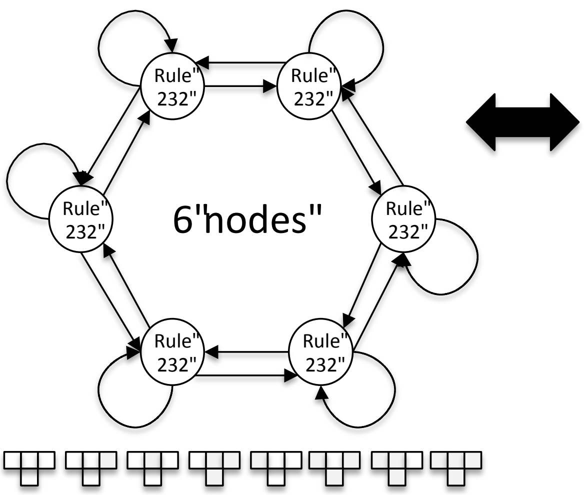
Supplementary Methods

In the following, we explain how cause and effect repertoires are calculated from the transition probability matrix (TPM). For further information, see Text S2 in [1] and the glossary in [2].

1. Transition Probability Matrix (TPM)

The transition probability matrix of a system of N elements is obtained by perturbing the system S into all 2^N possible states s_t and observing the probability distribution of resulting system states s_{t+1}. It thus describes all possible state transitions of the system and their probabilities assuming each state s_t with equal probability. Since the ECA considered in this article are deterministic, there is only one possible future state s_{t+1} for each state s_t. ECA cells are binary elements and conditionally independent. For these reasons, we can write the TPM in the state x element format (Figure 1), a compressed version of the typical state x state format. In the state x element TPM, each entry specifies the probability of the respective element (A-F, columns) to be in state “1” at time t + 1 following a perturbation into the indicated state (rows) at time t. While for deterministic ECA all entries in the TPM are either 0 or 1, in binary probabilistic systems any number from 0 to 1 is possible.

Rule 232 = “Majority”



State transition probability matrix (TPM)

t \ t+1 (ABCDEF)	A	B	C	D	E	F
000000	0	0	0	0	0	0
100000	0	0	0	0	0	0
010000	0	0	0	0	0	0
110000	1	1	0	0	0	0
001000	0	0	0	0	0	0
101000	0	1	0	0	0	0
011000	0	1	1	0	0	0
111000	1	1	1	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Figure 1. An N = 6 ECA, implementing rule 232, the majority rule, and its TPM in state x element format. TPM entries indicate p(element) = 1, the probability to be in state “1”.

Conditional independence is a requirement in the IIT formalism, excluding instantaneous causation. Considering the two elements A and B in the system A-F and their respective inputs (Figure 1 and Figure 9 in the main text), for example, conditional independence can be expressed mathematically as:

$$p(AB_{t+1}|ABCF_t) = p(A_{t+1}|ABF_t) \times p(B_{t+1}|ABC_t). \quad \text{(S1)}$$

In words, given their inputs at t, the probability distributions of the states of A and B at time t + 1 can be determined independently.

2. Cause Repertoire

The cause repertoire of a mechanism M_t in its current state m_t can be obtained from the TPM using Bayes' rule. This is demonstrated below, for the example mechanism $A_t = 1$, in the ECA system A-F_t = 111000 implementing rule 232 (Figure 9 in the main text). A_t can only possibly constrain its inputs ABF_{t-1}, but not the other elements in the system. Thus, we consider the cause repertoire $p_{cause}(ABF_{t-1}|A_t = 1)$, the distribution of possible past states of ABF_{t-1} conditioned on $A_t = 1$. Using Bayes' rule:

$$p_{cause}(ABF_{t-1}|A_t = 1) = \frac{p(A_t = 1|ABF_{t-1}) \cdot p_{uc}(ABF_{t-1})}{p(A_t = 1)} \quad \text{(S2)}$$

Here, $p(A_t = 1|ABF_{t-1})$ corresponds to A's column in the TPM (Figure 1), averaged across all elements not in the set of elements ABF, with $p(A_t = 1|ABF_{t-1}) = 0$ for ABF_{t-1} = {000, 100, 010, 001} and $p(A_t = 1|ABF_{t-1}) = 1$ for ABF_{t-1} = {110, 101, 011, 111}. This distribution is scaled by $p_{uc}(ABF_{t-1})$, which here is 1/8 for each past state, the uniform distribution over ABF, since the elements are perturbed into all states with equal probability; and $p(A_t = 1)$ = 0.5, since A implements the Majority function which results in state "1" for half of its 8 possible input states, again applied with equal probability. Altogether, this results in the cause repertoire $p_{cause}(ABF_{t-1}|A_t = 1)$ shown in Figure 9B of main text. Note that Equation (S2) is only defined for "possible" states of sets of elements with $p(M_t = m_t) > 0$, meaning states that can be reached at t + 1 following a system perturbation at t (excluding so-called "Garden of Eden" states of the set of elements). By definition, sets of elements in states with $p(M_t = m_t) = 0$ that cannot be reached from within the system itself also cannot specify cause repertoires. Consequently, cause-effect structures here are only calculated for possible system states.

The cause-repertoire of higher order mechanisms constituted of multiple system elements, such as mechanism AB_t = 11, is obtained from the product distribution of the individual mechanism elements, for example:

$$p_{cause}(ABCF_{t-1}|AB_t = 1) = \frac{1}{K} \cdot p_{cause}(ABF_{t-1}|A_t = 1) \times p_{cause}(ABC_{t-1}|B_t = 1), \quad \text{(S3)}$$

where K is a normalization factor so that the product probability distribution sums to 1.

The unconstrained cause repertoire is simply the uniform distribution of all possible states of a set of elements.

3. Effect Repertoire

The effect repertoire of a mechanism M_t in its current state m_t can be obtained from the TPM as demonstrated below for the example mechanism $A_t = 1$, in the ECA system A-F_t = 111000 implementing rule 232 (Figure 9 in the main text). A_t can only possibly constrain its outputs ABF_{t+1}, but not the other elements in the system. Thus, we consider the effect repertoire $p_{effect}(ABF_{t+1}|A_t = 1)$, which is obtained by fixing the state of element A_t to "1", while the remaining inputs to A_{t+1}, B_{t+1}, and F_{t+1} are perturbed independently into all possible states with equal probability. To eliminate effects of correlations through common inputs, the effect repertoire is defined as:

$$p_{effect}(ABF_{t+1}|A_t = 1) = p_{effect}(A_{t+1}|A_t = 1) \times p_{effect}(B_{t+1}|A_t = 1) \times p_{effect}(F_{t+1}|A_t = 1). \quad \text{(S4)}$$

The individual effect repertoires can be read out from the TPM columns of the respective elements, averaging over all rows in which $A_t = 1$ (*i.e.*, conditioning on $A_t = 1$). Since A, B, and F are Majority functions, the fact that one of their inputs (A_t) is in state “1” results in p(“1” at t+1) = 0.75 and p(“0” at t + 1) = 0.25 for each of them, assuming maximum entropy for the other two inputs. The resulting product distribution $p_{effect}(ABF_{t+1}|A_t = 1)$ is shown in Figure 9B in the main text.

For higher order mechanisms such as $AB_t = 11$, the effect repertoire is calculated in the same way, only now averaging over all rows in the TPM where $AB_t = 11$.

The unconstrained effect repertoire is the product distribution of all elements under no constraints, *i.e.*, averaging all rows of the TPM for the individual elements without conditioning and then taking the product distribution. Since all elements are Majority functions, their individual unconstrained probability distribution at t + 1 is uniform p(“1”) = p(“0”) = 0.5.

References

1. Oizumi, M.; Albantakis, L.; Tononi, G. From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0. *PLoS Comput. Biol.* **2014**, *10*, e1003588.
2. Tononi, G. Integrated information theory. *Scholarpedia* **2015**, *10*, 4164.