CHAPTER 1

# Information Accessibility and Cryptic Processes

## §1.1 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 [**?**, **?**]. 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 [**?**, **?**, **?**, **?**], in spin systems [**?**, **?**, **?**], in neurobiological systems [**?**, **?**], and even in language [**?**, **?**], 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* [**?**, **?**]. 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 [**?**, **?**]. Here we introduce a classification of processes using a systematic expansion of crypticity. This expansion will lead to a classification of processes orthogonal to that provided by Markov order.

Until recently, **E**, and consequently $\chi$, could not be as directly calculated from the $\epsilon$-machine as the process's entropy rate $h_\mu$ and its statistical complexity. References [**?**] and [**?**] solved this problem, giving a closed-form expression for the excess entropy:

$$\mathbf{E} = I[\mathcal{S}^+; \mathcal{S}^-] \,, \tag{1.1}$$

and an accompanying contructive algorithm 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.

The complementary viewpoint, which we will take in this paper, is also provided by this result. That is, we very straightforwardly have,

$$\begin{aligned} \chi^+ &= H[\mathcal{S}^+ | \overrightarrow{X}] \\ &= C_\mu - \mathbf{E} \\ &= C_\mu - I[\mathcal{S}^+; \mathcal{S}^-] \end{aligned}$$

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

$$\begin{aligned} \chi^+ &= H[\mathcal{S}^+ | \mathcal{S}^-] \text{ or} \\ \chi^- &= H[\mathcal{S}^- | \mathcal{S}^+] \,. \end{aligned}$$

In the following we will not concern ourselves with reverse representations and so can simplify the notation, using $C_\mu$ for $C_\mu^+$ and $\chi$ for $\chi^+$.

Here we show that, for a restricted class of processes, the crypticity in Eqn. **??** can be systematically expanded to give an alternative closed-form to the excess entropy in Eqn. 1.1. One ancillary benefit is a new and, we argue, natural hierarchy of processes in terms of information accessibility.

## §1.2 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 $\overrightarrow{X}_t$. Analogously, the causal state at time $t$ is denoted $\mathcal{S}_t$, and a sequence of states beginning at time $t$ that is $L$ states long is denoted $\mathcal{S}_t^L$.

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

$$H[\mathcal{S}_k | \overrightarrow{X}_0] = 0 \,. \tag{1.2}$$

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

**Definition.** *An* $\infty$-cryptic *process is one for which the process's $\epsilon$-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:

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

□

**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 $\chi_k = \{\mathcal{P} : \mathcal{P}$ is $k$-cryptic and not $(k-1)$-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. □

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. [**?**] and Ref. [**?**]), 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 cryptic order. One way to investigate this is to study the sequence of estimates of $\chi$ at different orders. To this end, we define the $k$-cryptic approximation.

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

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

### §1.2.1 The $k$-Cryptic Expansion

We will now develop a systematic expansion of $\chi$ 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*

$$\chi = \chi(k) + H[\mathcal{S}_k|\overrightarrow{X}_0] \,. \tag{1.3}$$

*Proof.* We calculate directly, starting from the definition, adding and subtracting the $k$-crypticity criterion term from $\chi$'s definition, Eqn. **??**:

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

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 $\epsilon$-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.

□

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

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

*Proof.* Given the order-$k$ expansion of $\chi$ just developed, we now assume the $k$-crypticity criterion is satisfied; viz., $H[\mathcal{S}_k|\overrightarrow{X}_0] = 0$. Thus, we have from Eqn. 1.3:

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

Likewise, assuming $\chi = \chi(k)$ requires, by Eqn. 1.3 that $H[\mathcal{S}_k|\overrightarrow{X}_0] = 0$ and thus the process is $k$-cryptic. □

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

*Proof.*

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

□

## §1.2.2 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 $k$. □

**Corollary 3.** *Once* $\chi(k)$ *reaches the value* $\chi$, $\chi(j) = \chi$ *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, $\chi(j) = \chi$. □

**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 $\chi$ at a finite $k$, remains at that value for all larger $k$.

**Proposition 3.** *The cryptic approximation $\chi(k)$ converges to $\chi$ 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 $\epsilon$ map from pasts to causal states of Eqn. **??**. Therefore, as $k \to \infty$, $\chi(k) \to \chi$. □

### §1.2.3 Block-State and State-Block Entropies

We briefly define two new entropy functions which we will make some use of for a concavity proof. These functions are interesting in their own right. In **ruro2**, we analyze presentations of processes that are *not* the $\epsilon$-machine. That work provides some interesting extensions of the material contained in this chapter. We will focus on the primary results.

**Definition.** *The* block-state entropy function *is defined as $H[X_0^L, \mathcal{S}_L]$.*

**Definition.** *The* state-block entropy function *is defined as $H[\mathcal{S}_0, X_0^L]$.*

**Lemma 3.** *The cryptic approximation is the difference between the state-block and the block-state entropy functions.*

*Proof.*

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

□

**Lemma 4.** *The state-block entropy function is linear for an $\epsilon$-machine. Specifically, it has the form $C_\mu + h_\mu L$.*

*Proof.*

$$\begin{aligned}H[\mathcal{S}_0, X_0^0] &= H[\mathcal{S}_0] = C_\mu \\ H[\mathcal{S}_0, X_0^{L+1}] - H[\mathcal{S}_0, X_0^L] &= H[X_L|\mathcal{S}_0, X_0^L] \\ &= H[X_L|\mathcal{S}_0, X_0^L, \mathcal{S}_L] \\ &= H[X_L|\mathcal{S}_L] \\ &= h_\mu\end{aligned}$$

□

**Lemma 5.** *The subtraction of a convex (concave) function from a linear one is concave (convex).*

*Proof.* Assume that $f(L)$ is convex. Then $(f(L+1) - f(L)) - (f(L) - f(L-1)) \geq 0$. Then let $g(L) = a + bL - f(L)$. Examining the consequence on $g$:

$$\begin{aligned}g(L+1) - g(L) &= a + b(L+1) - f(L+1) - a - bL + f(L) \\ g(L) - g(L-1) &= a + bL - f(L) - a - b(L-1) + f(L-1) \\ (g(L+1) - g(L)) - (g(L) - g(L-1)) &= -(f(L+1) - f(L)) - (f(L) - f(L-1))\end{aligned}$$

□

**Proposition 4.** *The block-state entropy is a convex function.*

*Proof.* The statement of convexity is,

$$H[X_0^{L+1}\mathcal{S}_{L+1}] - H[X_0^L\mathcal{S}_L] \geq H[X_0^L\mathcal{S}_L] - H[X_0^{L-1}\mathcal{S}_{L-1}]$$

By stationarity we have,

$$H[X_{-1}^{L+1}\mathcal{S}_L] - H[X_0^L\mathcal{S}_L] \geq H[X_{-1}^{L}\mathcal{S}_{L-1}] - H[X_0^{L-1}\mathcal{S}_{L-1}]$$

Equivalently,

$$H[X_{-1}|X_0^L\mathcal{S}_L] \geq H[X_{-1}|X_0^{L-1}\mathcal{S}_{L-1}]$$

We can use an I-diagram to help understand the convexity statement (Fig. 1.10). The convexity is now translated to,

$$\alpha + \gamma \geq \alpha + \beta$$

$$\gamma \geq \beta$$

Using the fact that the causal state is an optimal representation of the past, we have the following

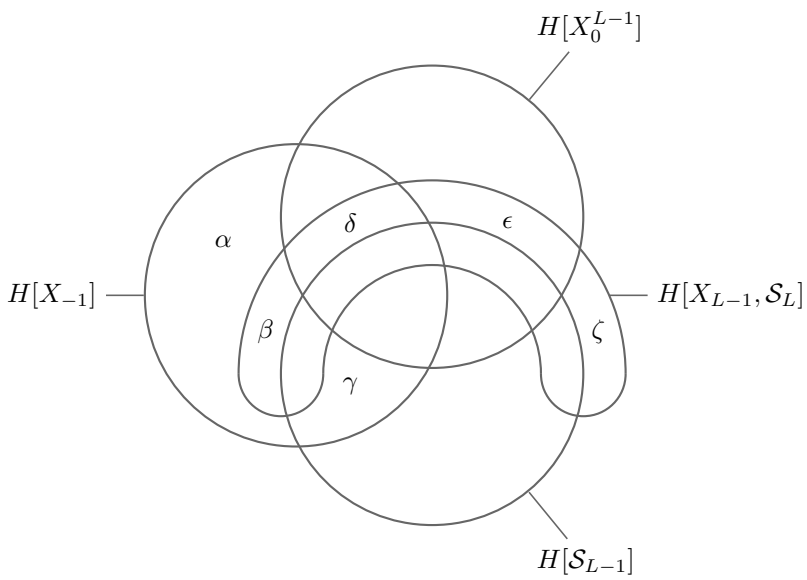


Figure 1.1: An I-diagram helps to organize the algebra. Note that we reduce the complexity of this diagram by making two of the variables aggregate variables. Also, we have opted for an alternate representation of the I-diagram keeping three of the regions circular.

equivalent expressions for the entropy rate:

$$h_\mu =$$

$$H[X_{L-1}\mathcal{S}_L|\mathcal{S}_{L-1}] = \beta + \epsilon + \delta + \zeta$$

$$H[X_{L-1}\mathcal{S}_L|\mathcal{S}_{L-1}X_0^{L-1}] = \beta + \zeta$$

$$H[X_{L-1}\mathcal{S}_L|\mathcal{S}_{L-1}X_{-1}] = \epsilon + \zeta$$

$$H[X_{L-1}\mathcal{S}_L|\mathcal{S}_{L-1}X_{-1}X_0^{L-1}] = \zeta$$

Note that we relied on the shielding property of the causal states, and also the unifilarity of the $\epsilon$-machine. These four relations together yield,

$$\zeta = h_\mu$$

$$\beta = \delta = \epsilon = 0$$

And combining, we see that the convexity statement is simply $\gamma \geq 0$. Since $\gamma$ is a conditional mutual information, and is therefore positive semidefinite, we have shown that the block-state entropy function is convex. □

**Corollary 7.** *The cryptic approximation $\chi(k)$ is a concave function.*

*Proof.* Since the cryptic approximation is the difference between the linear state-block entropy and the convex block-state entropy, as shown in the previous proofs, it is a trivial consequence that $\chi(k)$ is concave. □

**Proposition 5.**

*Proof.* □

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

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

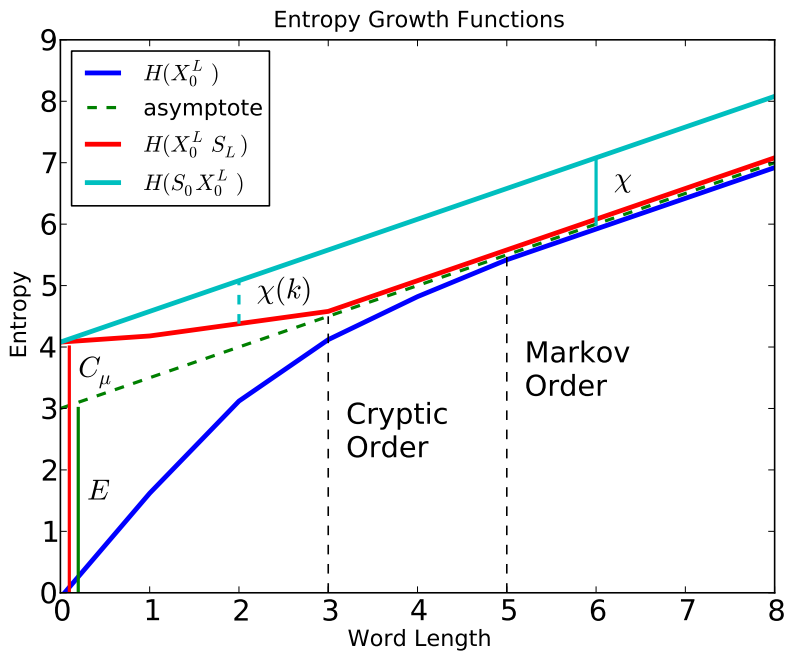


Figure 1.2: The entropy growth functions: block entropy $H[X_0^L]$, block-state entropy $H[X_0^L, \mathcal{S}_L]$, and state-block entropy $H[\mathcal{S}_0, X_0^L]$ provide a convenient way for understanding several of a process's properties. Previously, the entropy rate, excess entropy, and Markov order were seen on this diagram. We now add statistical complexity, crypticity, and cryptic order to that list. A pleasing feature of this figure is that it reproduces the I-diagram in Fig. **??** when viewed end on. To draw the analogy explicitly, just as Markov order describes the scale at which the block entropy $H[X_0^L]$ linearizes, the cryptic order describes the scale at which the block-state entropy $H[X_0^L, \mathcal{S}_L]$ linearizes.

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

**Corollary 9.** *A process $\mathcal{P}$ is* 0*-cryptic* if *and* only if $\mathbf{E} = C_\mu$.

*Proof.* If $\mathcal{P}$ is 0-cryptic, then $\mathbf{E} = C_\mu - \chi(0)$ and Cor. 2 says that $\chi(0) = 0$. To establish the opposite direction, note that $\mathbf{E} = C_\mu$ implies $\chi = 0$. Applying Cor. 2 shows $\chi = \chi(0)$, and so the process is 0-cryptic by Cor. 1. □

## §1.2.5 Crypticity of Spin Chains

Now, we provide results on the crypticity of one-dimensional spin chains to complement prior results on Markovity and excess entropy. First recall Eqn. **??**, which gives the excess entropy for

order-$R$ Markov processes:

$$\mathbf{E} = H[X_0^R] - R\, h_\mu \,.$$

By Prop. 1, such processes are also $R$-cryptic and so:

$$\mathbf{E} = C_\mu - \chi(R)\,.$$

One-dimensional spin chains are precisely those order-$R$ Markov processes for which the statistical complexity, $C_\mu \equiv H[S_R]$, equals the entropy over $R$-blocks, $H[X_0^R]$. Reference [**?**] stated a condition under which equality held in terms of transfer matrices. Here, we state a simpler condition by equating two chain-rule expansions of $H[X_0^R, \mathcal{S}_R]$:

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

Since the process is Markov, $H[\mathcal{S}_R|X_0^R] = 0$ and thus:

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

In words, spin chains are processes for which there exists a one-to-one correspondence between the $R$-blocks and the causal states, confirming the interpretation specified in Ref. [**?**].

The above equations also show that spin chains have $\chi(R) = Rh_\mu$. Here we provide another proof:

**Proposition 6.**

$$H[X_0^R|\mathcal{S}_R] = 0 \quad \Longleftrightarrow \quad \chi(R) = R\, h_\mu\,, \tag{1.4}$$

*where $h_\mu$ is the process's entropy rate.*

*Proof.* The proof is a direct calculation:

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

□

Furthermore, we actually have the form of the entire $\chi(L)$ function for spin chains. Since $H[X_0^L|\mathcal{S}_L] = 0$, we can marginalize and get $H[X_0^i|\mathcal{S}_i]$ for $0 \leq i \leq R$. Then, $H[X_0^L, \mathcal{S}_L] = C_\mu$ for

$0 \leq i \leq R$. The crypticity approximation, which is the difference between the state-block and block-state, is then $C_\mu + h_\mu L - C_\mu = h_\mu L$ for $0 \leq i \leq R$. Since $\chi(L)$ is nondecreasing and bounded by $\chi$, which in this case is $h_\mu R$, the full form of the crypticity approximation for spin chains is simply:

$$\chi(L) = \begin{cases} h_\mu L & \text{if } 0 \leq L \leq R, \\ h_\mu R & \text{if } x \geq R. \end{cases}$$

**Proposition 7.** *Periodic processes are* 0*-cryptic.*

*Proof.* Periodic processes are order-$R$ Markov spin chains, so $\mathbf{E} = C_\mu - Rh_\mu$. Since $h_\mu = 0$, $\mathbf{E} = C_\mu$. By Cor. 9 the process is 0-cryptic. □

**Proposition 8.** *An order-R spin chain with positive entropy rate 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. 6, 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 Eqn. **??** 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 9.** *An order-R spin chain with positive entropy rate is not* $k$*-cryptic for any* $0 \leq k < R$.

*Proof.* By Lem. 2, if the process where $k$-cryptic for some $0 \leq k < R$, then it would also be $(R-1)$-cryptic. By Prop. 8, this is not true. Therefore, the primitive orders of Markovity and crypticity are the same. □

### §1.2.6 Geometric Constraints

Returning to the class of general $\epsilon$-machines, there are a couple of things that can be said about the cryptic order that, when viewed in the context of block entropies, have a very geometric feel to them.

**Proposition 10.** *The cryptic order for a process with a known $\chi$ and $h_\mu$ is bounded below by* $\lceil \frac{\chi}{h_\mu} \rceil = \lceil \frac{C_\mu - \mathbf{E}}{h_\mu} \rceil$.

*Proof.* The nondecreasing block-state entropy can only intersect the block entropy linear asymptote $\mathbf{E} + h_\mu L$ at $L \geq \frac{\chi}{h_\mu}$, and the ceiling just takes into account the fact that we are only interested in integer word lengths. It is this intersection which defines the linearization of the block-state entropy and which establishes the cryptic order (see Fig. 1.3).

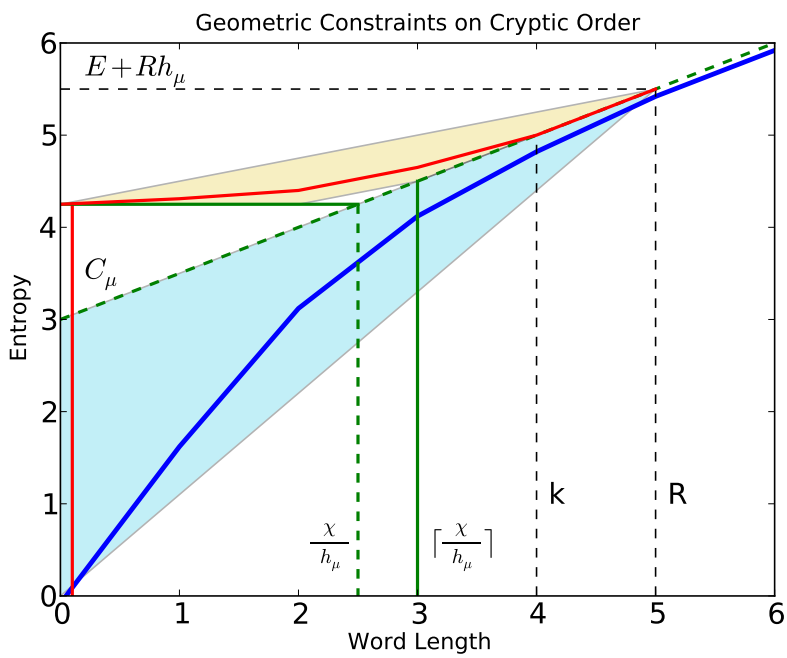


Figure 1.3: Given only $\chi$, and $h_\mu$, we can already place a lower bound on the cryptic order. The upper bound of course comes from the Markov order. The cyan and beige shaded areas indicate the bounds on the block and block-state entropy functions respectively. Notice the small area missing near the center of the diagram. This is a result of the block-state function being defined only for integers.

□

A specific case of this is when $H[X_0^{R-1}] < C_\mu \leq H[X_0^R]$. Since the cryptic order is discrete and bounded above by the Markov order R the $H[X_0^L \mathcal{S}_L]$ curve is forced to meet the asymptote at $R$ and thus the cryptic order is also $R$.

**Proposition 11.** *The cryptic order for a process with a known $\chi$, $h_\mu$ and Markov order $R$ is bounded below by* $\min(\lfloor \frac{\chi}{h_\mu} \rfloor, R)$.

*Proof.* This is just a slight modification of the previous result. First observe that for an order-$R$ Markov process, $(H[X_0^L \mathcal{S}_L] = C_\mu, \forall 0 \geq L \geq R) \iff (C_\mu = H[X_0^R])$. In words, spin chains are the only processes that have a constant block-state entropy function (for $0 \leq L \leq R$). This implies that if, in addition to $\chi$ and $h_\mu$, the Markov order $R$ is known, that $\min(\lfloor \frac{\chi}{h_\mu} \rfloor + 1, R)$ is actually a slightly better lower bound on cryptic order. □

## §1.3 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 $\epsilon$-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. [**?**, **?**, **?**].

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 $\infty$-cryptic.

It should be pointed out, though, that these examples were hand-chosen to illustrate some of the range of possible processes in terms of cryptic and Markov orders. If one were to encounter a process in the wild, its cryptic order would not be known and the calculation of crypticity would require that one determines the cryptic order. One can estimate the cryptic order by calculating the cryptic approximation until it appears to have converged or computational power has run out. Alternatively, one might deduce the order exactly via some other technique, as we do in the upcoming examples. Of course, we wish to note that Ref. [**?**] demonstrates how to calculate $\chi$ without any knowledge of the cryptic order.

### §1.3.1 Even Process: 0-Cryptic

Figure 1.4 gives the $\epsilon$-machine for the Even Process. The Even Process produces binary sequences in which all blocks of uninterrupted 1s are even in length, bounded by 0s. Further, after each even length is reached, there is a probability $p$ of breaking the block of 1s by inserting one or more 0s.

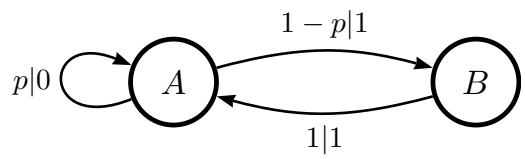


Figure 1.4: A 0-cryptic process: Even Process. The transitions denote the probability $p$ of generating symbol $x$ as $p|x$.

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

To see why the Even Process is 0-cryptic, first note that the semi-infinite string $\overrightarrow{X}_0 = 1, 1, 1 \ldots$ occurs with probability zero. So with probability one, a given future will have only a finite number of 1s before a 0 is seen. Once the 0 is seen, it is straightforward to count the number of 1s preceding it. If the number of 1s is even, then $\mathcal{S}_0$, the causal state that preceded this future, is $A$. Otherwise, it is $B$. In either case, we know the causal state with certainty, and so, $H[\mathcal{S}_0|\overrightarrow{X}_0] = 0$.

It is important to note that this process is *not* order-$R$ Markov for any finite $R$ [**?**]. 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.

### §1.3.2 Golden Mean Process: 1-Cryptic

Figure 1.5 shows the $\epsilon$-machine for the Golden Mean Process [**?**]. 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 $\epsilon$-machine looks remarkably similar to that for the Even Process. The informational analysis, however, shows that they have markedly different properties.

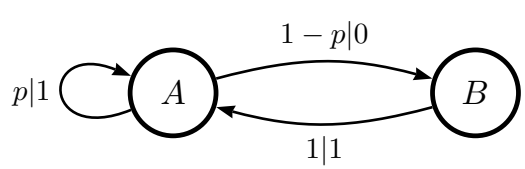


Figure 1.5: A 1-cryptic process: Golden Mean Process.

Reference [**?**] showed that the Golden Mean Process has the same statistical complexity and entropy rate as the Even Process: $C_\mu = H\left(1/(2-p)\right)$ 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, excess entropy $\mathbf{E} = \log_2(3) - \frac{4}{3}$ bits, and crypticity $\chi = \frac{2}{3}$ bits. Thus, the excess entropy differs from that of the Even Process. (As with the Even Process, these closed-form expressions for $C_\mu$ 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. [**?**], Eqn. **??** here, applies for an $R = 1$ Markov chain.

### §1.3.3 Butterfly Process: 2-Cryptic

The next example, the Butterfly Process of Fig. 1.6, 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 $\epsilon$-machine's structure. Most of the explanation does not require calculating much, if anything.

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 of futures beginning with 01.

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* two-symbol word $(X_0, X_1)$ 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 $\chi$ using Cor. 1:

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

From Cor. 8, 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.

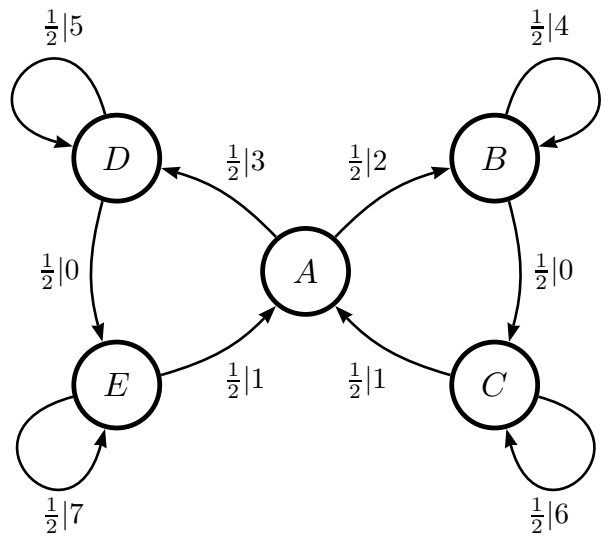


Figure 1.6: A 2-cryptic process: Butterfly Process over a 6-symbol alphabet.

### §1.3.4 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. 1.7 for its $\epsilon$-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 since each word of length $k$ induces just one causal state. Proposition 1 then implies it is (at most) $k$-cryptic. In order to show that it is not $(k-1)$-cryptic, consider the case $\overrightarrow{x}_0 = 1^k 0^\infty$. The first $(k-1)$ 1s will induce a mixture over states $k$ and 0. The following future $\overrightarrow{x}_k = 10^\infty$ is consistent with both states $k$ and 0. Therefore, the $(k-1)$-crypticity criterion is not satisfied. Therefore, it is $k$-cryptic.

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. 8:

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

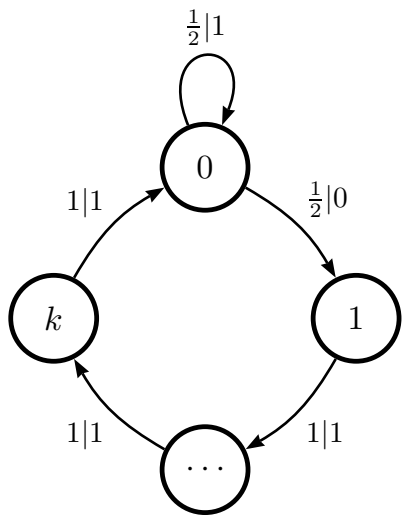


Figure 1.7: $k$-cryptic processes: Restricted Golden Mean Family.

which diverges with $k$. (Calculational details are found in Ref. [**?**].)

## §1.3.5 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 $k$. The $\epsilon$-machine for this process is shown in Fig. 1.8.

Again, it is straightforward to see that this process is order-$k$ Markov. To see that it is not 0-cryptic, note that:

$$\begin{aligned}
H[\mathcal{S}_0|\overrightarrow{X}_0] &= H[\mathcal{S}_0|X_0 = 0, \overrightarrow{X}_1] + H[\mathcal{S}_0|X_0 = 1, \overrightarrow{X}_1] \\
&\geq H[\mathcal{S}_0|X_0 = 1, \overrightarrow{X}_1] \\
&= \frac{2}{k+2} \sum_{\overrightarrow{x}_1} H[\mathcal{S}_0|X_0 = 1, \overrightarrow{X}_1 = \overrightarrow{x}_1] \\
&\geq \frac{2}{k+2} H[\mathcal{S}_0|\overrightarrow{X}_1 = 1^\infty] \\
&= \frac{2}{k+2} \\
&> 0 \,.
\end{aligned}$$

To see that this family 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 $\overrightarrow{x}_1 = 1^\infty$, then $\mathcal{S}_1 = k$. Similarly, if the future $\overrightarrow{x}_1 = 0^n 1^\infty$, then $\mathcal{S}_1 = k - n$.

This family provides an example for which the cryptic order is strictly less than the Markov order. In this case, the cryptic order is fixed at 1 for all $k$, while the Markov order is $k$. Note that the separation between the Markov and cryptic order can grow arbitrarily large and, thus, the two properties are clearly not redundant.

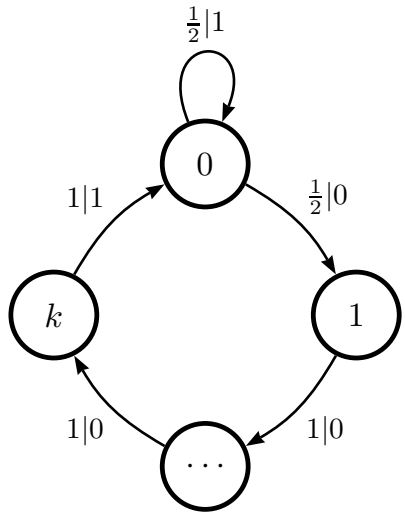


Figure 1.8: $k$-cryptic processes: Stretched Golden Mean Family.

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,

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

## §1.3.6 Nemo Process: ∞-Cryptic

We close our cryptic process bestiary with a (very) finite-state process that has infinite cryptic order: 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 $\epsilon$-machine is shown in Fig. 1.9.

Its stationary state distribution is

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

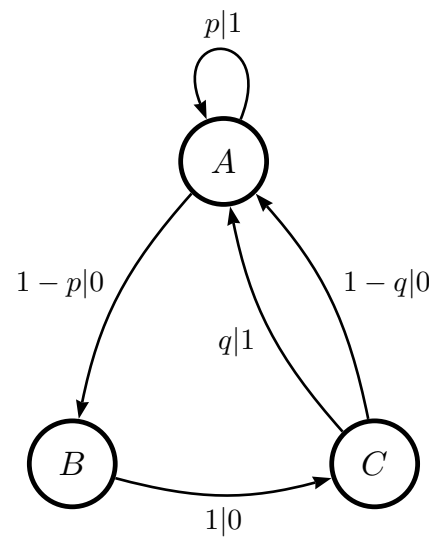


Figure 1.9: The $\infty$-cryptic Nemo Process.

from which one calculates the statistical complexity:

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

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 $\epsilon$-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 [**?**], shielding the past from

the future. Each of these quantities is given by:

$$\Pr(X_k^L = w | \mathcal{S}_k = \sigma) = [T^{(w)}\mathbf{1}]_\sigma$$

$$\Pr(X_0^k = w, \mathcal{S}_k = \sigma) = [\pi T^{(w)}]_\sigma$$

$$\Pr(X_0^L = w) = \pi T^{(w)}\mathbf{1}\,.$$

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 a 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:

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

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 $\mathcal{S}_{3n}$ 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:

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

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

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

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 $\infty$-cryptic. This means that we cannot make use of the

$k$-cryptic approximation to calculate $\chi$ or $\mathbf{E}$.

Fortunately, the techniques introduced in Refs. [**?**] and [**?**] do not rely on an approximation method. To avoid ambiguity, denote the statistical complexity we just computed as $C_\mu^+$. When those 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 are provided in Ref. [**?**].)

## §1.4 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 $\epsilon$-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 in which stratum we lie, we can rapidly calculate many quantities of interest.

Unfortunately, in the general case, one will not know a priori a process's cryptic order. Worse, as far as we are aware, there is no known finite method for calculating the cryptic 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

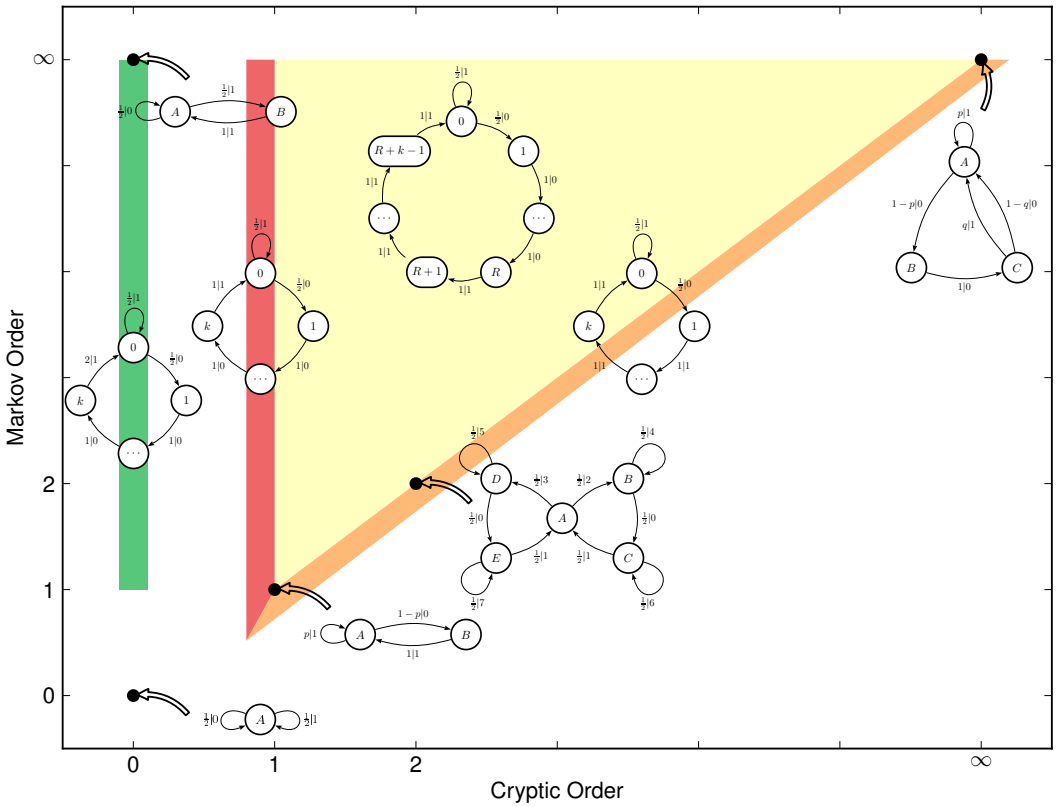


Figure 1.10: This figure shows a bird's-eye view of process space. Some sample processes were chosen and placed on a plot of Markov vs cryptic order. Some $\epsilon$-machines point to particular points in the space while others are parameterized $\epsilon$-machines and refer to a colored region. We can readily see that aside from the bound $R \geq k$, the space is filled. This means that the cryptic order is a nontrivial complement to Markov order.

computational complexity with cryptic 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 **E**, it provides us with a bound from above. This is complementary to the lower bound one finds using the typical expansion $\mathbf{E}(L) = H[X_0^L] - h_\mu L$ [**?**]. Using these upper and lower bounds, one may determine that for a given purpose, the estimate of $\chi$ or **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-$\infty$ 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 $\infty$-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 $\infty$-cryptic is rather sobering. 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 [**?**].

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