, and stored information. *submitted*, 2009. arxiv.org:0902.1209 [cond-mat].
[2] C. J. Ellison, J. R. Mahoney, and J. P. Crutchfield. Prediction, retrodiction, and the amount of information stored in the present. *arxiv: 0905.3587 [cond-mat]*, 2009.
[3] A. Fraser. Chaotic data and model building. In H. Atmanspacher and H. Scheingraber, editors, *Information Dynamics*, volume Series B: Physics Vol. 256 of *NATO ASI Series*, page 125, New York, 1991. Plenum.
[4] M. Casdagli and S. Eubank, editors. *Nonlinear Modeling*, SFI Studies in the Sciences of Complexity, Reading, Massachusetts, 1992. Addison-Wesley.
[5] J. C. Sprott. *Chaos and Time-Series Analysis*. Oxford University Press, Oxford, UK, second edition, 2003.
[6] H. Kantz and T. Schreiber. *Nonlinear Time Series Analysis*. Cambridge University Press, Cambridge, UK, second edition, 2006.
[7] D. Arnold. Information-theoretic analysis of phase transitions. *Complex Systems*, 10:143–155, 1996.
[8] J. P. Crutchfield and D. P. Feldman. Statistical complexity of simple one-dimensional spin systems. *Phys. Rev. E*, 55(2):1239R–1243R, 1997.
[9] D. P. Feldman and J. P. Crutchfield. Discovering noncritical organization: Statistical mechanical, information theoretic, and computational views of patterns in simple one-dimensional spin systems. 1998. Santa Fe Institute Working Paper 98-04-026.
[10] G. Tononi, O. Sporns, and G. M. Edelman. A measure for brain complexity: Relating functional segregation and integration in the nervous system. *Proc. Nat. Acad. Sci. USA*, 91:5033–5037, 1994.
[11] W. Bialek, I. Nemenman, and N. Tishby. Predictability, complexity, and learning. *Neural Computation*, 13:2409–2463, 2001.
[12] W. Ebeling and T. Poschel. Entropy and long-range correlations in literary english. *Europhys. Lett.*, 26:241–246, 1994.
[13] L. Debowski. On the vocabulary of grammar-based codes and the logical consistency of texts. *IEEE Trans. Info. Th.*, 2008.
[14] J. P. Crutchfield and K. Young. Inferring statistical complexity. *Phys. Rev. Let.*, 63:105–108, 1989.
[15] J. P. Crutchfield and C. R. Shalizi. Thermodynamic depth of causal states: Objective complexity via minimal representations. *Phys. Rev. E*, 59(1):275–283, 1999.
[16] C. R. Shalizi and J. P. Crutchfield. Computational mechanics: Pattern and prediction, structure and simplicity. *J. Stat. Phys.*, 104:817–879, 2001.
[17] J. P. Crutchfield and D. P. Feldman. Regularities unseen, randomness observed: Levels of entropy convergence. *CHAOS*, 13(1):25–54, 2003.
[18] J. P. Crutchfield. Unreconstructible at any radius. *Phys. Lett. A*, 171:52 – 60, 1992.
[19] K. Wiesner, M. Gu, E. Rieper, and V. Vedral. Information erasure lurking behind measures of complexity. 2009. arxiv.org:0905.2918 [physics.data-an].