Detecting emergent processes in cellular automata with excess information

David Balduzzi[1]

[1] Department of Empirical Inference, MPI for Intelligent Systems, Tübingen, Germany
david.balduzzi@tuebingen.mpg.de

Abstract

Many natural processes occur over characteristic spatial and temporal scales. This paper presents tools for (i) flexibly and scalably coarse-graining cellular automata and (ii) identifying which coarse-grainings express an automaton's dynamics well, and which express its dynamics badly. We apply the tools to investigate a range of examples in Conway's Game of Life and Hopfield networks and demonstrate that they capture some basic intuitions about emergent processes. Finally, we formalize the notion that a process is emergent if it is better expressed at a coarser granularity.

Introduction

Biological systems are studied across a range of spatiotemporal scales – for example as collections of atoms, molecules, cells, and organisms (Anderson, 1972). However, not all scales express a system's dynamics equally well. This paper proposes a principled method for identifying which spatiotemporal scale best expresses a cellular automaton's dynamics. We focus on Conway's Game of Life and Hopfield networks as test cases where collective behavior arises from simple local rules.

Conway's Game of Life is a well-studied artificial system with interesting behavior at multiple scales (Berlekamp et al., 1982). It is a 2-dimensional grid whose cells are updated according to deterministic rules. Remarkably, a sufficiently large grid can implement any deterministic computation. Designing patterns that perform sophisticated computations requires working with distributed structures such as gliders and glider guns rather than individual cells (Dennett, 1991). This suggests grid computations may be better expressed at coarser spatiotemporal scales.

The first contribution of this paper is a coarse-graining procedure for expressing a cellular automaton's dynamics at different scales. We begin by considering cellular automata as collections of spacetime coordinates termed occasions (cell n_i at time t). Coarse-graining groups occasions into structures called *units*. For example a unit could be a 3×3 patch of grid containing a glider at time t. Units do not have to be adjacent to one another; they interact through *channel* – transparent occasions whose outputs are marginalized over. Finally, some occasions are set as *ground*, which fixes the initial condition of the coarse-grained system.

Gliders propagate at 1/4 diagonal squares per tic – the grid's "speed of light". Units more than $4n$ cells apart cannot interact within n tics, imposing constraints on which coarse-grainings can express glider dynamics. It is also intuitively clear that units should group occasions concentrated in space and time rather than scattered occasions that have nothing to do with each other. In fact, it turns out that most coarse-grainings express a cellular automaton's dynamics badly.

The second contribution of this paper is a method for distinguishing good coarse-grainings from bad based on the following principle:

- *Coarse-grainings that generate more information, relative to their sub-grainings, better express an automaton's dynamics than those generating less.*

We introduce two measures to quantify the information generated by coarse-grained systems. Effective information, ei, quantifies how selectively a system's output depends on its input. Effective information is high if few inputs cause the output, and low if many do. Excess information, ξ, measures the difference between the information generated by a system and its subsystems.

With these tools in hand we investigate coarse-grainings of Game of Life grids and Hopfield networks and show that grainings with high ei and ξ capture our basic intuitions regarding emergent processes. For example, excess information distinguishes boring (redundant) from interesting (synergistic) information-processing, exemplified by blank patches of grid and gliders respectively.

Finally, the penultimate section converts our experience with examples in the Game of Life and Hopfield networks into a provisional formalization of the principle above. Roughly, we define a process as *emergent* if it is better expressed at a coarser scale.

The principle states that emergent processes are more than the sum of their parts – in agreement with many other approaches to quantifying emergence (Crutchfield, 1994;

Tononi, 2004; Polani, 2006; Shalizi and Moore, 2006; Seth, 2010). Two points distinguishing our approach from prior work are worth emphasizing. First, coarse-graining is *scalable*: coarse-graining a cellular automaton yields another cellular automaton. Prior works identify macro-variables such as temperature (Shalizi and Moore, 2006) or centre-of-mass (Seth, 2010) but do not show how to describe a system's dynamics purely in terms of these macro-variables. By contrast, an emergent coarse-graining is itself a cellular automaton, whose dynamics are computed via the mechanisms of its units and their connectivity (see below).

Second, our starting point is selectivity rather than predictability. Assessing predictability necessitates building a model and deciding what to predict. Although emergent variables may be robust against model changes (Seth, 2010), it is unsatisfying for emergence to depend on properties of *both* the process *and* the model. By contrast, effective and excess information depend only on the process: the mechanisms, their connectivity, and their output. A process is then emergent if its internal dependencies are best expressed at coarse granularities.

Probabilistic cellular automata

Concrete examples. This paper considers two main examples of cellular automata: Conway's Game of Life and Hopfield networks (Hopfield, 1982).

The Game of Life is a grid of deterministic binary cells. A cell outputs 1 at time t iff: (i) three of its neighbors outputted 1s at $t-1$ or (ii) it and two neighbors outputted 1s at $t-1$.

In a Hopfield network (Amit, 1989), cell n_k fires with probability proportional to

$$p(n_{k,t}=1|n_{\bullet,t-1}) \propto \exp\left[\frac{1}{T}\sum_{j\to k}\alpha_{jk}\cdot n_{j,t-1}\right] \qquad (1)$$

Temperature T controls network stochasticity. Attractors $\{\xi^1,\ldots,\xi^N\}$ are embedded into a network by setting the connectivity matrix as $\alpha_{jk}=\sum_{\mu=1}^{N}(2\xi_j^\mu-1)(2\xi_k^\mu-1)$.

Abstract definition. A *cellular automaton* is a finite directed graph X with vertices $V_X=\{v_1\ldots v_n\}$. Vertices are referred to as occasions; they correspond to *spacetime* coordinates in concrete examples. Each occasion $v_l\in V_X$ is equipped with finite output alphabet A_l and Markov matrix (or *mechanism*) $p_l(a_l|s_l)$, where $s_l\in S_l=\prod_{k\to l}A_k$, the combined alphabet of the occasions targeting v_l. The mechanism specifies the probability that occasion v_l chooses output a_l given input s_l. The input alphabet of the entire automaton X is the product of the alphabets of its occasions $X_{in}:=\prod_{l\in V_X}A_l$. The output alphabet is $X_{out}=X_{in}$.

Remark. The input X_{in} and output X_{out} alphabets are distinct copies of the same set. Inputs are causal interventions imposed via Pearl's $do(-)$ calculus (Pearl, 2000). The probability of output a_l is computed via the Markov matrix: $p_l\big(a_l|do(s_l)\big)$. The $do(-)$ is not included in the notation explicitly to save space. However, it is always implicit when applying any Markov matrix.

A Hopfield network over time interval $[\alpha,\beta]$ is an abstract automaton. Occasions are spacetime coordinates – e.g. $v_l=n_{i,t}$, cell i at time t. An edge connects $v_k\to v_l$ if there is a connection from v_k's cell to v_l's and the time coordinates are $t-1$ and t respectively for some t. The mechanism is given by Eq. (1). Occasions at $t=\alpha$, with no incoming edges, can be set as fixed initial conditions or noise sources. Similar considerations apply to the Game of Life.

Non-Markovian automata (whose outputs depend on inputs over multiple time steps) have edges connecting occasions separated by more than one time step.

Coarse-graining

Define a *subsystem* X of cellular automaton Y as a subgraph containing a subset of Y's vertices and a subset of the edges targeting those vertices. We show how to coarse-grain X.

Definition (coarse-graining)**.** *Let X be a subsystem of Y. The coarse-graining algorithm detailed below takes $X\subset Y$ and data $\mathcal{K}$ as arguments, and produces new cellular automaton $X_{\mathcal{K}}$. Data $\mathcal{K}$ consists of (i) a partition of X's occasions $V_X=\mathbf{G}\cup\mathbf{C}\cup\mathbf{U}_1\cup\cdots\cup\mathbf{U}_N$ into ground $\mathbf{G}$, channel $\mathbf{C}$ and units $\mathbf{U}_1\ldots\mathbf{U}_N$ and (ii) ground output $s^{\mathbf{G}}$.*

Vertices of automaton $X_{\mathcal{K}}$, the new coarse-grained occasions, are units: $V_{X_{\mathcal{K}}}:=\{\mathbf{U}_1\ldots\mathbf{U}_N\}$. The directed graph of $X_{\mathcal{K}}$ is computed in Step 4 and the alphabets $\mathbf{A}_l$ of units $\mathbf{U}_l$ are computed in Step 5. Computing the Markov matrices (mechanisms) of the units takes all five steps.

The ground specifies occasions whose outputs are fixed: the initial condition $s^{\mathbf{G}}$. The channel specifies unobserved occasions: interactions between units propagate across the channel. Units are macroscopic occasions whose interactions are expressed by the coarse-grained automaton. Fig. 1 illustrates coarse-graining a simple automaton.

There are no restrictions on partitions. For example, although the ground is intended to provide the system's initial condition, it can contain any spacetime coordinates so that in pathological cases it may obstruct interactions between units. Distinguishing good coarse-grainings from bad is postponed to later sections.

Algorithm. Apply the following steps to coarse-grain:

Step 1. *Marginalize over extrinsic inputs.*

External inputs are treated as independent noise sources; we are only interested in *internal* information-processing. An occasion's input alphabet decomposes into a product $S_l=S_l^X\times S_l^{Y\setminus X}$ of inputs from within and without the system. For each occasion $v_l\in V_X$, marginalize over external outputs using the uniform distribution:

$$p_l\big(a_l\big|s_l^X\big):=\sum_{S_l^{Y\setminus X}}p_l\big(a_l\big|s_l^X,s_l^{Y\setminus X}\big)\cdot p_{unif}(s_l^{Y\setminus X}). \qquad (2)$$

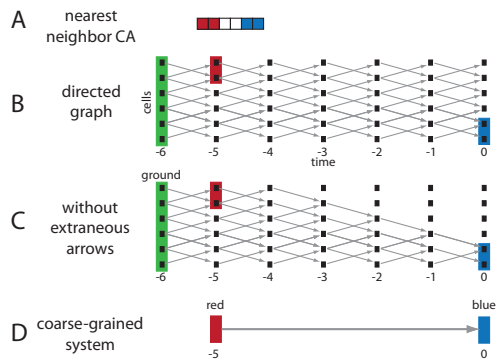

Figure 1: (A) An automaton of 6 cells connected to their immediate neighbors. (B): The directed graph of occasions over time interval $[-6, 0]$. Green occasions are ground. Red and blue occasions form two units. Other occasions are channel. (C): Edges whose signals do not reach the blue unit have no effect. (D): The coarse-grained system consists of two units (macro-occasions).

Step 2. *Fix the ground.*

Ground outputs are fixed in the coarse-grained system. Graining $\mathcal{K}$ imposes a second decomposition onto v_l's input alphabet, $S_l^X = S_l^{\mathbf{G}} \times S_l^{\mathbf{C}} \times S_l^{\mathbf{U}}$ where $\mathbf{U} = \cup_k \mathbf{U}_k$. Subsume the ground into v_l's mechanism by specifying

$$p_l^{\mathbf{G}}\left(a_l \middle| s_l^{\mathbf{C}}, s_l^{\mathbf{U}}\right) := p_l\left(a_l \middle| s_l^{\mathbf{G}}, s_l^{\mathbf{C}}, s_l^{\mathbf{U}}\right).$$

Step 3. *Marginalize over the channel.*

The channel specifies transparent occasions. Perturbations introduced into units propagate through the channel until they reach other units where they are observed. Transparency is imposed by marginalizing over the channel occasions in the product mechanism

$$p_{\mathcal{K}}(x_{out}^{\mathcal{K}} | x_{in}^{\mathcal{K}}) := \sum_{l \in \mathbf{C}} \prod_{l \in \mathbf{C} \cup \mathbf{U}} p_l^{\mathbf{G}}\left(x_{out}^l | x_{in}^l\right), \qquad (3)$$

where superscripts denote that inputs and outputs are restricted, for $\mathcal{K}$, to occasions in units in $\mathcal{K}$ (since channel is summed over and ground is already fixed) and, for each l, to the inputs and outputs of occasion v_l.

For example, consider cellular automaton with graph $v_a \to v_b \to v_c$ and product mechanism $p(c|b)p(b|a)p(a)$. Setting v_b as channel and marginalizing yields coarse-grained mechanism $\sum_b p(c|b)p(b|a)p(a) = p(c|a)p(a)$. The channel is rendered transparent and new mechanism $p(c|a)$ convolves $p(c|b)$ and $p(b|a)$.

Step 4. *Compute the effective graph of coarse-graining $X_{\mathcal{K}}$.*

The micro-alphabet of unit $\mathbf{U}_l$ is $\tilde{\mathbf{A}}_l := \prod_{k \in \mathbf{U}_l} A_k$. The mechanism of $\mathbf{U}_l$ is computed as in Eq. (3) with the product restricted to occasions $j \in \mathbf{C} \cup \mathbf{U}_l$, thus obtaining $p_{\mathbf{U}_l}(a_l | x_{in})$ where $a_l \in \tilde{\mathbf{A}}_l$.

Two units $\mathbf{U}_k$ and $\mathbf{U}_l$ are connected by an edge if the outputs of $\mathbf{U}_k$ make a difference to the behavior of $\mathbf{U}_l$. More precisely, we draw an edge if $\exists a_k, a_k' \in \tilde{\mathbf{A}}_k$ such that

$$p_{\mathbf{U}_l}(a_l | \overline{x_{in}}, a_k) \neq p_{\mathbf{U}_l}(a_l | \overline{x_{in}}, a_k') \text{ for some } a_l \in \tilde{\mathbf{A}}_l.$$

Here, $\overline{x_{in}}$ denotes the input from all units other than $\mathbf{U}_k$.

The effective graph need not be acyclic. Intervening via the $do(-)$ calculus allows us to work with cycles.

Step 5. *Compute macro-alphabets of units in $X_{\mathcal{K}}$.*

Coarse-graining can eliminate low-level details. Outputs that are distinguishable at the base level may not be after coarse-graining. This can occur in two ways. Outputs b and b' have indistinguishable effects if $p(a|b, c) = p(a|b', c)$ for all a and c. Alternatively, two outputs react indistinguishably if $p(b|c) = p(b'|c)$ for all c.

More precisely, two outputs u_l and u_l' of unit $\mathbf{U}_l$ are equivalent, denoted $u_l \sim_{\mathcal{K}} u_l'$, iff

$$p_{\mathcal{K}}(x_{out} | \overline{x_{in}}, u_l) = p_{\mathcal{K}}(x_{out} | \overline{x_{in}}, u_l') \text{ and}$$
$$p_{\mathbf{U}_l}(u_l | x_{in}^{\mathcal{K}}) = p_{\mathbf{U}_l}(u_l' | x_{in}^{\mathcal{K}}) \text{ for all } x_{out}, x_{in}.$$

Picking a single element from each equivalence class obtains the macro-alphabet $\mathbf{A}_l$ of the unit $\mathbf{U}_l$. The mechanism of $\mathbf{U}_l$ is $p_{\mathbf{U}_l}$, Step 4, restricted to macro-alphabets.

Information

This section extends prior work to quantify the information generated by a cellular automaton, both as a whole and relative to its subsystems (Balduzzi and Tononi, 2008, 2009).

Given subsystem $\mathfrak{m}$ of X, let $p_{\mathfrak{m}}(x_{out}|x_{in})$, or $\mathfrak{m}$ for short, denote its *mechanism* or Markov matrix. The mechanism is computed by taking the Markov matrix of each occasion in X, marginalizing over extrinsic inputs (edges not in X) as in Eq. (2), and taking the product. It is notationally convenient to write $p_{\mathfrak{m}}$ as though its inputs and outputs are x_{out} and x_{in}, even though $\mathfrak{m}$ does not in general contain all occasions in X and therefore treats some inputs and outputs as extrinsic, unexplainable noise. We switch freely between terms "subsystem" and "submechanism" below.

Effective information quantifies how selectively a mechanism discriminates between inputs when assigning them to an output. Alternatively, it measures how sharp the functional dependencies leading to an output are.

The *actual repertoire* $\hat{p}_{\mathfrak{m}}(X_{in}|x_{out})$ is the set of inputs that cause (lead to) mechanism $\mathfrak{m}$ choosing output x_{out}, weighted by likelihood according to Bayes' rule

$$\hat{p}_{\mathfrak{m}}\left(x_{in} | x_{out}\right) := \frac{p_{\mathfrak{m}}\left(x_{out} | do(x_{in})\right)}{p(x_{out})} \cdot p_{unif}(x_{in}). \qquad (4)$$

The $do(-)$ notation and hat $\hat{p}$ remind that we first *intervene* to impose x_{in} and then apply Markov matrix $p_{\mathfrak{m}}$.

For deterministic mechanisms, i.e. functions $f : X_{in} \to X_{out}$, the actual repertoire assigns $\hat{p} = \frac{1}{|f^{-1}(x_{out})|}$ to elements of the pre-image and $\hat{p} = 0$ to other elements of X_{in}.

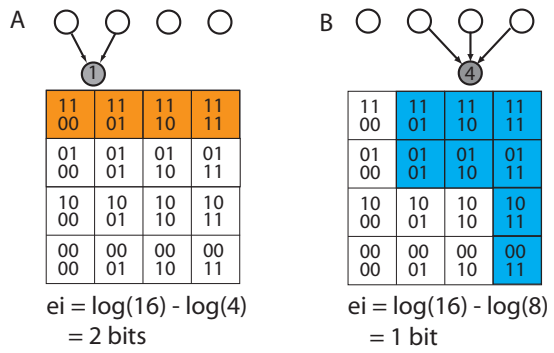

Figure 2: Categorization and information. Cells fire if they receive two or more spikes. The $16 = 2^4$ possible outputs by the top layer are arranged in a grid. (AB): Cells n_1 and n_4 fire when the output is in the orange and blue regions respectively. Cell n_1's response is more informative than n_4's since it fires for fewer inputs.

The shaded regions in Fig. 2 show outputs of the top layer that cause the bottom cell to fire.

Effective information generated when $\mathfrak{m}$ outputs x_{out} is Kullback-Leibler divergence ($KL[p\|q] = \sum_i p_i \log_2 \frac{p_i}{q_i}$),

$$ei(\mathfrak{m}, x_{out}) := KL\Big[\hat{p}_{\mathfrak{m}}\big(X_{in}|x_{out}\big)\Big\|p_{unif}(X_{in})\Big]. \quad (5)$$

Effective information is *not* a statistical measure: it depends on the mechanism and a *particular* output x_{out}.

Effective information generated by deterministic function f is $ei(f, x_{out}) = \log_2 \frac{|X_{in}|}{|f^{-1}(x_{out})|}$ where $|\cdot|$ denotes cardinality. In Fig. 2, ei is the logarithm of the ratio of the total number of squares to the number of shaded squares.

Excess information quantifies how much more information a mechanism generates than the sum of its submechanisms – how synergistic the internal dependencies are.

Given subsystem with mechanism $\mathfrak{m}$, partition $\mathcal{P} = \{M^1 \dots M^m\}$ of the occasions in $src(\mathfrak{m})$, and output x_{out}, define excess information as follows. Let $\mathfrak{m}^j := \mathfrak{m} \cap (M^j \times X)$ be the restriction of $\mathfrak{m}$ to sources in M^j. *Excess information* over $\mathcal{P}$ is

$$\xi(\mathfrak{m}, \mathcal{P}, x_{out}) := ei(\mathfrak{m}, x_{out}) - \sum_j ei(\mathfrak{m}^j, x_{out}). \quad (6)$$

Excess information (sans partition) is computed over the information-theoretic weakest link $\mathcal{P}^{MIP}$

$$\xi(\mathfrak{m}, x_{out}) := \xi(\mathfrak{m}, \mathcal{P}^{MIP}, x_{out}). \quad (7)$$

Let $A_{M^j} := \prod_{l \in M^j} A_j$. The minimum information partition[1] $\mathcal{P}^{MIP}$ minimizes normalized excess information:

$$\mathcal{P}^{MIP} := \arg\min_{\mathcal{P}} \frac{\xi(\mathfrak{m}, \mathcal{P}, x_{out})}{\mathcal{N}_{\mathcal{P}}}, \text{ where}$$
$$\mathcal{N}_{\mathcal{P}} := (m-1) \cdot \min_j \{\log_2 |A_{M^j}|\}.$$

[1] We restrict to bipartitions to reduce the computational burden.

Excess information is negative if any decomposition of the system generates more information than the whole.

Fig. 3 shows how two cells taken together can generate the same, less, or more information than their sum taken individually depending on how their categorizations overlap. Note the figure decomposes the mechanism of the system over *targets* rather than sources and so does not depict excess information – which is more useful but harder to illustrate.

Effective information and excess information can be computed for any submechanism of any coarse-graining of any cellular automaton.

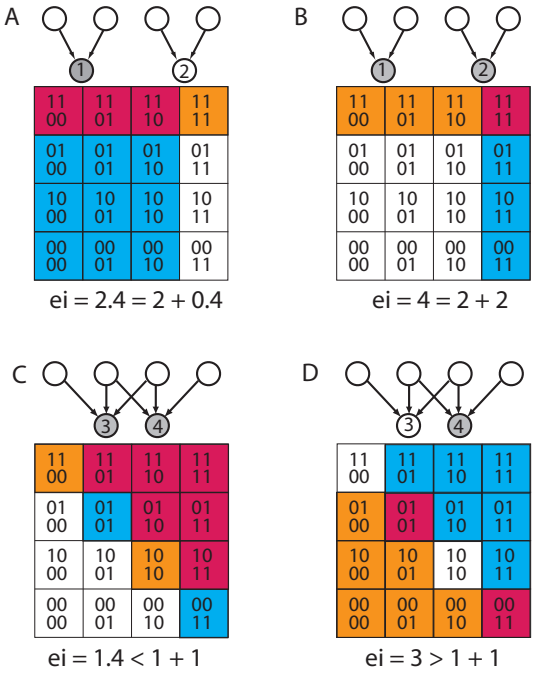

Figure 3: Independent, redundant and synergistic information. (AB): Independent. Orthogonal categorizations, orange+pink and blue+pink shadings respectively, by n_1 and n_2. (C): Partially redundant. Both cells fire; categorizations overlap (pink) more "than expected" and $ei(n_3n_4, 11) < ei(n_3, 1) + ei(n_4, 1)$. (D): Synergistic. Overlap is less "than expected"; $ei(n_3n_4, 01) > ei(n_3, 0) + ei(n_4, 1)$.

Application: Conway's Game of Life

The Game of Life has interesting dynamics at a range of spatiotemporal scales. At the atomic level, each coordinate (cell i at time t) is an occasion and information processing is extremely local. At coarser granularities, information can propagate through channels, so that units generate information at a distance. Gliders, for example, are distributed objects that can interact over large distances in space and time, Fig. 4A, and provide an important example of an emergent process (Dennett, 1991; Beer, 2004).

This section shows how effective and excess information quantifiably distinguish coarse-grainings expressing glider

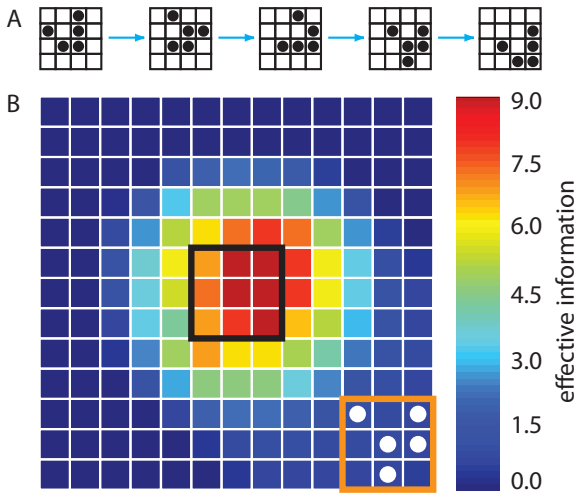

Figure 4: Detecting focal points. (A): A glider moves 1 diagonal square every 4 time steps. (B): Cells in the orange and black outlined 3×3 squares are units at $t = 0$ and $t = -20$ respectively, with x_{out} the glider shown. Cells at $t = -21$ are blank ground; other occasions are channel. Shifting the position of the black square produces a family of coarse-grainings. Effective information is shown as the black square's center varies over the grid.

dynamics well from those expressing it badly.

Effective information detects focal points. Fig. 4A shows a glider trajectory, which passes through 1 diagonal step over 4 tics. Fig. 4B investigates how glider trajectories are captured by coarse-grainings: if there is a glider in the 3×3 orange square at time 0, Fig. 4B, it must have passed through the black square at $t = -20$ to get there. Are coarse-grainings that respect glider trajectories quantifiably better than those that do not?

Fig. 4B fixes occasions in the black square at $t = -20$ and the orange square at $t = 0$ as units (18 total), the ground as blank grid at $t = -21$ and everything else as channel. Varying the spatial location of the black square over the grid, we obtain a *family* of coarse-grainings. Effective information for each graining in the family is shown in the figure. There is a clear focal point exactly where the black square intersects the spatiotemporal trajectory of the glider where ei is maximized (dark red). Effective information is zero for locations that are too far or too close at $t = -20$ to effect the output of the orange square at $t = 0$.

Effective information thus provides a tool analogous to a camera focus: grainings closer to the focal point express glider dynamics better.

Macroscopic texture varies with distance. The behavior of individual cells within a glider trajectory is far more complicated than the glider itself, which transitions through 4 phases as it traverses its diagonal trajectory, Fig. 4A. Does coarse-graining quantifiably simplify dynamics?

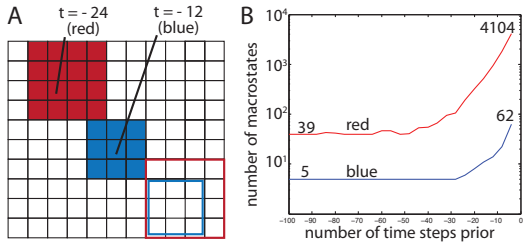

Figure 5: Macro-alphabets as a function of distance. (A): Consider two families of coarse-grainings with channel and ground as in Fig. 4. First, take the blue squares (filled and empty) as units at times $-4n$ and 0 where n is the diagonal distance between them. Second, repeat for the red squares. (B): Log-plot of the size of the filled squares' macro-alphabets as a function of $-4n$.

Fig. 5 constructs pairs of 3×3 units out of occasions at various distances from one another and computes their macro-alphabets. A 3×3 unit has a micro-alphabet of $2^9 = 512$ outputs. The macro-alphabet is found by grouping micro-outputs together into equivalences classes if their effect is the same after propagating through the channel. We find that the size of the macro-alphabet decreases exponentially as the distance between units increases, stabilizing at 5 macro-outputs: the 4 glider phases in Fig. 4A and a large equivalence class of outputs that do not propagate to the target unit and are equivalent to a blank patch of grid. A similar phenomenon occurs for pairs of 4×4 units, also Fig. 5.

Continuing the camera analogy: at close range the texture of units is visible. As the distance increases, the channel absorbs more of the detail. The computational texture of the system is simpler at coarser-grains yielding a more symbolic description where glider dynamics are described via 4 basic phases produced by a single macroscopic unit rather than 2^9 outputs produced by 9 microscopic occasions.

Excess information detects spatial organization. So far we have only considered grainings of the Game of Life that respect its spatial organization – in effect, taking the spatial structure for granted. *A priori*, there is nothing stopping us from grouping the 8 gray cells in Fig. 6A into a single unit that *does not* respect the spatial organization, since its constituents are separated in space. Are coarse-grainings that respect the grid-structure quantifiably better than others?

Fig. 6A shows a coarse-graining that does *not* respect the grid. It constructs two units, one from *both* gray squares at $t = 1$ and the other from *both* red squares at $t = 0$. Intuitively, the coarse-graining is unsatisfactory since it builds units whose constituent occasions have nothing to do with each other over the time-scale in question. Quantitatively, excess information over the obvious partition $\mathcal{P}$ of the system into two parts is 0 bits. It is easy to show $\xi \leq 0$ for any disjoint units. By comparison, the coarse-grainings in panels CD, which respect the grid structure, both generate positive excess information.

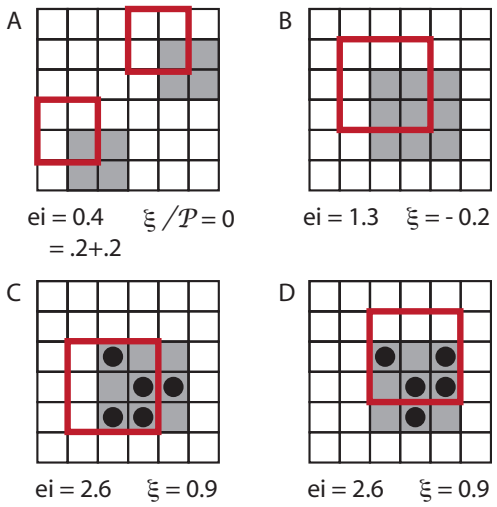

Figure 6: Detecting spatial organization. Units are the cells in the red (thick-edged) and gray (filled) squares at $t = 0$ and $t = 1$ respectively; other occasions are extrinsic noise. (A): $\xi = 0$. The coarse-graining groups non-interacting occasions into units. (B): $\xi < 0$. A blank grid is highly redundant. (CD): $\xi > 0$. Gliders perform interesting information-processing.

Thus we find that not only does our information-theoretic camera have an automatic focus, it also detects when processes hang together to form a single coherent scene.

Excess information detects gliders. Blank stretches of grid, Fig. 6B, are boring. There is nothing going on. Are interesting patches of grid quantifiably distinguishable from boring patches?

Excess information distinguishes blank grids from gliders: ξ on the blank grid is negative, Fig. 6B , since the information generated by the cells is redundant analogous to Fig. 3C. By contrast, ξ for a glider is positive, Fig. 6CD, since its cells perform synergistic categorizations, similarly to Fig. 3D. Glider trajectories are also captured by excess information: varying the location of the red units (at $t = 0$) around the gray units we find that ξ is maximized in the positions shown, Fig. 6CD, thus expressing the rightwards and downwards motions of the respective gliders.

Returning to the camera analogy, blank patches of grid fade into (back)ground or are (transparent) channel, whereas gliders are highlighted front and center as units.

Application: Hopfield networks

Hopfield networks embed energy landscapes into their connectivity. For any initial condition they tend to one of few attractors – troughs in the landscape (Hopfield, 1982; Amit, 1989). Although cells in Hopfield networks are quite different from neurons, there is evidence suggesting neuronal populations transition between coherent distributed states similar to attractors (Abeles et al., 1995; Jones et al., 2007).

	output		INT: $B \to B$		EXT: $A \to B$	
t	A	B	ei	$\max \xi$	ei	$\max \xi$
0	00000000	01010101				
1	10100011	01010101	**2.42**	**0.10**	0.31	0.04
2	10101010	00010101	1.85	0.08	**2.44**	**0.16**
3	10101010	00101011	1.96	0.12	**6.89**	**0.27**
4	10101010	00101010	**1.85**	0.08	1.60	**0.10**
5	10101010	10101010	**2.42**	**0.10**	0.90	0.06
6	10101010	10101010	**2.42**	**0.10**	0.31	0.04

Table 1: Analysis of unidirectionally coupled Hopfield networks $A \to B$ each containing 8 cells. The networks and coupling embed attractors $\{00001111, 00110011, 01010101\}$ and their mirrors. Temperature is $T = 0.25$. A sample run is analyzed using two coarse-grainings: INT captures B's effect on itself and EXT captures A's effect on B; see text.

Attractors are population level phenomena. They arise because of interactions between groups of cells – no single cell is responsible for their existence – suggesting that coarse-graining may reveal interesting features of attractor dynamics.

Effective information detects causal interactions. Table 1 analyzes a sample run of unidirectionally coupled Hopfield networks $A \to B$. Network A is initialized at an unstable point in the energy landscape and B in an attractor. A settles into a different attractor from B and then shoves B into the new attractor over a few time steps. Intuitively, A only exerts a strong force on B once it has settled in an attractor and before B transitions to the same attractor. Is the force A exerts on B quantitatively detectable?

Table 1 shows the effects of A and B respectively on B by computing ei for two coarse-grainings constructed for each transition $t \to t+1$. Coarse-graining INT sets cells in B at t and $t+1$ as units and A as extrinsic noise. EXT sets cells in A at t and B at $t+1$ as units and fixes B at time t as ground.

INT generates higher ei for all transitions except $1 \to 2 \to 3$, precisely when A shoves B. Effective information is high when an output is sensitive to changes in an input so it is unsurprising that B is more sensitive to changes in A exactly when A forces B out from one attractor into another. Analyzing other sample runs (not shown) confirms that ei reliably detects when A shoves B out of an attractor.

Macroscopic mechanisms depend on the ground. Fixing the ground incorporates population-level biases into a coarse-grained cellular automaton's information-processing.

The ground in coarse-graining EXT (i.e. the output of B at $t-1$) biases the mechanisms of the units in B at time t. When the ground is an attractor, it introduces tremendous inertia into the coarse-grained dynamics since B is heavily biased towards outputting the attractor again. Few inputs from A can overcome this inertia, so if B is pushed out of an attractor it generates high ei about A. Conversely, when B stays in an attractor, e.g. transition $5 \to 6$, it follows its internal bias and so generates low ei about A.

Excess information detects attractor redundancy. Following our analysis of gliders, we investigate how attractors are captured by excess information. It turns out that ξ is negative in all cases: the functional dependencies within Hopfield networks are redundant. An attractor is analogous to a blank Game of Life grid where little is going on. Thus, although attractors are population-level phenomena, we exclude them as emergent processes.

Excess information expresses attractor transitions. We therefore refine our analysis and compute the subset of units at time t that maximize ξ; maximum values are shown in Table 1. We find that the system decomposes into pairs of occasions with low ξ, except when B is shoved, in which case larger structures of 5 occasions emerge. This fits prior analysis showing transitions between attractors yield more integrated dynamics (Balduzzi and Tononi, 2008) and suggestions that cortical dynamics is metastable, characterized by antagonism between local attractors (Friston, 1997).

Our analysis suggests that *transitions* between attractors are the most interesting emergent behaviors in coupled Hopfield networks. How this generalizes to more sophisticated models remains to be seen.

Emergence

The examples show we can quantify how well a graining expresses a cellular automaton's dynamics. Effective information detects glider trajectories and also captures when one Hopfield network shoves another. However, ei does not detect whether a unit is integrated. For this we need excess information, which compares the information generated by a mechanism to that generated by its submechanisms. Forming units out of disjoint collections of occasions yields $\xi = 0$. Moreover, boring units (such as blank patches of grid or dead-end fixed point attractors) have negative ξ. Thus, ξ is a promising candidate for quantifying emergent processes.

This section formalizes the intuition that a system is emergent if its dynamics are better expressed at coarser spatiotemporal granularities. The idea is simple. Emergent units should generate more excess information, and have more excess information generated about them, than their sub-units. Moreover emergent units should generate more excess information than *neighboring* units, recall Fig. 4.

Stating the definition precisely requires some notation. Let $\mathfrak{src}_{v_l} = \{v_l\} \cup \{v_k | k \to l\}$ and similarly for $\mathfrak{trg}_{v_l}$. Let $\mathcal{J}$ be a subgraining of $\mathcal{K}$, denoted $\mathcal{J} \prec \mathcal{K}$, if for every $\mathbf{U}_j \in \mathcal{J}$ there is a unit $\mathbf{U}_k \in \mathcal{K}$ such that $\mathbf{U}_j \subsetneq \mathbf{U}_k$. We compare mechanism $\mathfrak{m} \subset \mathcal{K}$ with its subgrains via

$$\xi_{\mathcal{K}/\mathcal{J}}(\mathfrak{m}, x_{out}) := ei_{\tilde{\mathcal{K}}}(\mathfrak{m}, x_{out}) - \sum_{v_j \in \mathcal{J}} ei_{\tilde{\mathcal{J}}}(\mathfrak{m}^j, x_{out}),$$

where $\mathfrak{m}^j = \mathfrak{m} \cap \mathfrak{src}_{v_j}$ and $ei_{\tilde{\mathcal{K}}}$ signifies effective information is computed over $\mathcal{K}$ using micro-alphabets.

Definition (emergence)**.** *Fix cellular automaton X with output x_{out}. Coarse-graining*[2] *$\mathcal{K}$ is* emergent *if it satisfies conditions* E1 *and* E2*.*

E1. Each unit $\mathbf{U}_l \in \mathcal{K}$ generates excess information about its sources and has excess information generated about it by its targets, relative to subgrains $\mathcal{J} \prec \mathcal{K}$:

$$0 < \xi_{\mathcal{J}/\mathcal{K}}\big(\mathfrak{src}_{\mathbf{U}_l}, x_{out}\big) \text{ and } 0 < \xi_{\mathcal{J}/\mathcal{K}}\big(\mathfrak{trg}_{\mathbf{U}_l}, x_{out}\big). \tag{8}$$

E2. There is an *emergent* subgrain $\mathcal{J} \prec \mathcal{K}$ such that (i) every unit of $\mathcal{K}$ contains a unit of $\mathcal{J}$ and (ii) neighbors $\mathcal{K}'$ (defined below) of $\mathcal{K}$ with respect to $\mathcal{J}$ satisfy

$$\xi_{\mathcal{J}/\mathcal{K}'}\big(\mathfrak{src}_{\mathbf{U}'}, x_{out}\big) \leq \xi_{\mathcal{J}/\mathcal{K}}\big(\mathfrak{src}_{\mathbf{U}}, x_{out}\big) \tag{9}$$

for all $\mathbf{U} \in \mathcal{K}$, and similarly for $\mathfrak{trg}$'s.

If $\mathcal{K}$ has no emergent subgrains then E2 is vacuous.

Grain $\mathcal{K}'$ is a *neighbor* of $\mathcal{K}$ with respect to $\mathcal{J} \prec \mathcal{K}$ if for every $\mathbf{U} \in \mathcal{K}$ there is a unique $\mathbf{U}' \in \mathcal{K}'$ satisfying

N1. there is a unit $T \in \mathcal{J}$ such that $T \subset \mathbf{U}, \mathbf{U}'$, $\mathfrak{src}_T \subset \mathfrak{src}_{\mathbf{U}}, \mathfrak{src}_{\mathbf{U}'}$ and similarly for $\mathfrak{trg}$; and

N2. the alphabet of $\mathbf{U}'$ is no larger than $\mathbf{U}$: $\big|\prod_{k \in \mathbf{U}'} A_k\big| \leq \big|\prod_{l \in \mathbf{U}} A_l\big|$, and similarly for the combined alphabets of their sources and targets respectively.

The graining $\mathcal{E}_X$ that best expresses X outputting x_{out} is found by maximizing normalized excess information:

$$\mathcal{E}_X(x_{out}) := \arg \max_{\{\mathcal{K} \,|\, \text{emergent}\}} \frac{\xi(\mathcal{K}, x_{out})}{\mathcal{N}^{\mathcal{K}}_{\mathcal{P}^{MIP}}}. \tag{10}$$

Here, $\mathcal{N}^{\mathcal{K}}_{\mathcal{P}^{MIP}}$ is the normalizing constant found when computing the minimum information partition for $\mathcal{K}$.

Some implications. We apply the definition to the Game of Life to gain insight into its mechanics.

Condition E1 requires that interactions between units and their sources (and targets) are synergistic, Fig. 6CD. Units that decompose into independent pieces, Fig. 6A, or perform highly redundant operations, Fig. 6B, are therefore not emergent.

Condition E2 compares units to their neighbors. Rather than build the automaton's spatial organization directly into the definition, neighbors of $\mathcal{K}$ are defined as coarse-grainings whose units overlap with $\mathcal{K}$ and whose alphabets are no bigger. Coarse-grainings with higher ξ than their neighbors are closer to focal points, recall Fig. 4 and Fig. 6CD, where ξ was maximized for units respecting glider trajectories. An analysis of glider boundaries similar in spirit to this paper is (Beer, 2004).

[2] Ground output $s^{\mathbf{G}}$ is x_{out} restricted to ground occasions.

Finally, Eq. (10) picks out the most expressive coarse-graining. The normalization plays two roles. First, it biases the optimization towards grainings whose MIPs contain few, symmetric parts following (Balduzzi and Tononi, 2008). Second, it biases the optimization towards systems with simpler macro-alphabets. Recall, Fig. 5, that coarse-graining produces more symbolic interactions by decreasing the size of alphabets. Simplifying alphabets typically reduces effective and excess information since there are less bits to go around. The normalization term rewards simpler levels of description, so long as they use the bits in play more synergistically.

Discussion

In this paper we introduced a flexible, scalable coarse-graining method that applies to any cellular automaton. Our notion of automaton applies to a broad range of systems. The constraints are that they (i) decompose into discrete components with (ii) finite alphabets where (iii) time passes in discrete tics. We then described how to quantify the information generated when a system produces an output (at any scale) both as a whole and relative to its subsystems. An important feature of our approach is that the output x_{out} of a graining is incorporated into the ground and also directly influences ei and ξ through computation of the actual repertoires. Coarse-graining and emergence therefore capture some of the *suppleness* of biological processes (Bedau, 1997): they are context-dependent and require many *ceteris paribus* clauses (i.e. background) to describe.

Investigating examples taken from Conway's Game of Life and coupled Hopfield networks, we accumulated a small but significant body of evidence confirming the principle that *expressive coarse-grainings generate more information relative to sub-grainings*. Finally, we provisionally defined emergent processes. The definition is provisional since it derives from analyzing a small fraction of the possible coarse-grainings of only two kinds of cellular automata.

Hopfield networks and the Game of Life are simple models capturing some important aspects of biological systems. Ultimately, we would like to analyze emergent phenomena in more realistic models, in particular of the brain. Conscious percepts take 100-200ms to arise and brain activity is (presumably) better expressed as comparatively leisurely interactions between neurons or neuronal assemblies rather than much faster interactions between atoms or molecules (Tononi, 2004). To apply the techniques developed here to more realistic models we must confront a computational hurdle: the number of coarse-grainings that can be imposed on large cellular automata is vast. Nevertheless, the approach developed here may still be of use. First, manipulating macro-alphabets provides a method for performing approximate computations on large-scale systems. Second, for more fine-grained analysis, initial estimates about which coarse-grainings best express a system's dynamics can be fine-tuned by comparing them with neighbors.

Acknowledgements. The author thanks Dominik Janzing for many useful comments on an earlier draft, Giulio Tononi for stimulating conversations and Virgil Griffiths for emphasizing the importance of excess information.

References

Abeles, M., Bergman, H., Gat, I., Meilijson, I., Seidemann, E., Tishby, N., and Vaadia, E. (1995). Cortical activity flips among quasi-stationary states. *Proc. Nat. Acad. Sci.*, 92:8616–8620.

Amit, D. (1989). *Modelling brain function: the world of attractor neural networks*. Cambridge University Press.

Anderson, P. W. (1972). More is different. *Science*, 177(4047):393–6.

Balduzzi, D. and Tononi, G. (2008). Integrated Information in Discrete Dynamical Systems: Motivation and Theoretical Framework. *PLoS Comput Biol*, 4(6):e1000091.

Balduzzi, D. and Tononi, G. (2009). Qualia: the geometry of integrated information. *PLoS Comput Biol*, 5(8):e1000462.

Bedau, M. A. (1997). Emergent models of supple dynamics in life and mind. *Brain Cogn*, 34(1):5–27.

Beer, R. D. (2004). Autopoiesis and cognition in the game of life. *Artif Life*, 10(3):309–26.

Berlekamp, E., Conway, J., and Guy, R. (1982). *Winning Ways for your Mathematical Plays*, volume 2. Academic Press.

Crutchfield, J. (1994). The calculi of emergence: Computation, dynamics, and induction. *Physica D*, 75:11–54.

Dennett, D. C. (1991). Real Patterns. *J. Philosophy*, 88(1):27–51.

Friston, K. (1997). Transients, metastability and neuronal dynamics. *Neuroimage*, 5:164–171.

Hopfield, J. (1982). Neural networks and physical systems with emergent computational properties. *Proc. Nat. Acad. Sci.*, 79:2554–2558.

Jones, L. M., Fontanini, A., Sadacca, B. F., Miller, P., and Katz, D. B. (2007). Natural stimuli evoke dynamic sequences of states in sensory cortical ensembles. *Proc Natl Acad Sci U S A*, 104(47):18772–18777.

Pearl, J. (2000). *Causality: models, reasoning and inference*. Cambridge University Press.

Polani, D. (2006). Emergence, intrinsic structure of information, and agenthood. *Int J Complex Systems*, 1937.

Seth, A. K. (2010). Measuring autonomy and emergence via Granger causality. *Artif Life*, 16(2):179–96.

Shalizi, C. and Moore, C. (2006). What is a macrostate: Subjective observations and objective dynamics. *http://arxiv.org/abs/condmat/0303625*.

Tononi, G. (2004). An information integration theory of consciousness. *BMC Neurosci*, 5:42.