Engineering Emergence

Abel Jansma[*1,2,3] and Erik Hoel[†4]

[1]Dutch Institute for Emergent Phenomena, the Netherlands
[2]Institute for Logic, Language and Computation, and
[3]Institute of Physics, University of Amsterdam, the Netherlands
[4]Allen Discovery Center, Tufts University, Medford, MA, USA

October 15, 2025

Abstract

A defining property of complex systems is that they have multiscale structure. How does this multiscale structure come about? We argue that within systems there emerges a hierarchy of scales that contribute to a system's causal workings. An intuitive example is how a computer can be described at the level of its hardware circuitry (its microscale) but also its machine code (a mesoscale) and all the way up at its operating system (its macroscale). Here we show that even simple systems possess this kind of emergent hierarchy, which usually forms over only a small subset of the super-exponentially many possible scales of description. To capture this formally, we extend the theory of causal emergence (version 2.0) so as to analyze how causal contributions span the full multiscale structure of a system. Our analysis reveals that systems can be classified along a taxonomy of emergence, such as being either "top-heavy" or "bottom-heavy" in their causal workings. From this new taxonomy of emergence, we derive a measure of complexity based on a literal notion of "scale-freeness" (here, when causation is spread equally across the scales of a system) and compare this to the standard network science definition of "scale-freeness" based on degree distribution, showing the two are closely related. Finally, we demonstrate the ability to engineer not just the degree of emergence in a system, but to control it with pinpoint precision.

[*]a.a.a.jansma@uva.nl
[†]erik.hoel@tufts.edu

1 The paradox of scale

A central reason why complex systems are considered complex is because they have multiscale structure [1, 2]. However, this multiscale structure is regularly hidden, simply because when scientists attempt to understand how a system works, they will often write down the equivalent of a single diagram—such as a causal model (e.g., a structural equation model or directed acyclic graph [3]), or a discrete state machine (e.g., a Markov chain). Writing down only a single set of variables, and charting their influence on each other, is certainly natural [4]. Yet this necessarily limits the analysis to a single scale of description (consisting of just those variables, or elements, or mechanisms, or states—depending on the model), which in turn hides a great deal of structure: All the other scales that are being ignored in favor of the one. Descriptions of systems at a single scale are like taking a 2D slice of a higher dimensional object and presenting that slice as if it were the object itself.

One reason this is common is that, when confronted with the full set of possible scales of description, it is a set so astronomically large it at first seems of little use, if not outright confusing. Even for a system with just 8 discrete states, the set of all possible partitions of those states is 4,140 (see Figure 1). This set forms a partially ordered lattice, or *mereology* [5], wherein some partitions are finer, or coarser, than other partitions. The ordered lattice can also be thought of as many paths that increase coarseness, spanning from the microscale of the system (its most fine-grained possible description) to the ultimate macroscale (its most coarse-grained possible description). Each horizontal slice of the lattice corresponds to some particular dimensionality (or level) of the system, usually with many possible partitions at each level.

Notably, the spatiotemporal hierarchy of the sciences (from microphysics to macroeconomics) suggests that among the astronomical number of ways to partition a system, there is a much smaller subset of actually causally-relevant scales that best (or better) explain a system's workings [6], at least when compared to random scales. This is true for individual systems too: e.g., the causally-relevant scales of a computer can also be described sensibly as a hierarchy, with its physical logic gates as a microscale, its machine code as a mesoscale, and its operating system as a macroscale.

A system can differ in important ways depending on what scale it is being described at. E.g., the mechanisms of a system, such as the kind of logic gates the system contains, can change at different scales of descriptions [7], and consequentially so can the types of computations [8, 9]. Broadly, how

a system functions can be described at different scales; e.g., brain function may be modeled as a large set of interacting individual neurons, but also as a (smaller) set of interacting cortical minicolumns or brain regions [10].
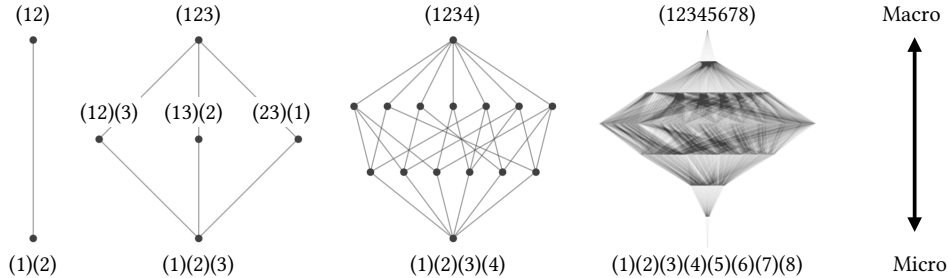

Figure 1: The multiscale structure of a system is based on a partitioning of its states. Partitions can be partially ordered by *refinement* to construct partitions lattices. The associated Hasse diagrams are shown here for 2, 3, 4, and 8-state systems. For any two partitions π_a, π_b we set $\pi_a \leq \pi_b \iff \forall b \in \pi_b : \exists a \in \pi_a$ s.t. $a \subseteq b$. This puts coarse partitions with few blocks (macroscales) at the top, and fine partitions with many blocks (microscales) at the bottom. The number of possible partitions of n states (the number of elements in the unpartitioned microscale at the bottom) grows superexponentially with n, namely, as the nth Bell number (the lattices above contain respectively 2, 5, 15, 4140 partitions).

We are certainly not the first to notice that complex systems have multiple descriptions at different scales: the observation has a long tradition in disparate fields, particularly in cognitive sciences (like Marr's three levels of analysis [11]); it also extends to questions of how software arises from hardware [12], or how macroscales are dynamically related to their microscales [13]. Additionally, there already exist various information-theoretic proposals for methods to identify particular macroscales that may operate as especially informative compressions of microscale dynamics [14, 15]. However, if macroscales add nothing beyond successful compression, this suggests they are not truly causally-relevant at all, and useful only for an observer's convenience.

When analyzing the causal workings of a system directly, the astronomical size of the set of scales creates a seeming paradox when it comes to understanding "What does what?", i.e., how the system actually works. The existence of the full set of scales entails that everything that happens in the system is massively overdetermined [16, 17] (to put it in lay terms: for

3

any one event, there is an astronomically large set of possible explanations across the full set of possible scales).

Confronted with this embarrassment of multiplicity, one option is to assume relativism, i.e., that when it comes to complex systems, there is "no privileged level of causation" [18]. However, this cannot account for the fact that some scales seem to truly capture a system's workings better than others, much like how the many special sciences beyond physics seem to give unique insight into the phenomena they study [19]. An alternative but equally extreme option is to assume reductionism, i.e., that only the microscale truly matters for the system's causal workings [20]. However, this too leads to absurdities (no higher level truly *does* anything)—indeed, this would indicate that most elements of science, the majority of which are macroscales of some sort, cannot truly add anything to a system's workings, and so are epiphenomenal [21].

2 How systems get their scales

Herein, we demonstrate a framework for resolving this paradox that enables a non-reductive, but also non-mysterious, rich view of multiscale structure. Specifically, we propose a method for identifying the subset of scales (defined as coarse-grainings of a system based on partitions) that contribute—non-redundantly and uniquely—to the system's overall causal workings. We judge this subset of scales to be causally emergent [22], since the causation they contribute cannot be fully reduced to any underlying scale (i.e., any scale below them along any micro→macro path).

To establish this framework, first we note that we can formalize each individual partition (such as those in Figure 1) as representing an individual scale, which is expressible as a causal model that describes the system's workings. Here, we focus on systems defined by transition probability matrices (TPMs) that reflect what the system would do (how it would transition next) in any given state X_t, or under any given intervention possible in the statespace (such as $do(X_t = x)$).

Normally, identifying causal models involves inference under a myriad of difficulties and limitations. For simplicity, both conceptual and mathematical, we therefore confine our analysis to causal models that are already known in their entirety, and are represented as Markov chains (as described by a given TPM—see SI S1 for details). We also assume that a system's microscale is well-defined, in that it forms a specific TPM, and that all the true conditional probabilities (the state transitions) are a priori known.

For a system that meets those conditions, this still results in an enormous set of TPMs representing scales beyond the microscale, any one of which are on the way to other (more macroscale) TPMs (that is, they lie along some micro→macro path). This can be seen by coarse-graining the original microscale TPM of a system according any one of its many possible partitions, such that, e.g., for partition $(123)(456)$ the microstates $1, 2, 3$ and $4, 5, 6$ would be coarse-grained together, and the probabilities of their possible effects (next states) and their possible causes (previous states) would also be summarized together, resulting in a macroscale TPM that has only two macrostates (see [7, 23] for more details and methods on coarse-graining microscale TPMs into macroscale TPMs).

Some of these many TPMs, each corresponding to a scale of description, are informative about the causal workings of the system as it moves from t to $t + 1$. Most are not. This is because a randomly chosen scale, and the constructed TPM it specifies, is often junk: by which we mean that the conditional probabilities are garbled and noisy. For a randomly chosen scale, upon examining its TPM that defines its dynamics from t to $t + 1$, causes do not deterministically lead to effects (in the form of future states), and for any given effect, there are many possible common causes (previous states). One can think of this problem intuitively as: "What would happen if you randomly coarse-grained parts of a computer together?"—most such coarse-grainings will average over random parts from different transistors together, resulting in obscure variables and confusing relationships. Conversely, for a much smaller subset of scales, their TPMs are intuitively informative and interesting, and bring the higher-scale causal relationships of the system into focus.

Specifically, we can judge the causal relationships at a particular scale of the system (again, defined as a TPM) based on interventions of the type $do(X_t = x)$, performed at some time t, and their effects observed at $t + 1$. For causal relationships to be strong (a property which is sometimes called "powerful," or "informative," or other synonyms) they must be deterministic (individual interventions should have few possible effects), and non-degenerate (most effects should be unique).

Assuming the simple case of Markovian systems, fully characterized by a TPM that captures the conditional probability of the state of the system at $t + 1$ (which we refer to as an effect e), given the state of the system at time t (which we refer to as a cause c): $T_{ce} = p(e|c)$, then *determinism* is defined based on some individual cause c (here, a state in a TPM), and a possible set of future effects (here, over the set of states possible at the next timestep).

$$\text{determinism}(c) = 1 + \frac{1}{\log_2 n} \sum_{e \in E} p(e|c) \log_2 p(e|c) \tag{1}$$

$$= 1 - \frac{H(E|c)}{\log_2 n} \tag{2}$$

Being based on the entropy, H, it is an information-theoretic quality that captures the degree of certainty about effects, given a cause, and is therefore sensitive to the strength of the conditional probabilities of a system [7]. Herein, we will refer only to the system-wide determinism, which is calculated based on the average occurrence of any given transition from c to e (a system transition from t to t_{+1}), given some prior distribution over causes, or interventions, $p(C)$, for which we use a uniform distribution at each scale [24]. For a system with transition probability matrix T, the cause-specific and system-wide determinism are given by

$$\text{determinism}_T(c) = 1 + \frac{1}{\log_2 n} \sum_{e \in E} T_{ec} \log_2 T_{ec} \tag{3}$$

$$\text{determinism}_T = \mathbb{E}_{p(C)} \left[\text{determinism}_T(c) \right] \tag{4}$$

$$= 1 + \frac{1}{\log_2 n} \sum_{c \in C} \sum_{e \in E} p(c) T_{ec} \log_2 T_{ec} \tag{5}$$

$$= 1 - \frac{H(E|C)}{\log_2 n} \tag{6}$$

Conversely, *degeneracy* captures the degree of overlap in the effects of causes, given a prior $p(C)$, calculated as:

$$\text{degeneracy}_T = 1 - \frac{H(E)}{\log_2(n)} \tag{7}$$

$$= 1 + \frac{1}{\log_2 n} \sum_{e \in E} T_e \log_2 T_e \tag{8}$$

where

$$T_e = \sum_{c \in C} p(c) T_{ec} \tag{9}$$

Determinism is 1 and degeneracy is 0 only in cases of a TPM that is a permutation matrix; determinism is 1 and degeneracy is 1 when all states of a TPM

transition to the same single state (all causes have the same common effect); and both are 0 when the transitions of a TPM are completely random. For this reason, determinism and degeneracy are closely related to the dynamical reversibility [25]. Additionally, the two terms are also closely related to the axiomatic definition of causation [22], being an information-theoretic generalization of the concepts of the sufficiency and necessity, respectively; when causes are always completely sufficient for their effects, determinism is high, and when causes are always necessary for their effects, degeneracy is low.

Measures of probabilistic causation, which capture to what degree something is a cause, [26], are generally reducible to the sufficiency and necessity, and therefore are closely related to the determinism and degeneracy as well [24]. Therefore, we refer to the determinism and the degeneracy together as the *causal primitives* (CP) of a scale, as in CE 2.0 [22], to reflect that a well-tuned chosen measure of causation (like a measure of causal power, or strength, or degree, or informativeness, or constraint, or some other interpretation) will set some analog of these two terms in relationship, such that causes are more powerful (or stronger, or more informative, etc.) when determinism is closer to 1 and degeneracy is closer to 0.

For simplicity's sake, we define the inverse of the degeneracy $(1-degeneracy)$ as the *specificity*. Then, we capture the two terms' relationship by simply adding the two terms (determinism and the inverse of degeneracy) together, referring to this summation as CP, and this is calculated at each scale of the system (i.e., for each TPM), in a way that normalizes to a $[0, 1]$ range:

$$CP = \text{determinism}_T + \text{specificity}_T - 1. \qquad (10)$$

CP represents a particular scale's contribution to the system's causal workings, and is equivalent to the *effectiveness* defined in [7].

An example of calculating the combined causal primitives for different scales of a system is shown in Figure 2. Two scales out of the set of all possible scales of a system were chosen at the same level (the same degree of dimension reduction), and each were used to construct two different TPMs. Notably, the two respectively differ significantly in their determinism and degeneracy.
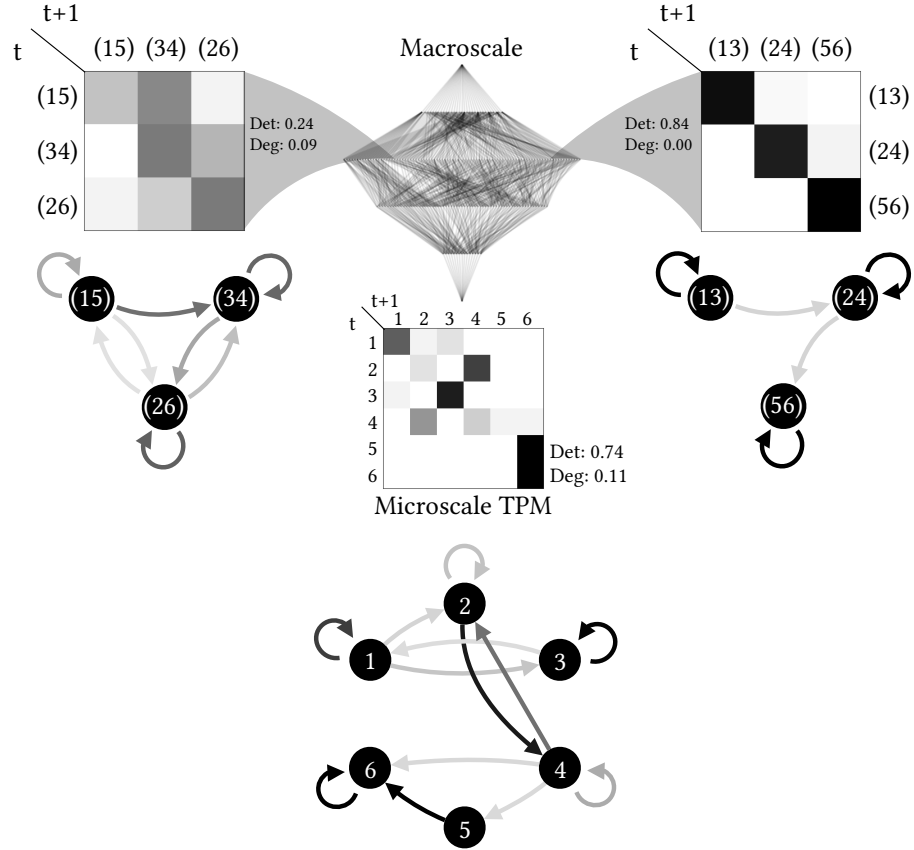
7

Figure 2: A demonstration of calculating CP at specific scales out of the full set of scales of a 6-state system (here a Markov chain described by its microscale TPM, the probabilities of which are shown in grayscale in the middle). Above the microscale TPM we show the associated Hasse diagram. Moving up the lattice traverses all coarse-grainings, proceeding from smaller groups to larger one. *Left:* a generic coarse-graining that reduces CP (by lowering determinism) indicating that scale fails to add anything to the system's causal workings. *Right:* an especially informative coarse-graining that groups the six states into appropriate pairs, increasing determinism and decreasing degeneracy. This is reflected in the associated graph by the dominance of self-loops.

The microscale TPM has a CP value of 0.63. Meanwhile, the macroscale TPM highlighted on the left in Figure 2 has a CP value of only 0.13, reflecting that the causal structure is represented in a noisy manner, with

transitions that are highly uncertain. Comparatively, the macroscale TPM on the right has a CP value of 0.84, reflecting a positive gain in CP, because the causal relationships are more deterministic and less degenerate at that scale, compared to the underlying microscale. That is, while the scales themselves are reducible to their underlying scales (being a coarse-graining of the microscale), they can still add to the system's causal workings (here reflected by the joint CP value) by increasing the determinism and decreasing the degeneracy. Mathematically, this is like adding error correction to causal relationships due to the multiple-realizability of macroscales [21].

Note that measuring these quantities (the system-wide determinism and degeneracy) requires a prior distribution over causes $p(C)$ to reconstruct the joint distribution from T. This corresponds to an *intervention distribution* that specifies the probabilities with which each of the interventions are performed at this scale (it can also be thought of as the set of possible counterfactuals in the causal analysis). It is important that $p(C)$ is not solely based on the observed distribution, since in most cases interventions/counterfactuals are required for capturing causation [3] (here we use a uniform distribution, to reflect treating counterfactuals/interventions equally—see the Discussion for more details).

3   How hierarchies emerge across scales

Previous research introducing CE 2.0 examined how CP could increase across a single micro→macro path [22]. Here, we extend CE 2.0 to be calculated over all possible micro→macro paths that traverse a system: that is, we demonstrate how to distribute the determinism and degeneracy of a system out across the full set of scales. This carves away all but the subset of scales that non-redundantly contribute to the system's causal workings: this remainder we call the *emergent causal hierarchy* (or *emergent hierarchy* for short), defined as the subset of scales that contribute non-zero and non-redundant gains in CP. We introduce a taxonomy for these emergent hierarchies: some are "top-heavy," others "bottom-heavy," and others form peaks at mesoscales.

This taxonomy can be quantified by further information-theoretic measures that capture how complex the emergence in a system is, based on how causal contributions are spread across scales. Essentially, we introduce a literal notion of "scale-freeness" wherein if no one scale dominates the causal workings of a system, it has a high degree of emergent complexity. We also demonstrate how shifting network growth rules can in turn shift a system's

9

emergent hierarchy from "bottom" to "top-heavy," and this reveals how our notion of emergent complexity overlaps with the traditional notion of "scale-freeness" from network theory. To simulate these networks, and to expand our methods to larger systems, we introduce a heuristic that operates as a branching greedy sampling procedure that sets the stage for performing our analyses on empirical data.

Finally, we explore how our methods can be used to engineer the degree of emergence in a system with pinpoint precision. Specifically, we demonstrate the ability to design systems that possess only a single emergent scale of description in their emergent hierarchy, ensuring that the system's workings are driven by just one lone macroscale and no others.

3.1 Detecting and visualizing the emergent hierarchy of a system

Identifying the emergent hierarchy of a system is much like carving away until the underlying causal skeleton that spans scales is revealed. To ensure the clearest conceptual explanation, we first show an overview of our method on a small 5-state system (Figure 3), and then present some example systems which at the microscale are small enough (8 states) that we can calculate via brute force all possible partitions, TPMs, and CP values, but that nonetheless contain distinct emergent hierarchies (Figure 4).

For a system defined by some microscale TPM (the transitions of which are known as ground truth) and with all partitions and their associated TPMs and CP in hand, we use a *causal apportioning schema* that appropriately distributes out the CP across all the scales.

For any given scale, its CP is compared to all the CP values of all the scales of the system that are on a path below it in the lattice of scales. Its non-redundant and unique ΔCP value is its positive gain compared to the maximum CP value of any scale below it, as long as it the comparative scale is along any given micro→macro path that leads through it (panel A in Figure 3). This is somewhat like a game-theoretic quantity, asking how much each "player" (here, a scale) contributes to the overall game that is the system's workings (see the pseudocode in SI Section 2 for a formal description of this process). The result is the sublattice of remaining scales that have non-zero ΔCP values, which we refer to as the emergent hierarchy (panel B in Figure 3). It is still organized such that partitions of a certain size (a certain dimensionality) are represented in rows on the sublattice (such that for any given dimension, there can be a number of scales that share that same dimensionality, or level).
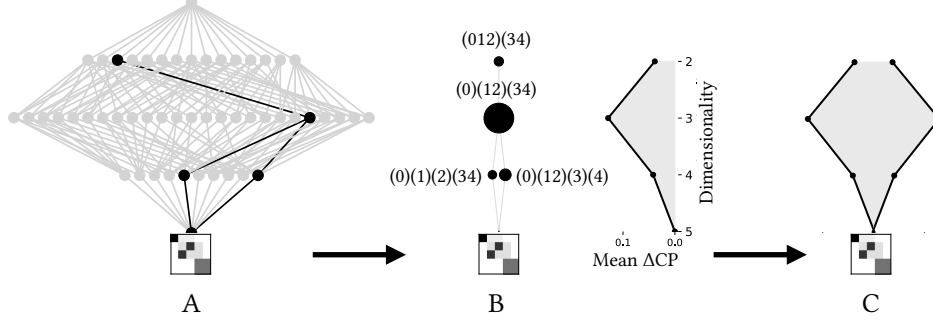
Figure 3: A visualization of our method on a small system with 5 states that correspond, respectively, to a source state, two cycle states, and two sink states. The microscale TPM is shown at the bottom of the panels. Panel A shows the full partition lattice over 5 states, which consists of 52 possible coarse-grainings. Highlighted in black are the scales with a positive ΔCP score that increase the CP value beyond any of the scales below them. Panel B shows the sublattice that contains only these five scales (the emergent hierarchy) with the corresponding partitions indicated next to the nodes. In the visualization of the emergent hierarchy, the size of the node reflects its ΔCP value. Note that the intuitively "correct" scale that partitioned the system into source, cycle, and sink nodes indeed contributes the most ΔCP. This is further shown by the vertical plot next to it that shows the average ΔCP value per row (per dimensionality of the coarse grained systems). Finally, panel C shows an abstracted representation of the hierarchy, which is just a symmetrized version of the vertical ΔCP plot. This representation is used in later figures to illustrate the difference between top-heavy, bottom-heavy, and complex systems.

While we refer to this procedure as an apportioning schema, it is generally not the case that the sum of all ΔCP values adds up to the maximum attainable CP of 1; however, we do note that in practice the sum of ΔCP reflects the total possible CP, and that this method has both the advantage of conceptual simplicity and resulting in a clearly distinguishable taxonomy of emergent hierarchies.

The emergent hierarchies for a variety of 8-state TPMs are shown in Figure 4 (top). There is a significant diversity between them: they form a kind of garden of different degrees and types of emergence, the taxonomy of which ranges from "top-heavy" (wherein most CP is concentrated near the ultimate macroscale), to more "bottom-heavy," to more "mesoscale" (wherein

CP peaks at midway, or thereabouts, along the dimensions of the system); some even look like "balloons" in that there is only a single emergent scale that contributes to the workings of the system. This taxonomy can also be visualized by looking at how the average non-zero ΔCP is distributed across the dimensions or levels of the system, plots of which are shown in Figure 4 (bottom).
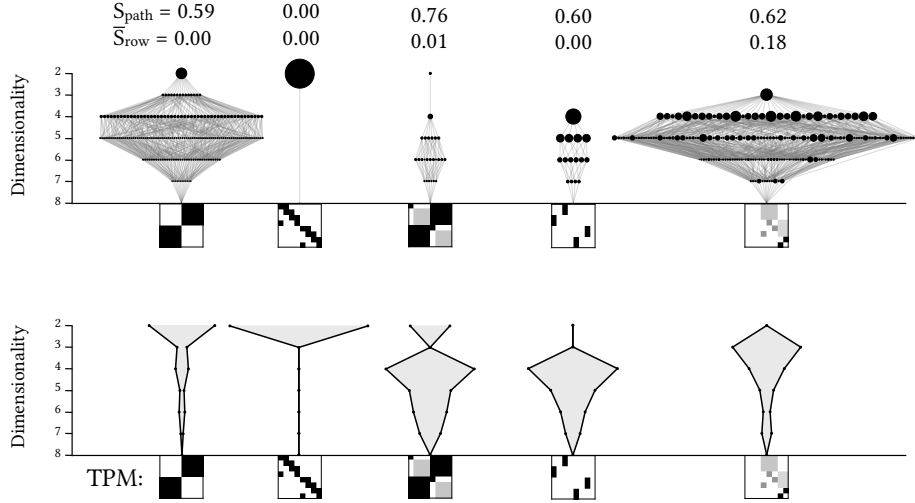

Figure 4: Systems can display causal emergence in many ways, here this is showcased as a "garden" of five systems (all the same size at the microscale) that have different emergent hierarchies. *Top:* The emergent hierarchies visualized across the eight levels of coarse-graining. We show the sublattice of the full partition lattice with positive ΔCP, i.e., the nodes that correspond to coarse-grainings that contribute CP beyond any of its ancestors in the lattice. The size of the node is proportional to ΔCP. From left to right, the systems correspond to i) a system composed of two equivalence classes of states, ii) a system with two length-4 cycles (explored in more detail in Figure 6), iii) a system with a mesoscale (taken from [22]), iv) a system with fully deterministic but degenerate transitions, and v) a system with three modules: three source states, a 3-cycle processing module, and 2 alternating sink states. *Bottom:* The width of the shaded region above each TPM corresponds to the average ΔCP at that level (a particular row in the sublattice) of the emergent hierarchy shown in the panel above it (as in Figure 3C) .

3.2 Complexity as literal scale-freeness

While the difference between emergent hierarchies can be seen qualitatively, here we introduce two mathematical tools which, in combination, quantify the complexity of an emergent hierarchy.

For the first tool, we follow the intuition that if a system's emergent hierarchy is heavily concentrated at just a few dimensionalities, it must be somewhat simple. This can be assessed by considering the paths that traverse the emergent hierarchy from the bottom microscale to the top macroscale (all the micro→macro paths). For a system with an L-dimensional microscale, these are paths of length L, with corresponding ΔCP values $(\Delta\mathrm{CP}_1, \ldots, \Delta\mathrm{CP}_L)$. We normalize these along the path to define:

$$p_i \;=\; \frac{\Delta\mathrm{CP}_i}{\sum_{j=1}^{L} \Delta\mathrm{CP}_j} \quad \text{for } i = 1, \ldots, L \tag{11}$$

which results in a normalized distribution, $\{p_1, \ldots, p_L\}$ over a single path, with associated entropy:

$$H(p) = -\sum_{i=1}^{L} p_i \log_2(p_i). \tag{12}$$

Note that this is identical to the emergent complexity measure defined for a single micro→macro path in CE 2.0 [22]. However, as we are extending CE 2.0 to beyond measuring a single path, we take the average of these path entropies across the full set of paths that traverse the emergent hierarchy and define $S_{\mathrm{path}} = \sum_p H(p)$, where the sum is over all paths from the bottom to the top of the emergent hierarchy.

Additionally, now that CE 2.0 can handle multiple paths, there is also more information about how complex the emergent hierarchy is beyond its spread across levels, since there may be multiple scales within a level (e.g., like $(1,2)(3,4)$ and $(1,3)(2,4)$), which share the same dimensionality and occupy the same row on the sublattice of the emergent hierarchy.

Therefore, we introduce a second way to judge the emergent complexity: how much differentiation there is between the scales at a particular level (i.e., to what degree the ΔCP values vary within rows in the sublattice of the emergent hierarchy).

Let Δ_i^l denote the ΔCP value for the ith l-dimensional partition. We normalize this per row as $\bar{\Delta}_i^l = \Delta_i^l / \sum_j \Delta_j^l$ to define

$$S_l = -\sum_i \bar{\Delta}_i^l \log_2 \bar{\Delta}_i^l \tag{13}$$

and define the row entropy as its average

$$S_{\mathrm{row}} = \frac{1}{L} \sum_l S_l \qquad (14)$$

When S_{row} is large, then there is a large degree of symmetry between the different scales at that level. To capture complexity, we therefore take its informational inverse, the *row negentropy*:

$$\overline{S}_{\mathrm{row}} = \log_2 L - S_{\mathrm{row}} \qquad (15)$$

Thus if the emergent hierarchy has positive ΔCP spread across its different levels (large S_{path}), *and* there is significant differentiation in terms of the size of ΔCP within those levels (large $\overline{S}_{\mathrm{row}}$), then that system possesses a complex emergent hierarchy, and no one scale dominates.

In Figure 4, for example, the system on the far left and on the far right are both "top-heavy," and therefore their path entropy is similar (shown at the top of the figure for each system). However, the system on the far right has much more differentiation in terms of ΔCP between the scales at any given level, and so has greater emergent complexity. In this, the emergent complexity measures how *literally scale-free* a system is in terms of its causal workings.

3.3 Growing literally scale-free systems

To demonstrate how our notion of emergent complexity is related to literal scale-freeness, we make use of network growth rules, using the derived networks to construct different microscale TPMs that have different emergent hierarchies with different properties.

Specifically, across a set of growing model networks we parameterize the preferential attachment, α, which changes the way a network grows as new nodes are added. For each different α, this creates a different network, which here we treated as a directed network representing a set of state transitions (assuming a uniform split of transition probabilities for any node with multiple outputs); i.e., in this interpretation, each node of the network represents a state, and each connection in the network represents a possible transition. In other words, instead of using the growth rule to just grow networks, we use it to grow TPMs by normalizing the adjacency matrix to derive transition probabilities.

During this process, if $\alpha = 1$ then the network is a classic Barabási-Albert network, also known as a "scale-free" network [27]. Scale-free networks are networks wherein connectivities between nodes follow a scale-free

power-law distribution, and thus there are a small number of extremely high-connectivity nodes (intuitively, this traditional notion of "scale-freeness" refers how the degree distribution doesn't change when "zooming out").

Such networks have important properties: they are commonly found in the real data and they have a tight link to criticality [28]. However, if during network growth $\alpha > 1$ then the resultant networks have superlinear preferential attachment, and are no longer scale-free, instead congealing into distinct clusters as α increases (eventually forming a starlike TPM where all states transition to the same state and therefore form one large macrostate). The other regime, where $\alpha < 1$, represents sublinear preferential attachment, which forms sparse and branching TPMs.

Differences in growth rules are best seen in larger networks, so therefore we developed a heuristic to reveal the emergent hierarchies of larger networks (this also sets the stage for our methods to eventually be used to analyze real data). Rather than pre-calculating the full set of possible scales and their CP, we developed a "branching greedy algorithm" which can start at the microscale and, by repeatedly traversing the set of system scales in different ways, average over the paths of the system to reveal how ΔCP is distributed (see SI Section 2 for the algorithm's details).
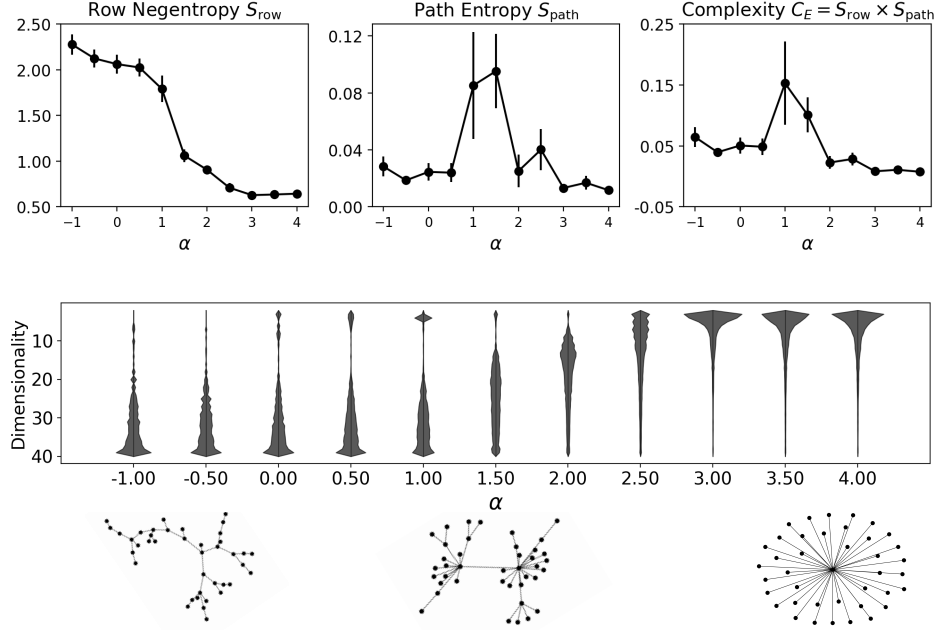
Figure 5: Preferential attachment networks show different emergent hierarchies depending on the preference parameter α, and display a phase transition around the $\alpha = 1$ scale-free point. The number of connections for newly introduced nodes is set to $m = 1$ throughout, in order to yield sparse and more hierarchical tree-like networks. *Top:* We show three measures that contribute to the system's complexity. The average negentropy of ΔCP across the levels of the system (Equation 15) is shown on the left (capturing the differentiation of ΔCP within levels). In the middle are plotted the average entropies of ΔCP across 100 uniformly sampled paths on the sublattice (capturing how uniformly ΔCP is spread across levels). For each α value, we report the mean and standard error across five sampled networks with 40 nodes each. The row negentropy sharply drops beyond $\alpha = 1$, which reflects a strong decrease in the differentiation of ΔCP within levels. On the right is plotted the multiplied values of both row negentropy and path entropy. Previous methods, like [22] and [23], do not take into account parallel paths and therefore couldn't capture this horizontal heterogeneity. *Bottom:* We show the mean ΔCP value for every dimension (in the same representation as in the bottom of Figure 4), averaged across five sampled networks at each α. The distribution of ΔCP can be seen to go from "bottom-heavy" at low α to "top-heavy" at high α, with the transition taking place between $\alpha = 1$ and $\alpha = 1.5$.

16

Our results demonstrate how the degree of causal emergence, and its distribution across the emergent hierarchy, changes dramatically under different regimes of preferential attachment. When $\alpha < 1$ the networks are in a bottom-heavy regime, where the scales that causally contribute cluster around the microscale. As α approaches 1, there is a phase shift wherein causal contribution distributions begin to shift upward (as captured by the increase in the path entropy) and also there is rapid increase in differentiation across the scales at any given level (captured by the negentropy of the distribution of ΔCP at a particular dimensionality, i.e., any row in the sublattice). Together, this means that the emergent complexity peaks at around the regime of $\alpha = 1$. However, in the superlinear regime of attachment, emergent hierarchies become increasingly top-heavy (a low path entropy) and at any given level ΔCP is quite uniform, and so the emergent complexity decreases.

Interestingly, the traditional notion of scale-freeness is linked to criticality [28], which is in turn tied to properties like information processing and computation capabilities [29], as well as adaptability and evolvability [30]. However, that traditional definition of "scale-free" is defined based solely on the power-law distribution of connectivity [27]. Our results identify an overlapping, but likely not identical, regime wherein networks have emergent hierarchies that are highly complex, spanning the levels of a system and varying within a level. That is, we find that the regime of connectivity wherein emergent hierarchies are complex represents a regime that is *literally* scale-free, in that no one scale of description of the network dominates, instead of just the degree distribution following a power law.

3.4   Engineering "pinpoint emergence"

In the previous section, differently chosen growth rules biased the kind of emergence exemplified by the system, leading to different emergent hierarchies. This indicates that emergence can be steered, and brings up the question: Can one engineer emergence with pinpoint precision?

An intriguing hint is an example within the garden of emergent hierarchies already shown in Figure 4, which is the "balloon"—where all the positive ΔCP in the entire emergent hierarchy is contained within an individual macroscale. Here we show two general methods for how construct such examples of pinpoint emergence.

Specifically, we make use of the fact that a cycle of transitions has no hierarchical structure: there exists no intermediate scale between the individual nodes and the full cycle at which the dynamics simplify. Contracting

17

the whole cycle, in contrast, yields a very efficient simplification, namely by representing the whole cycle as a single node with only a self-loop, drastically increasing CP. Figure 6 demonstrates examples of this phenomenon. To isolate the emergent scale at level l (that is, the coarse-grained system has l states), we introduce either a cycle of size $c \leq l$ and a set of $l - c$ permutations (possibly representing other levels that have already been simplified), or a total of l disjoint cycles. The result is that, out of an astronomical number of scales of description (even the small systems in Figure 6 have a total of 877 possible scales), only one single macroscale contributes to the causal workings of the system.
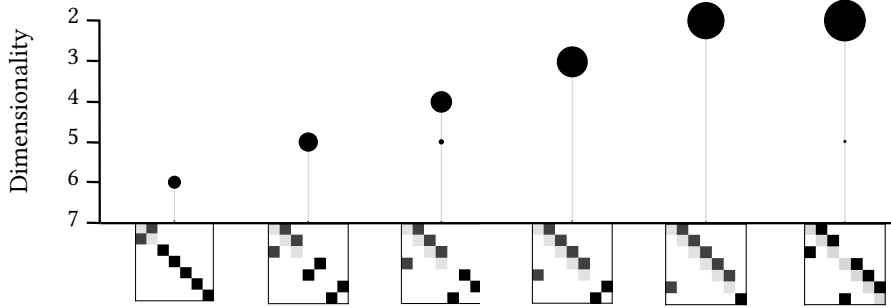

Figure 6: By introducing a set of cycles and/or permutations, the emergent scale can be engineered to be at any preferred dimensionality. Shown here are directed diffusion cycles with a probability of staying in a state or transitioning to the next point along the cycle of 0.2 and 0.8, respectively.

4   Discussion

There are significant benefits to being able to engineer emergence. First, emergent systems may be more robust, in that they are more resistant to knockouts at the microscale [31], and also in that their macroscales are cryptic [32]. As we have also seen here, complex emergent hierarchies appear to have a close relationship with scale-freeness and criticality [27, 28], even offering a new—quite literal—definition of scale-freeness as when no one scale dominates the causal workings of a system.

There has been increasing research on causal emergence (see Yuan et al. 2024 [33] for a review), based in the idea that macroscale causation can under some conditions be stronger (i.e., that macroscale causal relationships

are more robust, less noisy, more necessary) and ultimately this stems from a kind of causal error-correction via multiple-realizability [6, 34] . Recently there has been the introduction of the theory of causal emergence 2.0, an update which is more theoretically and conceptually complete than its predecessor, and which negates previous mathematical criticisms of the theory (e.g., [35]) that no longer apply to the new formalism; most importantly, as we demonstrate here, CE 2.0 distinguishes a taxonomy of different types of emergence, wherein a system's workings may be "top-heavy," "bottom-heavy," peak at a mesoscale, or be highly complex and so spread across the system's different levels (that is, at different degrees of dimension reduction).

It is natural to, for various systems, desire different emergent outcomes. E.g., we may want a system to have maximal emergent complexity (and given the prevalence of classic scale-freeness in real data [36], we should expect real systems to also show high degrees of emergent complexity in the same manner). In other cases, it is natural to want certain systems to operate precisely and solely at particularly scales of description, such that only those scales, and those alone, drive the causal workings of the system. Important examples might be, for instance, engineering brain-computer interfaces to operate in synchrony with the relevant brain scales that correlate to human psychology or the neural correlates of consciousness [37, 38].

More generally, we posit it is a requisite for human freedom and autonomy that the scale of psychology (or the neural correlates of consciousness) is a contributing member of the emergent hierarchy of the brain and body, adding significantly to the system's total causal workings. Critically, we point out that the causal relevancy of human psychology must be preserved in any sort of AI-human merger, such as those being attempted via brain implants [39], and that the ability to engineer emergence may be necessary to prevent loss of human control, and that this area of research is currently acutely understudied.

Several limitations of our paper are worth discussing, mostly in that it is focused on proof-of-principle results rather than large-scale empirical data. Even confined to systems of states we study, these are either small (so that we can brute force compare all the scales) and, even when analyzing the network growth rules, still not very large (only 40 states), so that we can easily run the branching greedy algorithm. However, in the future, we plan to apply this to larger networks, as currently our algorithm (detailed in SI Section 3) scales in subexponential time, and further speedups using parallelization are feasible.

Throughout our analysis we use a $p(C)$ that is uniform for any TPM we consider, which reflects the assumption that each scale is treated inde-

pendently during causal analysis (e.g., you don't count the atoms in a light switch to assess the impact of the intervention of do(light switch = UP) on a connected light bulb, you instead compare it to do(light switch = DOWN)). A different choice would be to instead coarse-grain the microscopic $p(C)$; however, then the coarse-grained versions of $p(C)$ would merely correspond to an observational distribution based on interventions at a system's microscale (which violates how causation should go beyond observation) [3]). To avoid this, we follow the equivalent of performing a randomized controlled trial (the heart of the scientific understanding of causation [40]) at each scale to understand what its states do. While this is a defensible mathematical assumption for dealing with causation, it is important to note that, unlike its earlier version CE 1.0 [34, 35, 41], the mere existence of emergence in CE 2.0 does not depend on any specific assumptions regarding $p(C)$, e.g., such as using the uniform distribution. Indeed, gains in the causal primitives (and the many measures of causation based on them) can be identified across different assumptions of what $p(C)$ should be [24], and causal emergence can still occur even in cases where our assumption of scale-independence is violated, such as when the $p(C)$ used at a macroscale is purely observational and so is just a direct coarse-graining of the microscale $p(C)$ [22].

It's also important to note that not all macroscales herein are *dynamically consistent* [42]. Unlike in previous research on CE 1.0 and 2.0, we do not restrict our analysis to the set of scales which are dynamically consistent: i.e., such that the dynamics of the microscale are preserved (albeit coarse-grained) up at the macroscale. This is simply because we found that checking for dynamical consistency did not alter our results in a significant way (data not shown).

Future research will explore opportunities for synthesis with existing theories of emergence or alternative or improved methods for how to partition or apportion causation across scales. For instance, in terms of causal emergence, there has been a number of efforts to provide alternative versions to CE 1.0, such as those based instead on information decomposition [43] or dynamical reversibility [25]. These methods will likely need to be expanded to better reflect the improvements of CE 2.0. Additionally, there have recently been proposed novel methods for decomposing interventional causality [44], as well as a new approach to Shapley values and Möbius inversions [45] that might serve as a future potential causal apportioning schema wherein the total CP is preserved when summed across the sublattice.

Overall, our results show that CE 2.0 framework can be extended to account for multiple paths from the microscale to the macroscale, revealing the full emergent hierarchy of scales. This emergent hierarchy—the sub-

set of all possible scales that actually causally contribute to the system's workings—has evocative and rich mathematical properties to explore, and gives significant information about the nature of emergence occurring in the system. Additionally, it opens the door for designing systems to possess desired emergent hierarchies, with important implications for science and engineering and the construction of causal models.

5   Acknowledgments

Supplementary Information

S1 Ground-truth TPMs as causal models

TPMs are often inferred from observational data with incomplete knowledge of the causal graph, and therefore generally do not encode causal effects. However, we here use them to *encode* a causal model by assuming ground-truth knowledge, in that the system at time t is fully characterized by a state X_t (there are no hidden confounders), and the dynamics are Markovian, so that the (somewhat trivial) causal graph can be written as

$$\ldots \to X_{t-1} \to X_t \to X_{t+1} \to X_{t+2} \to \ldots \tag{16}$$

Interventions of the type $do(X_t = x)$ correspond to putting the system in a certain state x. By the second rule of the do-calculus:

$$p(X_{t+1}|do(X_t = x)) = p(X_{t+1}|X_t = x) \tag{17}$$

if X_{t+1} is independent of X_t in the graph with the outgoing arrows of X_t removed. This is clearly true in this case:

$$\ldots \to X_{t-1} \to X_t \qquad X_{t+1} \to X_{t+2} \to \ldots \tag{18}$$

Since a system's TPM is simply a way to summarize $p(X_{t+1}|X_t = x)$ for all x, we refer to such relationships as *causal*. Since the states are always over the entirety of the system, we refer to any prior state as a *cause*, and any of its immediately following states an an *effect*.

For simplicity, we focus solely on the causal relationships between the states of a system (and their macroscales, in the form of sets of macrostates). Admittedly, such TPMs do not uniquely specify a given system's mechanisms or elements. Therefore, this work could be extended to be over causal models consisting of mechanisms and elements as well (e.g., as in [7, 46]), rather than just causal models of states, although we feel for an initial treatment this adds unnecessary complexity (both conceptual and otherwise).

S2 Pseudocode of CE 2.0's calculation across all possible scales of a system.

Algorithm 1: Calculating $\Delta\text{CP}(T)$

Input : Row-stochastic matrix $T \in \mathbb{R}^{n \times n}$;
Output: Hasse diagram H over partitions;
            CPdict : partition $\mapsto \mathbb{R}$;
            ΔCPdict : partition $\mapsto \mathbb{R}$;
            Emergent: partitions with $\Delta\text{CP} > \varepsilon$.
$n \leftarrow$ number of states in T

1) Enumerate coarse-grainings
$\mathcal{P} \leftarrow \textsc{GenerateAllCoarseGrainings}(n)$

2) Compute CP on each partition
CPdict $\leftarrow \emptyset$
foreach $P \in \mathcal{P}$ **do**
    $\lfloor$ CPdict$[P] \leftarrow \text{CP}(\textsc{CoarseGrain}(T, P))$

3) Build refinement lattice (Hasse diagram)
$H \leftarrow \textsc{BuildRefinementGraph}(\mathcal{P})$

4) ΔCP relative to ancestors
ΔCPdict $\leftarrow \emptyset$
foreach $P \in \textsc{Nodes}(H)$ **do**
    $\mathcal{A} \leftarrow \textsc{Ancestors}(H, P)$
    if $\mathcal{A}$ *is empty* **then**
       $\mid \quad b \leftarrow 0$
    else
       $\lfloor \quad b \leftarrow \max\{\text{CPdict}[Q] : Q \in \mathcal{A}\}$
    ΔCPdict$[P] \leftarrow$ CPdict$[P] - b$

5) Emergent set (positive ΔCP)
Emergent $\leftarrow \{\, P \in \mathcal{P} \mid \Delta\text{CP}[P] > \varepsilon \,\}$

return (H, CP, ΔCP, Emergent)

S3   A branching greedy algorithm for estimating the emergent hierarchy of a system.

Algorithm 2: GreedyCompletion from a starting partition

Function GreedyCompletion(T, P_{start})**:**

    $P \leftarrow P_{\text{start}}$;   Path $\leftarrow [P]$;

    CPlist $\leftarrow [\,\text{CP}(\text{CoarseGrain}(T, P))\,]$

    while P *has more than one block* **do**

        $k \leftarrow$ number of blocks in P; $(\text{best}, P^{\star}) \leftarrow (-\infty, \bot)$

        foreach *unordered pair* (i, j) *of block indices in* $\{1, \dots, k\}$ **do**

            $Q \leftarrow$ merge blocks i and j in P;

            $e \leftarrow \text{CP}(\text{CoarseGrain}(T, Q))$;

            if $e > best$ **then**

                $\text{best} \leftarrow e$; $P^{\star} \leftarrow Q$

        append $P^{\star}$ to Path;   append best to CPlist; $P \leftarrow P^{\star}$

    return (Path, CPlist)

Algorithm 3: $\Delta\mathrm{CP}(T)$ from parallel greedy paths

---

Input : Row-stochastic matrix T; n_paths $\in \mathbb{N}$ (default 3).
Output: H: refinement DAG over sampled partitions;
　　　　　CPdict: partition $\to$ CP;
　　　　　ΔCPdict (via Alg. 1, Steps 4–5).

1) Parallel greedy sampling of high-scoring branches
$P_{\mathrm{cur}} \leftarrow \textsc{FinestGraining}(T)$
$\mathrm{CPdict} \leftarrow \{\, P_{\mathrm{cur}} \mapsto \mathrm{CP}(T) \,\}$
while P_{cur} *has more than one block* **do**
    $k \leftarrow$ number of blocks in P_{cur};
    $\mathcal{C} \leftarrow \emptyset$
    foreach *unordered pair (i,j) of block indices in $\{1,\ldots,k\}$* **do**
        $Q \leftarrow$ merge blocks i and j in P_{cur};
        $e \leftarrow \mathrm{CP}(\textsc{CoarseGrain}(T,Q))$;
        append (e,Q) to $\mathcal{C}$
    $\mathrm{Top} \leftarrow$ the $\min(\text{n\_paths}, |\mathcal{C}|)$ pairs in $\mathcal{C}$ with largest e
    $\mathcal{S} \leftarrow [\, Q : (e,Q) \in \mathrm{Top}\,]$ // selected starting merges
    foreach $Q \in \mathcal{S}$ **do**
        $(\mathrm{Path}, \mathrm{CPlist}) \leftarrow \textsc{GreedyCompletion}(T,Q)$
        for $t = 1$ **to** $|Path|$ **do**
            if $Path[t] \notin \mathrm{CPdict}$ **then**
                $\mathrm{CP}[\mathrm{Path}[t]] \leftarrow \mathrm{CPlist}[t]$
    $P_{\mathrm{cur}} \leftarrow$ first element of $\mathcal{S}$

2) Build refinement lattice over sampled nodes
$H \leftarrow \textsc{BuildRefinementGraph}(\mathrm{keys}(\mathrm{CPdict}))$

3) Apply 1, Steps 4–5, with (H, CPdict) to obtain ΔCPdict.

References

[1] Yaneer Bar-Yam. *Dynamics of complex systems*. crc Press, 2019.

[2] Herbert A Simon. The architecture of complexity. In *The Roots of Logistics*, pages 335–361. Springer, 2012.

[3] Judea Pearl. *Causality*. Cambridge university press, 2009.

[4] Judea Pearl and Dana Mackenzie. *The book of why: the new science of cause and effect*. Basic books, 2018.

[5] Abel Jansma. Mereological approach to higher-order structure in complex systems: From macro to micro with möbius. *Physical Review Research*, 7(2):023016, 2025.

[6] Erik Hoel. *The world behind the world: Consciousness, Free Will, and the Limits of Science*. Simon and Schuster, 2024.

[7] Erik P Hoel, Larissa Albantakis, and Giulio Tononi. Quantifying causal emergence shows that macro can beat micro. *Proceedings of the National Academy of Sciences*, 110(49):19790–19795, 2013.

[8] Andrés Gómez-Emilsson and Chris Percy. The "slicing problem" for computational theories of consciousness. *Open Philosophy*, 5(1):718–736, 2022.

[9] Joshua Bongard and Michael Levin. There's plenty of room right here: Biological systems as evolved, overloaded, multi-scale machines. *Biomimetics*, 8(1):110, 2023.

[10] Rafael Yuste. From the neuron doctrine to neural networks. *Nature reviews neuroscience*, 16(8):487–497, 2015.

[11] David Marr. *Vision: A computational investigation into the human representation and processing of visual information*. MIT press, 2010.

[12] Fernando E Rosas, Bernhard C Geiger, Andrea I Luppi, Anil K Seth, Daniel Polani, Michael Gastpar, and Pedro AM Mediano. Software in the natural world: A computational approach to emergence in complex multi-level systems. *arXiv preprint arXiv:2402.09090*, 2024.

[13] Lionel Barnett and Anil K Seth. Dynamical independence: discovering emergent macroscopic processes in complex dynamical systems. *Physical Review E*, 108(1):014304, 2023.

[14] Carlos Gershenson and Nelson Fernández. Complexity and information: Measuring emergence, self-organization, and homeostasis at multiple scales. *Complexity*, 18(2):29–44, 2012.

[15] Emiliano Marchese, Guido Caldarelli, and Tiziano Squartini. Detecting mesoscale structures by surprise. *Communications Physics*, 5(1):132, 2022.

[16] Jaegwon Kim. *Mind in a physical world: An essay on the mind-body problem and mental causation*. MIT press, 2000.

[17] Thomas D Bontly. The supervenience argument generalizes. *Philosophical Studies*, 109:75–96, 2002.

[18] Denis Noble. A theory of biological relativity: no privileged level of causation. *Interface focus*, 2(1):55–64, 2012.

[19] Jerry A Fodor. Special sciences, or the disunity of science as a working hypothesis. In *The language and thought series*, pages 120–133. Harvard University Press, 1980.

[20] Ned Block. Do causal powers drain away? *Philosophy and Phenomenological Research*, 67(1):133–150, 2003.

[21] Erik P Hoel. Agent above, atom below: how agents causally emerge from their underlying microphysics. *Wandering towards a goal: how can mindless mathematical laws give rise to aims and intention?*, pages 63–76, 2018.

[22] Erik Hoel. Causal emergence 2.0: Quantifying emergent complexity. *arXiv preprint arXiv:2503.13395*, 2025.

[23] Brennan Klein and Erik Hoel. The emergence of informative higher scales in complex networks. *Complexity*, 2020(1):8932526, 2020.

[24] Renzo Comolatti and Erik Hoel. Consilience in causation: Causal emergence is found across measures of causation. *Entropy*, 27(8):825, 2025.

[25] Jiang Zhang, Ruyi Tao, Keng Hou Leong, Mingzhe Yang, and Bing Yuan. Dynamical reversibility and a new theory of causal emergence based on svd. *npj Complexity*, 2(1):3, 2025.

[26] Branden Fitelson and Christopher Hitchcock. *Probabilistic measures of causal strength.*

[27] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[28] Maximino Aldana and Philippe Cluzel. A natural class of robust networks. *Proceedings of the National Academy of Sciences*, 100(15):8710–8714, 2003.

[29] Woodrow L Shew and Dietmar Plenz. The functional benefits of criticality in the cortex. *The neuroscientist*, 19(1):88–100, 2013.

[30] Jorge Hidalgo, Jacopo Grilli, Samir Suweis, Miguel A Munoz, Jayanth R Banavar, and Amos Maritan. Information-based fitness and the emergence of criticality in living systems. *Proceedings of the National Academy of Sciences*, 111(28):10095–10100, 2014.

[31] Brennan Klein, Erik Hoel, Anshuman Swain, Ross Griebenow, and Michael Levin. Evolution and emergence: higher order information structure in protein interactomes across the tree of life. *Integrative Biology*, 13(12):283–294, 2021.

[32] Erik Hoel and Michael Levin. Emergence of informative higher scales in biological systems: a computational toolkit for optimal prediction and control. *Communicative & Integrative Biology*, 13(1):108–118, 2020.

[33] Bing Yuan, Jiang Zhang, Aobo Lyu, Jiayun Wu, Zhipeng Wang, Mingzhe Yang, Kaiwei Liu, Muyun Mou, and Peng Cui. Emergence and causality in complex systems: A survey of causal emergence and related quantitative studies. *Entropy*, 26(2):108, 2024.

[34] Erik P Hoel. When the map is better than the territory. *Entropy*, 19(5):188, 2017.

[35] Frederick Eberhardt and Lin Lin Lee. Causal emergence: When distortions in a map obscure the territory. *Philosophies*, 7(2):30, 2022.

[36] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.

[37] Francis Crick and J Clark. The astonishing hypothesis. *Journal of Consciousness Studies*, 1(1):10–16, 1994.

[38] Anil K Seth and Tim Bayne. Theories of consciousness. *Nature reviews neuroscience*, 23(7):439–452, 2022.

[39] Elon Musk et al. An integrated brain-machine interface platform with thousands of channels. *Journal of medical Internet research*, 21(10):e16194, 2019.

[40] Lloyd D Fisher. Advances in clinical trials in the twentieth century. *Annual Review of Public Health*, 20(1):109–124, 1999.

[41] Scott Aaronson. Higher-level causation exists (but I wish it didn't), jun 2017.

[42] Paul K Rubenstein, Sebastian Weichwald, Stephan Bongers, Joris M Mooij, Dominik Janzing, Moritz Grosse-Wentrup, and Bernhard Schölkopf. Causal consistency of structural equation models. *arXiv preprint arXiv:1707.00819*, 2017.

[43] Pedro AM Mediano, Fernando E Rosas, Andrea I Luppi, Henrik J Jensen, Anil K Seth, Adam B Barrett, Robin L Carhart-Harris, and Daniel Bor. Greater than the parts: a review of the information decomposition approach to causal emergence. *Philosophical Transactions of the Royal Society A*, 380(2227):20210246, 2022.

[44] Abel Jansma. Decomposing interventional causality into synergistic, redundant, and unique components. In *Proceedings of the 39th Conference on Neural Information Processing Systems (NeurIPS 2025)*, 2025. (forthcoming) Preprint at arXiv:2501.11447.

[45] Patrick Forré and Abel Jansma. Möbius transforms and shapley values for vector-valued functions on weighted directed acyclic multigraphs, 2025.

[46] William Marshall, Larissa Albantakis, and Giulio Tononi. Black-boxing and cause-effect power. *PLoS computational biology*, 14(4):e1006114, 2018.