Associative conditioning in gene regulatory network models increases integrative causal emergence

Check for updates

Federico Pigozzi[1], Adam Goldstein[2] & Michael Levin [1,3] ✉

How does learning affect the integration of an agent's internal components into an emergent whole? We analyzed gene regulatory networks, which learn to associate distinct stimuli, using causal emergence, which captures the degree to which an integrated system is more than the sum of its parts. Analyzing 29 biological (experimentally derived) networks before, during, after training, we discovered that biological networks increase their causal emergence due to training. Clustering analysis uncovered five distinct ways in which networks' emergence responds to training, not mapping to traditional ways to characterize network structure and function but correlating to different biological categories. Our analysis reveals how learning can reify the existence of an agent emerging over its parts and suggests that this property is favored by evolution. Our data have implications for the scaling of diverse intelligence, and for a biomedical roadmap to exploit these remarkable features in networks with relevance for health and disease.

When is a system more than the sum of its parts? When and how do the properties of active components enable the emergence of a high-level, integrated decision-making entity[1–5]? These questions bear on issues in ecology, philosophy mind, psychiatry, swarm robotics, and developmental biology[6–13]. In a sense, all intelligence is collective intelligence[14,15] because even human minds supervene on a collection of cells which are themselves active agents. One practical way to define integrated emergent systems is by the fact that they have goals, memories, preferences, and problem-solving capabilities that their parts do not have. For example, while individual cells solve problems in metabolic, physiological, and transcriptional spaces, what makes an embryo more than a collection of cells is the alignment of cellular activity toward a specific outcome in anatomical morphospace[16]. Here, we focus on one aspect of emergent agency: integrated, distributed memory.

When a rat learns to press a lever to receive a reward, the cells at the paw touch the lever, those in the gut receive the delicious food - no individual cell has both experiences. The "rat" is the owner of the associative memory that none of its parts can have. This ability to bind together individual experiences of their parts is a hallmark of emergent agents. The rat can do associative learning because it has the right causal architecture (implemented by the nervous system) to integrate information across space and time within its body. However, this ability is not unique to brainy animals—various kinds of problem-solving and learning occur in single cells (reviewed in refs. 17–19) because biology

fundamentally exploits a multi-scale competency architecture in which the molecular components within a cell are likewise integrated to provide system-level context-sensitive responses.

Regardless of specific material implementation, certain functional topologies exhibit high emergent integration. In recent years, this topic has moved from philosophical debates over supervenience and downward causation to empirical science, as quantitative methods have been developed to measure a degree to which a system is more than its parts and possesses higher levels of organization that do causal work distinct from its lowest level mechanisms[20–23]. This now enables a study of the relationship between minimal cognition and collective intelligence. A degree of integration among parts is required for any amount of cognitive function, such as learning. Here, we explore the inverse hypothesis: could the process of learning increase integration within a system? That is, could training a system reify and strengthen the existence of it as a unified, emergent, virtual governor[24]?

To study this question in the most minimal model system, in which all the components are well-defined, deterministic, and transparent, we chose Gene Regulatory Networks. GRN models represent sets of gene products that up- or down-regulate each other's activity based on a given functional connectivity map[25]. These networks are very important topics in biomedicine[26–28], evolutionary developmental biology[29–31], and synthetic biology[32–35]. It is essential to be able to not only predict their behaviors, but also to induce desired dynamics for interventions in regenerative medicine

[1]Allen Discovery Center at Tufts University, Medford, MA, USA. [2]Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, UK. [3]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA, USA. ✉e-mail: michael.levin@allencenter.tufts.edu

and bioengineering[36–41]. These networks are well-recognized to have emergent properties[42–48], yet their control remains challenging[49–55]. We recently showed that such networks can be trained by providing stimuli on chosen nodes and reading out responses on other nodes. Well-known paradigms from behavioral and cognitive science can be used to predictably change future responses as a function of experience. Biological networks show several different kinds of learning, including Pavlovian conditioning. Thus, we sought to measure emergent integration in these networks before, during, and after training, to determine what effect the induction of memory has on the network as a coherent agent.

To quantify this property, we took advantage of tools from neuroscience, where several measures of "integrated information" have been proposed to explain how the activity of individual neurons gives rise to a unified emergent mind[56–59]. Different flavors of integrated information have been applied to study information dynamics in, among the others, natural evolution[60], genetic information flow[61], and biological and artificial neural systems[62], and have even been suggested as indicators of "consciousness"[63]. Without seconding any of its claims about consciousness, we adopt the framework of Integrated Information Decomposition (ΦID)[64] to specifically define a measure of *causal emergence*: it quantifies to what extent the whole system provides information about the future evolution that cannot be inferred by any of its individual components, or, in other words, the extent to

which the system behaves as a collective whole[65]. Intuitively, the higher causal emergence, the stronger the integration (or inseparability) of a collective of components. ΦID provides a rigorous framework to study causal emergence in a variety of different systems, from Conway's game of life[66] to human and primate neural dynamics[67], including for the comatose[68] and the brain-injured[69].

Here, we analyze in silico simulations showing how causal emergence increases in response to associative training in specific GRNs, characterize the GRNs' integration behaviors, and uncover a relationship between this phenomenon and the underlying biology (phylogeny and gene ontology of specific networks).

Results

We adopted the recent framework of Integrated Information Decomposition ΦID[64], an established approach to quantify causal emergence, in the dynamics of 29 GRNs (described as Ordinary Differential Equations, or ODEs) from the BioModels database[70], before and after training the networks for an associative memory task. We assessed how learning strengthens causal emergence in basal biological systems like regulatory networks, and how this results in qualitatively different patterns of behavior of the GRNs related to the biology (phylogeny and gene ontology) of the networks.
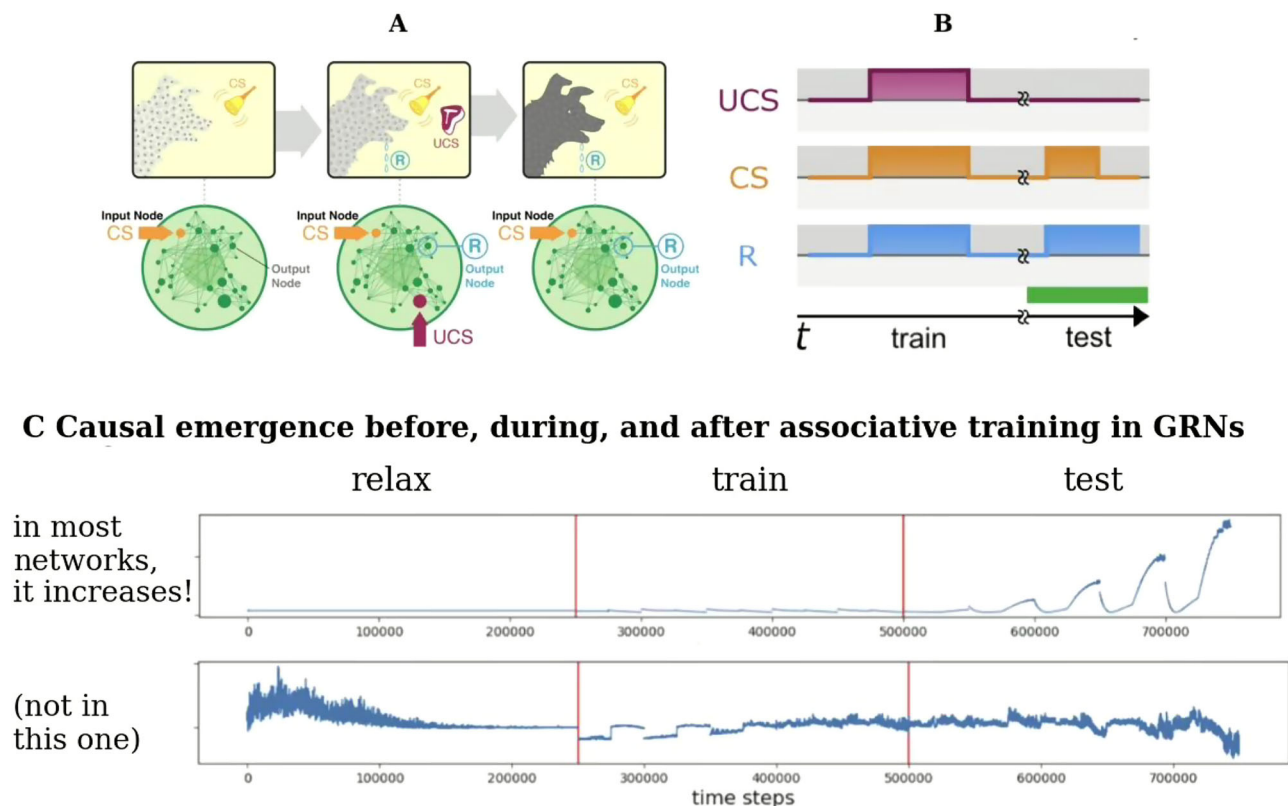

Fig. 1 | Learning in gene-regulatory networks. A Pavlovian conditioning and applications to gene regulatory networks. The standard paradigm is that the Conditioned Stimulus (CS) is initially neutral in that it does not elicit a Response (R), while an Unconditioned Stimulus (UCS) does. Some systems, like dogs, eventually show a response to the CS after it has been presented together with the UCS (forming an association between those stimuli). The same paradigm has been used with GRNs and pathways by mapping the CS, UCS, and R roles to specific nodes in the network and stimulating them by transiently raising their activity level (for example, by increasing their expression via a chemical ligand). Associative memory requires a degree of integration within the network, and here we explore the hypothesis that associative conditioning reifies the learning agent by increasing its emergent integration metrics (schematized by the dog progressively becoming more of a unified, centralized agent and less a collection of cells). Reported with permission from[86]. **B** Associative conditioning in simulated gene regulatory networks; simulation time is on the x-axis, while gene expression levels are on the y-axis. During the training phase, we pair UCS and CS stimuli to regulate a response R. If associative conditioning has taken place, we observe the CS alone (i.e., with no UCS stimulation) regulates R. In the schematic, we illustrate stimulation/regulation as upticks of the expression levels over a baseline value; in reality, gene expression can have a quantitatively different shape, but the principle remains the same. Used with permission from[87]. **C** How causal emergence (quantified with Integrated Information Theory[64]) changes over simulated time in two sample gene regulatory networks. For the network on top, it increases due to associative training; for the network on bottom, on the other hand, the same is not true. The diverse relationships between training and causal emergence are studied across the set of networks we analyzed in subsequent figures.
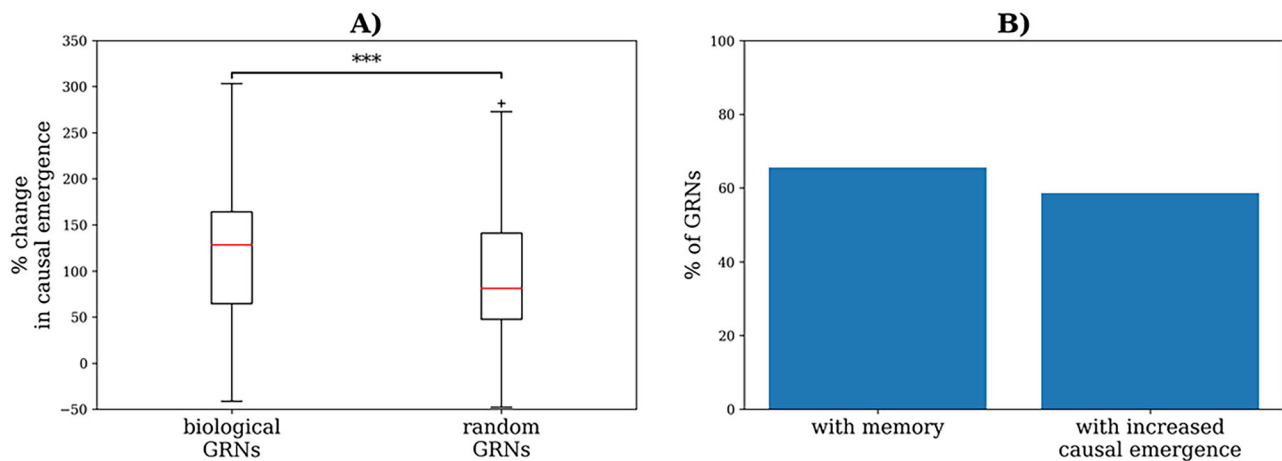
Fig. 2 | Biological networks with memory exhibit an increase of causal emergence during training and do so more than random networks. A Error bars for the % change in causal emergence from before to after training for biological and random networks; each point corresponds to one circuit of one network. Biological networks are significantly more causally emergent after training. The asterisks above the brackets indicate significance at $p < 0.001$ with the Mann-Whitney test. **B** Bars for the % of those networks that have memory and those that, if they have memory, show an increase in causal emergence after training. Associative training results in increased causal emergence among almost all networks with memory.

We illustrate how associative memory works in Fig. 1A, B. For each network, we pre-tested every triplet of its nodes (a circuit in which we assign nodes to the roles of unconditioned stimulus (UCS), a neutral stimulus (NS), and a response (R)) to determine whether it passed the test for associative memory[70]; similarly to Pavlovian conditioning in animals. We did so by: ensuring that stimulating the UCS alone triggers an increase in R, ensuring that stimulating the NS alone does not trigger R, and then: (1) "relaxing" the network in the initial phase; (2) stimulating both unconditioned and neutral stimuli during the subsequent training phase; (3) verifying that after the paired exposure, stimulation of the NS alone regulated the R (see "Materials and Methods"). Out of all the circuits, 808 (among 19 networks) passed this pretest and we considered these for all subsequent analyzes.

We then applied the ΦID to compute causal emergence on the genes' expression signals, exploiting it as an exhaustive measurement of all the ways a macroscopic (i.e., whole network) feature could affect the future of any parts of the network (including the network itself). Intuitively, this definition quantifies the degree to which the whole system influences the future in a way not discernible by considering the parts only (schematized by the dog of Fig. 1 progressively becoming more of a unified, centralized agent and less a collection of cells). Causal emergence is a numerical quantity measured in natural units of information; the higher it is, the more "emergent" the system is, in the sense that the more the "macro" beats the "micro" in explaining the system dynamics[71,72]. In Fig. 1C, we report a key finding: that associative training raises causal emergence in most of the GRNs studied.

Emergence increases after training

We set out to study how causal emergence changes before and after training GRNs for associative memory, and specifically to test the conjecture that the experience of learning an associative task raises the causal integration of the system. Because we also wanted to know whether biological networks have any unique properties in this regard, we constructed a set of 145 random networks as controls, by the established gene circuit method (see "Materials and Methods")[71], which involves randomizing the topology and connection strengths of these pseudo-biological networks with different random seeds. We computed the average % change in causal emergence from before to after "training" with paired stimuli (Fig. 2A), where each point corresponds to one circuit of one network for biological and random networks. We also plotted in Fig. 2B the ratio of biological GRNs that had any associative memory and witnessed an increase in causal emergence from before to after "training" in most networks. Finally, we verified the change in causal emergence was persistent by simulating the networks for longer without any stimulation and found the trend to persist and not be different from what was observed during the test phase.

In total, causal emergence was strengthened in 17 out of 19 networks. The change amounted, on average, to a $128.32 \pm 81.31\%$ increase from before to after training. This result was significant ($p < 0.001$) with the Wilcoxon signed-rank test for paired data, indicating that the two samples of before and after training came from statistically different distributions, confirming that training for associative memory increased causal emergence for most of the networks studied. Random networks had an average change of $56.25 \pm 51.40\%$ and the difference with biological networks was significant ($p < 0.001$). Instead, random networks had higher levels of absolute causal emergence before training. In other words, random networks started with higher emergence but did not increase it, while biological networks started with lower emergence but increased it with experience.

Previous network structure and function classifications do not capture emergence

We next sought to test whether our ΦID results captured new information about these networks that could not be derived with other previously existing metrics. To this end, we characterized the GRNs along existing structural and functional metrics. For all the circuits with memory, we computed established properties from network theory (in-degree, out-degree, betweenness centrality, PageRank, and HITS scores) for the structure and from dynamical systems literature (sample entropy, Lyapunov exponents, correlation dimension, detrended fluctuation analysis, and generalized Hurst exponent) for the function. We found no significant or only weak correlation with the change in causal emergence (Fig. 3A, B). These findings reveal that our experimental protocol uncovers new information that existing classifications of networks do not capture: how training GRNs affects causal emergence.

We next tested whether ΦID corresponds to mere network "activity" (i.e., checking that periods of low integration did not merely correspond to quiescent periods of low signaling). We characterized overall network "activity" as the first derivative of the ODE state variables and found no correlation with % change in causal emergence (Fig. 3C), revealing that causal emergence is not the same as network activity. If, for example, causal emergence spikes up, it is not an artifact of increased network activity. Similarly, if causal emergence drops, it is not attributable to a drop in signaling among nodes. In other words, we found that the dynamics of causal emergence are distinct from the levels of native or induced activity within the networks.
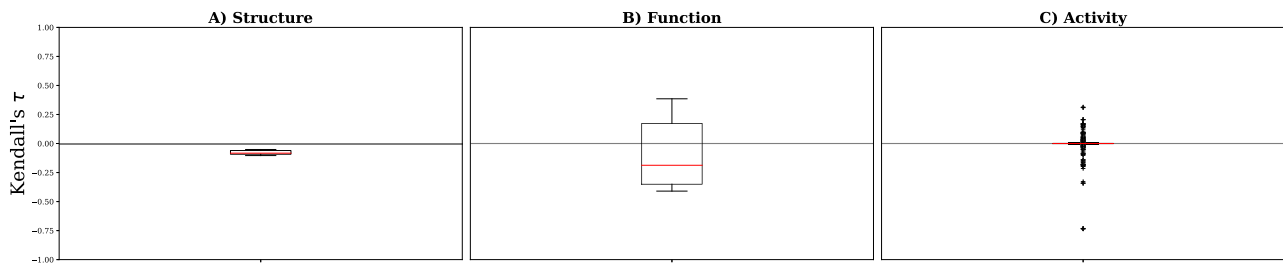
Fig. 3 | Change in causal emergence is not correlated with established structure, function, and activity metrics for networks. Kendall's rank correlation coefficient between % change in causal emergence and network **A** structure properties, **B** function properties, and **C** activity (measured as the first derivative of the state variables). Our causal emergence metric captures an aspect that neither established network properties nor mere activity encompass: how gene regulatory networks react to training.

Fig. 4 | GRNs exhibit distinct patterns of causal emergence with respect to the effects of training. Panels show five sample causal emergence trajectories (one gene regulatory network per row) of the five automatically discovered behaviors during the test phase. Each behavior represents trajectories sharing a specific and distinct temporal pattern.
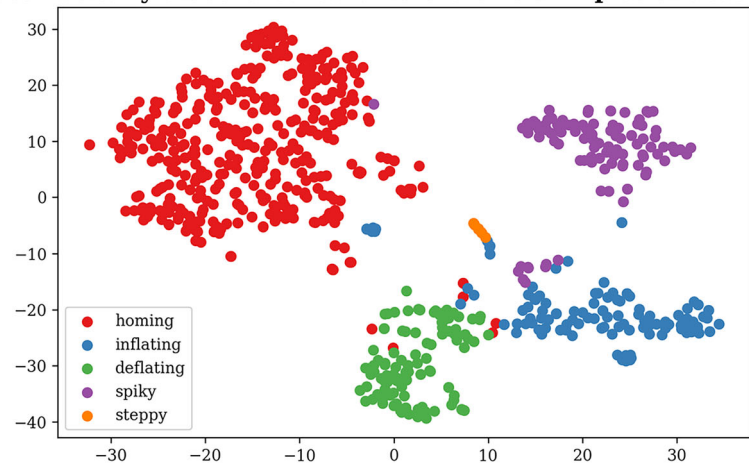
Automatic classification of emergence trajectories into behaviors finds five "species"

We observed several different ways in which integration of a network changed due to the training phase (Fig. 4). We wondered if the effects of training on integration across networks were highly diverse (forming a smooth space of possible effects of training), or whether there would be discrete categories of effects which define a kind of "species" with respect to how GRNs' causal emergence affected training. We first described each test phase's causal emergence trajectory using behavioral descriptors; alternatively, we could have extracted learned features through a neural network, but these would have been less interpretable[72]. We settled on seven descriptors we found (after manual search) to be, at the same time, the most expressive (in terms of quality of the classification) and compact: trend, monotonicity, flatness, number of peaks, average distance among peaks, average difference among peaks, and range (see "Materials and Methods" for detailed procedure).

We applied k-means[73], an established unsupervised learning technique, to automatically classify the extracted behavioral descriptors from each test phase of the circuits that have memory. This technique revealed discrete clusters in terms of the behavior descriptors (how emergence changes due to training), in effect classifying networks into "species" of individuals that exhibited similar effects of training upon their causal emergence. We tuned the number of species to be the optimal one according to the Silhouette coefficient of quality of a clustering. We found five optimal "species" of behavior and nicknamed them *homing, inflating, deflating, spiky,* and *steppy* after their characteristics (see next section). We plotted the t-SNE[74] embedding in 2D of the descriptors of each test phase trajectory, colored by the assigned behavior, in Fig. 5. We found that behaviors corresponded to very clear separation in the t-SNE embedding, with some overlapping

around the intersection of the five behaviors. The Silhouette coefficient of 0.5 indicates a good partitioning, i.e., a partitioning that is significantly different from random assignment (in which case the coefficient would be 0). We report the value counts per species in Table 1. This analysis showed that the species distribution was uneven, with a slight plurality of homing individuals, and Steppy being the rarest of all species among the GRNs that have memory.

Finally, we tested whether different circuits within the same network have different behaviors or consistently fall into one; in other words, we tested whether the same network can have multiple different "personalities" with respect to housing circuits that respond differently to training. We performed a chi-squared test for independence between the network identifier and its behavior label, which revealed ($p < 0.001$) that circuits belonging to the same network preferentially adopted one species of behavior.

Visualization of the five behaviors reveals relevant patterns

We then set out to study how the five types of behaviors differed. We plotted five sample trajectories for each behavior in Fig. 4, where the x-axis corresponds to simulated time. Each behavior represented trajectories sharing a distinct pattern, as if the GRN agents were adopting a specific behavior in the emergence space. Homing trajectories frequently oscillated around their mean, as if the GRN was "numb". Inflating individuals had an overall positive trend, as the name suggests, as if training made them more and more emergent, whereas the opposite was true for deflating individuals. The spiky behavior consisted of a few periodic, extreme bursts of emergence that left the trend unchanged in the long term. Finally, Steppy was about having a few prolonged (not short, like inflating) bursts of emergence as if they were strides.
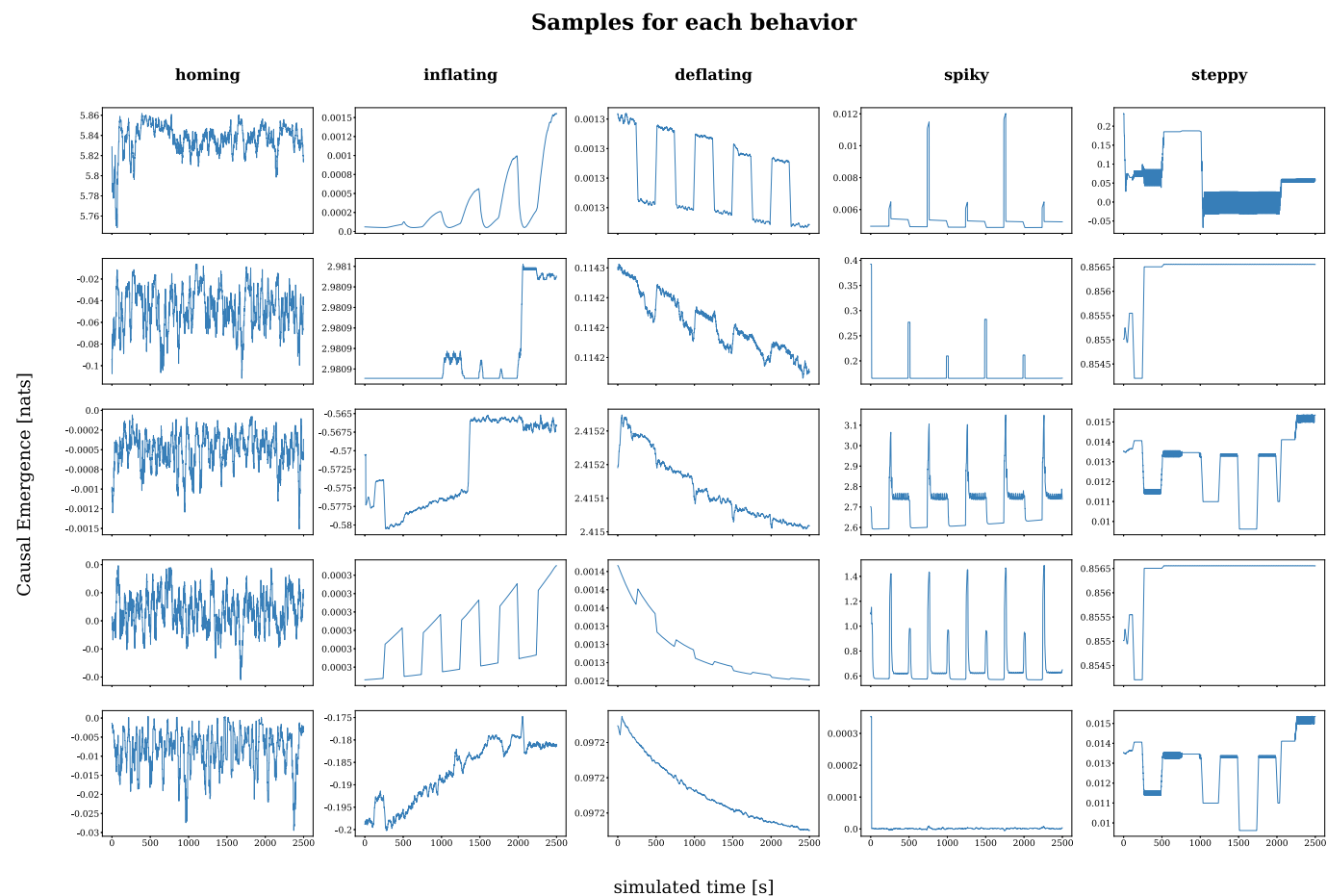
Fig. 5 | Effects of training on causal emergence occur in five distinct types. t-SNE embedding in 2D of the behavior descriptors for the test trajectories, as classified by the k-means unsupervised learning algorithm; each point corresponds to the test phase for one circuit of one network. The behaviors segregate well into distinct clusters.

Table 1 | Number of memory circuits per behavior class

Homing	378
Inflating	125
Deflating	114
Spiky	107
Steppy	6

Breakdown by behavior of the number of GRN circuits with memory. Behavior distribution is uneven, with a plurality of homing individuals and a rarity of steppy individuals.

The descriptor histograms in Fig. 6 illustrate the effects of training on emergence. They show, for example, how homing individuals were non-monotonous and flat; spiky individuals were also flat but had, on average, more distanced peaks and a larger range; finally, deflating individuals were more negatively monotonous, whereas inflating ones had positively monotonic effects on causal emergence.

Behaviors differ by phylogeny and gene ontology
Having seen that the integrated nature of biological networks displayed different responses to training, we wondered if these classes corresponded to distinct biology (phylogeny and gene ontology of the network): might networks belonging to different types of processes or species exhibit different responses with respect to how much they are reified by the associative conditioning? We extracted this information directly from the BioModels website and visualized the results in Fig. 7 with heatmaps. Figure 7A shows the relative occurrence (i.e., all the

behaviors sum to 1) of the five automatically classified behaviors by phylogeny (top row) and gene ontology (bottom row). Some cells are marked with "N/A" because not every behavior is represented in every phylogeny or gene ontology. From the tables, we saw that there existed a relationship between phylogeny (or gene ontology) and behavior occurrence. For example, lower vertebrates (which mostly include *Xenopus laevis*) have the highest diversity, followed by insects and plants, whereas mammals show the least. Slime molds broke down similarly to mammals, but it was hard to draw conclusions from that comparison because only one network fell under this taxon (from *Physarum polycephalum*), but we still included its results for completeness. When looking at gene ontology, the MAPK cascade and the mitotic cell cycle were the most diverse, and stem cell differentiation was the least. Similarly to slime molds, some gene ontologies (the far-red right signaling in *P. polycephalum* and the sucrose biosynthetic process in *Saccharum officinarum*) were represented by only one network, but we included their results for completeness, even though few inferences could be made.

A similar result was revealed analyzing the average (across the circuits) % change in causal emergence from before to after training in Fig. 7B, which includes the margins. Plants showed the greatest increase in emergence and insects the greatest decrease, while also having the highest behavior diversity (Fig. 7A). On the other axis, the sucrose biosynthetic process showed the greatest increase in causal emergence and the regulation of the circadian rhythm showed the greatest decrease. We performed a chi-squared test for independence in two-way tables for all the tables of Fig. 7 and found the results to be significant, confirming our hypothesis that the specific ways in which causal emergence is potentiated by learning correlate with specific phylogeny and gene ontology of the networks.

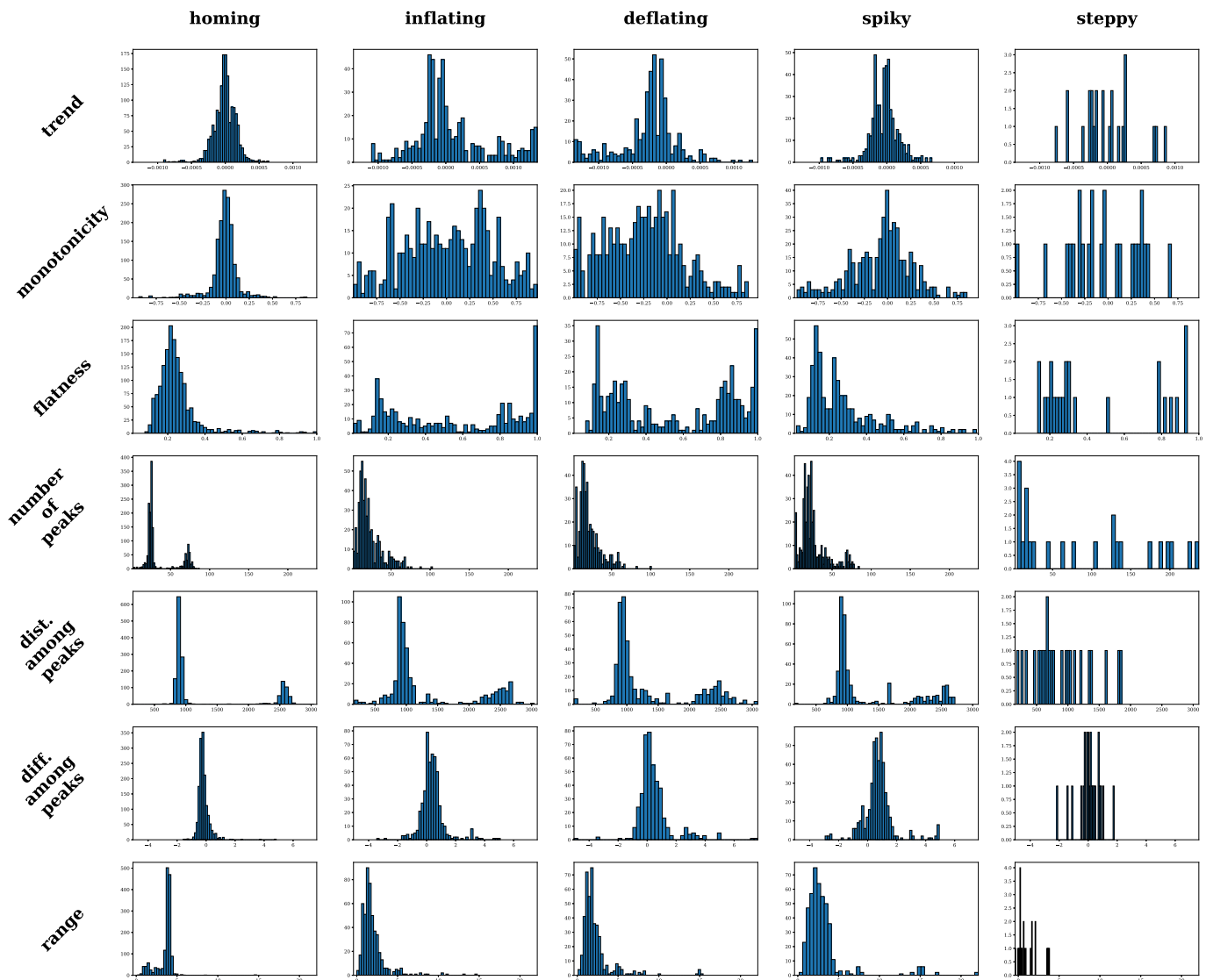## Descriptor distributions per behavior

Fig. 6 | Causal emergence types are characterized by different traits. Histograms for the seven descriptors (features of causal emergence during testing; one per row) by the five automatically discovered causal emergence types (*behaviors*; one per column). Behaviors correspond to different descriptors' distributions, meaning that our classification captures different manners in which causal emergence reacts to training.

Discussion

How could a system's supervenience over its parts be increased? Here we showed that re-engineering its hardware architecture (physical topology) is not required to accomplish this. Rather, by using an in silico model of a minimal agent, we found that training a network for associative memory can increase its integrated causal power. We also demonstrated another surprising result: the relationship between causal integration and learning can be bi-directional: not only is integration needed for a system to be capable of associating the experiences of its parts into associative learning in the collective, but it turns out that conversely, associative conditioning experiences can potentiate the collectivity and integration of networks.

We have shown that these results hold for a specific type of system, ODE GRNs, which has relevance to mechanisms necessary for physiological regulation and cognition in animals, both at the evolutionary and individual behavioral scale. Our work was motivated by two considerations. The first was aimed at the nascent field of diverse intelligence and unconventional cognition[75], seeking to understand the dynamics necessary and sufficient for the emergence of integrated selves on a scale from the most minimal matter

to human metacognition and beyond[76]. The second is a roadmap toward new ways to manipulate biological matter for biomedical and bioengineering applications, that moves beyond rewiring biochemical details toward programming, communicating with, and motivating living tissues toward desired system-level outcomes[6].

Network theory[77] and dynamical systems theory[78] provide metrics to analyze networks, including computational models of GRNs[79]. These existing tools allow the inference of biological pathways from data[80], as well as algorithms for predicting how they will respond to new inputs[81]. While very useful[32,35,82], advances are limited by the assumption that structure fully explains function[83,84]: a view of molecular pathways as mechanical machines inevitably focuses attention on the hardware and approaches to modify it, such as CRISPR, protein engineering, and the editing of promoters to create novel connections between them. However, network rewiring is hard and time-consuming, with many challenges for the synthetic biologist[85] or the designer of gene therapies. Recent research has shown that GRNs can demonstrate a variety of unanticipated behaviors, such as associative memory[86,87]—and that such behavior arises from changes in the *signaling* within a specific network rather than changes in its wiring (which has
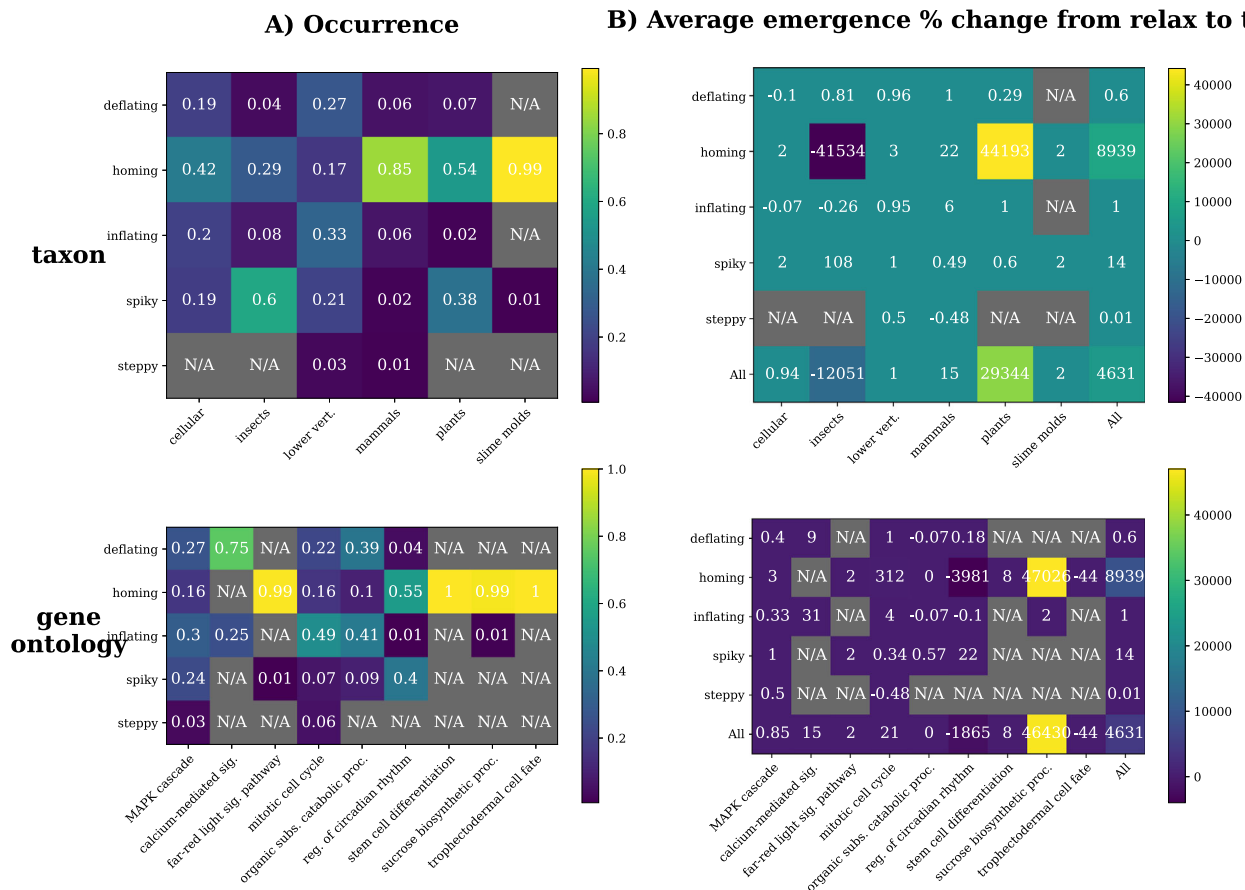
Fig. 7 | Relationship between behavior class and the biological nature of each network. A Two-way table of relative occurrences of behaviors for circuits (ways in which emergence changes upon training) for each taxon (top row) and gene ontology (bottom row); columns sum to one. **B** Two-way table of average % causal emergence change from before to after training for each taxon and gene ontology, including the margins (averages over rows/columns). Some cells are marked as "N/A" because not every taxon or gene ontology is represented for a given behavior. The occurrence of behavior is related to taxa and gene ontologies[87].

significant implications for the development of biomedical applications that use patterns of stimuli and do not rely on gene therapy[88]).

More broadly, the field of diverse intelligence research[17,18,89–92] seeks to understand cognition in unfamiliar guises and implementations. Beyond traditional studies of organisms with brains, it seeks to understand ways in which learning, decision-making, and different degrees of intelligent problem-solving can be implemented in a wide range of media. Especially important are ways in which evolutionary and engineering processes can scale up the basal cognitive capacity of minimal active matter[76] and lead to the emergence of new agents that are in crucial ways "more than the sum of their parts". Thus, in addition to the biomedical motivation for finding new ways to induce and manipulate memories in GRNs, we seek to use memory in gene regulatory networks as a model system in which to understand the origin of unconventional, minimal cognitive systems. By expanding concepts from neuroscience (such as measures of integration) to novel substrates[93], we hope to understand the relationship between learning and the causal structures that implement active agents composed of parts (i.e., all of us). In doing so, we found a surprising result, in which training reifies the causal potency of a distributed system.

Training a biological network for associative memory increases causal emergence by 128.32%, on average, meaning a factor of almost two-and-a-half. This does not happen for every network: some GRNs remain as they were natively, before their training, while others (the vast majority) adapt based on what they experience. Borrowing an analogy from circuit components, some GRNs behave like resistors (which function the same regardless of their history or the frequency of the incoming signal), while others behave like memristors[94] (which have a high degree of hysteresis and are also frequency-dependent). Interestingly, the specific way in which

causal emergence rises after training is similar within the various circuits of a given network, suggesting that a network has a consistent "personality" with respect to how stimulation of its various input nodes affects its integrated nature.

This increased causal emergence contrasts dramatically with what we observed in random networks, which have an average increase of only 56.25%. This result suggests that evolution may have selected biological networks to be more responsive to training in a way that is not reducible to random network dynamics. In other words, unlike some generic network properties (attractors, stability, etc.) that are found in even random networks[95], the ones we describe have been (directly or indirectly) reinforced by the processes of life[96]. There was no correlation between causal emergence and mere GRN activity, meaning that our findings cannot be derived from established metrics on networks. This result is consistent with studies on integrated information and psychedelics[97,98], which have shown that measurements of "how much activity" exists in the brain do not correlate well with the richness of corresponding conscious experience.

We automatically categorized causal emergence trajectories into behaviors and investigated how the behavior depends on the biology of the network, in particular, phylogeny and gene ontology. We found a dependence between the two, with different phylogenies and gene ontologies responding with different characteristic behaviors. However, we found no relationship between evolutionary history and behavior. Mammals and slime molds are, respectively, the most recent and one of the most ancient taxa in our study, yet they share low levels of diversity and relatively higher levels of increase in causal emergence. Plants (an ancient taxon[99]) and insects (a less ancient taxon[100]) have, respectively, the greatest positive and negative change of causal emergence. In the future, we will investigate what happens

when considering a wider repertoire of ontologies for each phylogeny. Crucially, there is also no relationship between "intelligence" when measured as the number of neurons (if any)[101] and the effects of training on integration. Mammals (which, in our study, mostly consist of human GRNs) are dominated by plants (that do not even have a nervous system[99]) in terms of causal emergence change, and slime molds (having the most primitive nervous systems[102]) dominate insects. Subsequent research may identify other biological parameters that map more tightly to the different classes of response that we found in these GRNs (but what we know already is that popular ways to categorize networks are not sufficient to capture the dynamics we observed).

The relationship between causal emergence and gene ontology also deserves mention. The MAPK cascade is not only the most diverse ontology here, but it also corresponds to an almost doubling of causal emergence after training. This finding is relevant considering that, in addition to regulating response to a wide array of stimuli and being found in most eukaryotes[103], this pathway pre-organizes pathway segments so that they respond faster and stronger to subsequent stimuli—that is, that they form new memories more readily[104]. MAPK is especially interesting given its central role in stress response and memory, which are very relevant as a kind of cognitive glue[14,105,106] that helps bind active subunits (such as cells and molecular networks) toward a common purpose in multicellular organisms navigating a range of problem spaces[16].

Similarly, the mitotic cell cycle (the second most diverse ontology) also plays a key role in the reproduction of every cell in the organism. On the other hand, the regulation of the circadian rhythm results in the greatest negative change in emergence after training, contrary to what happens with most ontologies. One possible reason is that the circadian GRNs are not persuadable, or, in other words, hold to their priors more strongly, but it is hard to draw conclusions in the absence of more experiments. In general, our intuitions are limited by the subset of networks analyzed and would be stronger if tested on a larger and more diverse pool of GRNs, consisting of, for example, phylogenies not considered so far in this work. It is a limitation of this area of inquiry that biologically accurate, fully parameterized network models are not plentiful.

There are several essential areas for future work. ODEs are a convenient formalism to study continuous-time GRNs[107], but it is possible they provide spurious behaviors (within certain parameter ranges) that may not map to observable phenotypes[108]. Also, ODEs model GRNs that operate in isolation and do not consider the biological noise and interactions coming from the intracellular matrix (future work will model the matrix as an environment for receiving and sending feedback to GRNs). Our tests were done in a highly simplified model, focusing on just one layer of biological control, specifically to show the minimal features sufficient to couple learning to increases of causal emergence. Next steps include the integration of this analysis with models of bioelectric[109–114] and biomechanical[115–117] aspects of cellular function, to see how these other layers compare with the biochemical one we analyzed here, with respect to the relationship of memories and integration.

It is, of course, essential to test these findings in real cells; technologies such as optogenetics and mesofluidics now exist that can provide the necessary level of temporal control[118]. Future work will also consider how causal emergence differs across other dimensions, such as the age of the host (i.e., GRNs in uteri, middle-aged adults, and elderly patients). An intriguing hypothesis is that cancer networks may react differently than tissue networks to training. Such insights would then inform biomedical control for the development of more efficacious drugs with fewer side effects. We have previously suggested that taking advantage of the decision-making, problem-solving, and other competencies of living material, such as by understanding how experiences affect its causality as an integrated whole distinct from the collection of parts, will lead to a much different therapeutic landscape[6,88].

Alternative approaches to model causality exist, especially in the field of free energy minimization[119]. However, while useful for inferring interactions from data, they either rely on generative models and priors (e.g., dynamic causal models) or do not inherently address causal emergence but mainly represent statistical structure (e.g., Bayesian networks), further justifying the ΦID as the preferred measure of causal emergence. Still, studies on the interlink between integrated information and the free energy principle have improved our understanding of the human mind[120] and virtual agents and environments[121], and future work will fully merge causal emergence in GRNs with free energy minimization.

Three other areas offer immediate opportunities for further investigation. First is evolution. Significant work has looked at the interplay between intelligence and evolvability[122–129]. Simulations and in vitro experiments could now look at the effects of learning-induced potentiation of integrated agency on the evolutionary process. We hypothesize that such experiments could reveal bootstrapping and feedback loop dynamics in which causal emergence improves learning ability, which in turn increases causal emergence, leading to an intelligence ratchet that could potentiate the emergence of agency in the world. Second, the study of the phylogenetic and ontogenetic origins of Selfhood and integrated minds may be enriched by a better understanding of relevant dynamics not tied to a specific neural basis. Third, an expanded survey of training modes (besides associative conditioning), and subject matter (pathways, and other networks in biological and technological systems) should be examined for these dynamics.

Furthermore, these data may suggest that GRN modification (i.e., almost all pharmacological interventions) may need to be part of the discussions that regulate the tradeoff between the needs of human patients and those of biologicals used for research. This includes both those used for basic science experiments, and those that produce therapeutics—animals making antibodies and other drugs, humanized pigs and other sources of heterologous transplantation materials. In other words, to whatever extent integration metrics may relate to the existence of an inner perspective in a system[57,63,130–132], the work described here provides an additional tool for discussions of ethics in the practice of biology and medicine.

Conclusions

Questions about the capabilities of living matter, and the applicability of tools from computational and cognitive sciences outside the neural domain, form the basis of an exciting emerging field that encompasses efforts to understand basal cognition, diverse intelligence, active matter, and unconventional computing[15,102,133–146]. We suspect it is likely that the ability to modulate integrated emergent agency by specific training experiences (that do not require physical rewiring) will have implications for not only evolutionary biology and philosophy of mind but also for engineering efforts involving biological, engineered, and hybrid agents.

Materials and methods
Biological models and simulation
We curated a dataset of 29 peer-reviewed biological network models from the BioModels open database[70] encompassing all the strata of life (from bacteria to humans), the same adopted in ref. 87. Each model describes a GRN whose nodes are proteins, metabolites, or genes (the species of a network) and whose edges are mutual reactions. Each network is modeled over time according to chemical rate law ODEs, and so each one is a continuous-time dynamical system. Species take values on a continuous domain (like protein concentrations or gene expression levels). We relied on the SBMLtoODEjax Python library[147], which parses System Biology Markup Language specification files into JAX programs. We then simulated each model in AutoDiscJAX (https://github.com/flowersteam/autodiscjax), which is not only written in the highly efficient and compute-optimized JAX language but also enables interventions on GRNs like applying stimuli. For all experiments, simulations were integrated with the fourth-order Runge-Kutta method using a step size of 0.01.

Random models
We built a set of 145 (5 different seeds per biological model) random networks by the gene circuit method[71], as in ref. 87. For each biological network, we created five different random networks by sampling—with

different random seeds—the parameters (i.e., connection strengths, initial concentrations, and constants) from a uniform distribution $U(0, 1)$, where we set the bounds to be consistent with the empirical distribution of parameters in the biological models. So, we created a total of 145 random networks. These models had the same distribution of structural properties, in particular topology and network theoretic properties (in-degree, out-degree, betweenness centrality, PageRank, HITS scores) as the biological models they are built from, thus representing a class of synthetic biological networks with structured properties and random weights.

Memory evaluation

Of the types of memory identified in ref. 87, here we focused on *associative* memory, since it most emphasizes the need for a network to integrate experiences across different nodes, and because it is the most interesting clinically (offering the possibility of associating powerful but toxic drugs with "placebo" triggers)[88]. Associative memory is analogous to Pavlovian training in animals; we present an illustration in Fig. 1A, with the corresponding training schedule for our ODE GRNs in Fig. 1B. Associative memory involves a triplet of nodes (a *circuit*): a target R, an UCS that regulates R, and a NS that does not. We first "relaxed" the GRN to allow it to settle on a steady state and have a baseline for its pre-training behavior (*relaxation* phase) by simulating it for t_s time steps without any stimuli. We then trained it by stimulating UCS and NS simultaneously (*training* phase) for another t_s time steps. If, finally, stimulation of NS only regulated (*testing* phase) for other t_s time steps, we said the network had "learned" to associate UCS with NS, turning NS into a conditioned stimulus CS. After preliminary experiments, we found $t_s = 250{,}000$ (2500 s of simulated time) to be sufficient for all the networks to settle on steady values.

For each network, we tested every possible triplet of nodes for associative memory. Since biological entities take on continuous values in ODE networks, we can either up-stimulate (increase the value to some extent) or down-stimulate (decrease the value up to some extent) for a specific species. Similarly, stimulation can up-regulate or down-regulate R, depending on whether stimulation increases or decreases its value. Following[87], we up-stimulated and down-stimulated by setting the quantity of a stimulus ST to $e_{\max}^{ST}\!\times\!100$ and $e_{\min}^{ST}/100$, respectively, where $e_{\max}^{ST}$ and $e_{\min}^{ST}$ were the maximum and the minimum values the species attained. We called R up-regulated if the mean value during testing was at least twice that during relaxation, and down-regulated if it was no more than half. In line with[87], we found these values for stimulation and regulation to result in associative learning in our network and simulate the delivery of real-world drugs.

Stimulation through the delivery of real-world drugs did not take place in one persistent bout, but in several time-delayed dispensations. For this reason, we applied stimulation in pulses: we partitioned the phase (whether training or testing) into five equally sized time intervals and alternated between applying the stimulus (i.e., at the first, third, and fifth intervals) and not applying it. When a stimulus was not applied, the species followed their intrinsic dynamics as dictated by the network; in the Results, we verify that this pulsed stimulation was not correlated with causal emergence and, thus, could not alone explain the increase in emergence after training.

Of all the triplets of the 29 biological networks considered, the 808 that passed the pretest belonged to 19 networks (more than half), in line with ref. 87. We considered these circuits for all the analyses and visualizations.

Data preprocessing

We applied the following preprocessing steps from ref. 148 to highlight underlying structures in the GRN simulation data and allow for more meaningful inferences. For each simulated ODE trajectory, we performed global signal regression by regressing out the mean at each time step across the species to filter out global artifacts; these could be of biological significance but, when computing information flows, we are interested in changes from the baseline and not global trends[149]. Second, we removed autocorrelation. Indeed, biological signals are known to be autocorrelated, which could inflate pairwise dependencies between time series by reducing

the effective degrees of freedom[150]. We followed the approach of [151] and performed the following steps independently for each species: we computed the linear least-squares regression between time $t-1$ and t, computed the predicted values at time given the regression results, and finally obtained the residuals as our preprocessed signal.

Information theory and partial information decomposition

Information theory, originally introduced to study the transmission capacity of communication channels, has over the years emerged as a principled language to evaluate dependencies in complex systems, including biological[152]. The basic object of study is Shannon's entropy:

$$H(X) = -\sum_{x} p(x)\ln p(x)$$

Where the summation is over the support of X, and it quantifies the amount of uncertainty about a random variable X. We can then define for a process consisting of a "source" variable X and a "target" variable Y the *mutual information* as the uncertainty that is left on Y after observing X, i.e., how much information observing X discloses about Y.

What if there is more than one source variable, as happens in complex systems like regulatory networks? For a finer-grained understanding of information, we must consider all the different ways information flows across a system. Intuitively, let us consider the case of stereoscopic vision in humans: with one eye open, we perceive a *unique* set of visual features for each, and also *redundant* features both of our eyes capture, but, finally, depth perception, which can only be captured if both eyes are open simultaneously, corresponds to *synergistic* information. The seminal work on Partial Information Decomposition (PID) provides a framework to parcel out mutual information into various information atoms (redundant, unique, and synergistic)[153].

Mutual information and PID, by themselves, are instantaneous measures of integration; they fail to capture the temporal and causal aspects of information dependencies up to and including all future time steps, a crucial aspect for dynamical systems that evolve over time[65], like our models of biological regulation.

Causal emergence and integrated information decomposition

We investigated how associative learning impacts the causal emergence of a system, meant as the ability of a whole to influence the future of its parts. Several operationalizations to quantify the integration of a system have blossomed, like the various Φ measures, with groundbreaking results in neuroscience[154]. Still, they are unidimensional and so tend to behave inconsistently[155], like the famed Tononi's Φ. We pursued a multidimensional decomposition and relied on the recent framework of Integrated Information Decomposition ΦID[64]; it is a finer decomposition of the PID and captures all informational dependencies of a system in space (macro- and micro-scale) and time (instantaneous up to and including all future time steps). According to general assumptions outlined in ref. 66, a system's capacity to display emergence depends on how much information the whole provides about the future evolution that cannot be inferred by any subset of parts. The ΦID formal apparatus then tells us that we can decompose emergence capacity as the sum of two terms:

1) *Downward causation*: the amount of information that the whole predicts about the future of the single components;
2) *Causal decoupling*: the amount of information that the whole predicts about the future of the whole.

This definition previously appeared to quantify the reduction in emergence capacity between healthy and brain-injured patients[69]; we chose it as our measure of *causal emergence* in biological systems since it considers all types of influences the whole can have on the future of a system. We present a schematic of how causal emergence is derived from GRN data in Fig. 8.
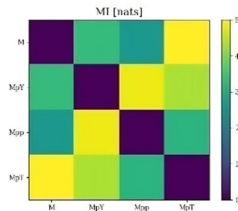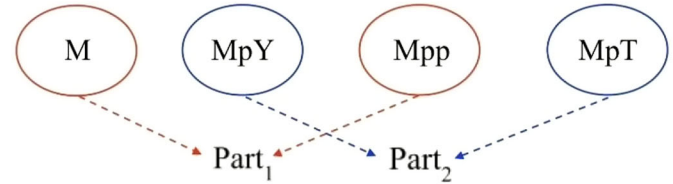
1) Simulate Gene Regulatory Network
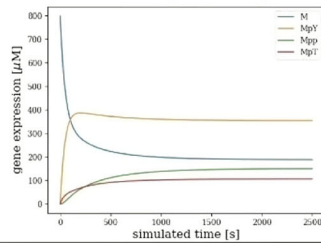
2) Preprocess gene expression series

a) compute z-score
b) regress-out global signal
c) remove autocorrelation

3) Compute Mutual Information matrix

4) Partition with Minimum Mutual Information

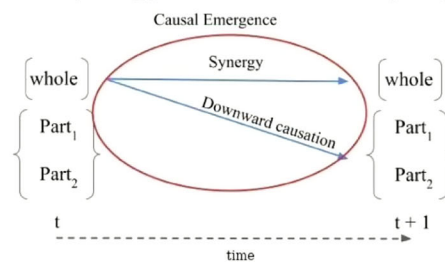**5) Compute Causal Emergence from synergy and causal decoupling**

Fig. 8 | From GRN data to causal emergence. This schematic shows our pipeline to arrive at the causal emergence values through the ΦID decomposition from GRN data. (1) We simulate a network to collect its gene expression profiles over time. (2) We preprocess the data with the techniques of[126]. (3) We compute the lag-1 mutual information matrix between the time series of every species in the network. (4) Using the results from the last step, we group the species into two partitions according to the minimum mutual information bipartition[137]. (5) Finally, we compute causal emergence as the sum of synergy (predictive power of the whole with respect to the whole) and downward causation (predictive power of the whole with respect to the individual parts) of the two partitions over time[69]; we remark that, as seen in this picture, causal emergence does not consider the predictive power of the individual parts with respect to the individual parts.

Gaussian information theory

Our models are continuous-valued biological dynamical systems, while information theory was originally defined for discrete random variables. We used the continuous generalization of Shannon's entropy, the differential entropy:

$$H(X) = \int_x p(x)\ln p(x)dx$$

Where the integral is over the support of X. This generalization is, in general, hard to compute because it requires estimating $p(x)$. But, if we assume that $p(x)$ is distributed as a Gaussian, we can leverage closed-form estimators for the entropy and, as a result, all the other information measures[156]. Indeed, the bivariate mutual information (in natural units) becomes:

$$I(X, Y) = -\frac{\ln(1 - \rho^2)}{2}$$

Where ρ is the Pearson correlation coefficient between X and Y. While the Gaussian hypothesis is limiting in the general case, for our specific data we verified, through the Shapiro-Wilk test, that our preprocessed data were significantly described by the Gaussian distribution.

Most practical computations of causal emergence converge on the same simple form for Gaussian continuous variables that we adopted here. We first computed the lag-1 mutual information matrix among all pairs of nodes in the system using the equation above[157]. Since we cannot handle systems with several elements because of the combinatorics involved[158], we reduced the dimensionality by the Minimum Information Bipartition (MIB)[159]. The MIB bisects the system into two components by approximating the bisection through the Fiedler vector (the eigenvector of the graph Laplacian corresponding to the smallest non-zero eigenvalue). After bisecting the graph with the Fiedler vector, we averaged within each component and compared the dynamics of the two parts to the whole. In essence, we sliced a watermelon along its longest axis and measured the extent to which the average number of seeds in one half predicted the average number of seeds in the other. Finally, we solved a linear system of equations relating the mutual information to the atoms the ΦID is decomposed in.

Behavior descriptors

We extracted seven descriptors from each causal emergence trajectory to be fed to an unsupervised learning pipeline for automatic classification into behaviors. The descriptors are:

1) *Trend*: the slope of the least squares fit of the trajectory. A positive slope indicates an increasing trend, while a negative slope indicates a decreasing trend.
2) *Monotonicity*: the Kendall's tau coefficient between the trajectory and the sequence of its time stamps. Kendall's tau is a standard statistic to measure ordinal association between two quantities, and in our case, it is the highest for a perfectly monotonically increasing trajectory, and the lowest for a perfectly monotonically decreasing one, with values

around zero corresponding to the trajectory fluctuating independently of the time axis.

3) *Flatness*: how flat the trajectory is and does not locally deviate from the mean. We divided the trajectory into consecutive intervals and approximated the trajectory with the mean over each interval. We computed flatness as the r-squared coefficient of this approximation: the higher the coefficient, the better fit are the local means, meaning the trajectory was locally flat (though jumps may have existed in correspondence with the interval boundaries). After preliminary experiments, we found an interval size of 100 to correctly capture the intuition behind the flatness of a trajectory.

4) *Number of peaks*: the number of local minima and maxima of the trajectory. To detect the maxima of a trajectory, we searched the time step list for values that are higher than the values neighboring them, and, to exclude weak maxima, filtered out those that were not equal to the maximum over a centered window of size 100 to correctly capture the intuition behind a peak. To detect the local minima, we repeated the same procedure but for values that are smaller.

5) *Average distance among peaks*: the average distance among all the peaks from 4) (or 0 if there were none).

6) *Average difference among peaks*: the average difference in causal emergence value of the peaks from 4) (or 0 if there were none).

7) *Range*: the difference in causal emergence value between the maximum and the minimum peaks.

Statistics and reproducibility

General information on what statistical tests were carried out is made explicit in the text whenever a test is mentioned. For all tests, we used 0.05 as the confidence level and reported the sample sizes in the description of the corresponding experiment. When random data were simulated, five different random seeds were used to seed the replicates.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The biological gene regulatory networks employed in this study are freely available as part of the BioModels project[70].

Code availability

We released all the code necessary to reproduce and replicate the experiments of this paper at https://github.com/pigozzif/IntegratedInformation GeneRegulation.

References

1. Juarrero, A. Top-down causation and autonomy in complex systems. In *Downward Causation and the Neurobiology of Free Will* (eds Murphy N., Ellis G. F. R., O'Connor T.) (Springer, 2009).
2. Ellis, G. F. R. Top-down causation and the human brain. In *Downward Causation and the Neurobiology of Free Will* (eds Murphy N., Ellis G. F. R., O'Connor T.) 63–81 (Springer, 2009).
3. Auletta, G., Ellis, G. F. & Jaeger, L. Top-down causation by information control: from a philosophical problem to a scientific research programme. *J. R. Soc. Interface* 5, 1159–1172 (2008).
4. Walker S., Cisneros L., Davies P. C. W. Evolutionary transitions and top-down causation. In *Proc. ALIFE 2012: The Thirteenth International Conference on the Synthesis and Simulation of Living Systems* (ASME, 2012).
5. Pezzulo, G. & Levin, M. Top-down models in biology: explanation and control of complex living systems above the molecular level. *J. R. Soc. Interface* 13, 20160555 (2016).
6. Lagasse, E. & Levin, M. Future medicine: from molecular pathways to the collective intelligence of the body. *Trends Mol. Med.* 29, 687–710 (2023).
7. McMillen, P. & Levin, M. Collective intelligence: a unifying concept for integrating biology across scales and substrates. *Commun. Biol.* 7, 378 (2024).
8. Couzin, I. Collective minds. *Nature* 445, 715 (2007).
9. Deisboeck, T. S. & Couzin, I. D. Collective behavior in cancer cell populations. *Bioessays* 31, 190–197 (2009).
10. Werfel J. Collective construction with robot swarms. In: *Morphogenetic Engineering* (eds Doursat R., Sayama H., Michel O.) (Springer, 2012).
11. Abdel-Rahman, A., Cameron, C., Jenett, B., Smith, M. & Gershenfeld, N. Self-replicating hierarchical modular robotic swarms. *Commun. Eng.* 1, 35 (2022).
12. Slavkov, I. et al. Morphogenesis in robot swarms. *Sci. Robot.* 3, eaau9178 (2018).
13. Strassmann, J. E. & Queller, D. C. The social organism: congresses, parties, and committees. *Evolution* 64, 605–616 (2010).
14. Levin, M. The computational boundary of a "self": developmental bioelectricity drives multicellularity and scale-free cognition. *Front. Psychol.* 10, 2688 (2019).
15. Levin, M. Technological approach to mind everywhere: an experimentally-grounded framework for understanding diverse bodies and minds. *Front. Syst. Neurosci.* 16, 768201 (2022).
16. Fields, C. & Levin, M. Competency in navigating arbitrary spaces as an invariant for analyzing cognition in diverse embodiments. *Entropy* 24, 819 (2022).
17. Baluška, F. & Levin, M. On having no head: cognition throughout biological systems. *Front. Psychol.* 7, 902 (2016).
18. Lyon, P. The cognitive cell: bacterial behavior reconsidered. *Front. Microbiol.* 6, 264 (2015).
19. Eckert, L. et al. Biochemically plausible models of habituation for single-cell learning. *Curr. Biol.* 34, 5646–5658 e5643 (2024).
20. Marshall, W., Kim, H., Walker, S. I., Tononi, G. & Albantakis, L. How causal analysis can reveal autonomy in models of biological systems. *Philos. Trans. A* 375, 20160358 (2017).
21. Juel, B. E., Comolatti, R., Tononi, G. & Albantakis, L. When is an action caused from within? Quantifying the causal chain leading to actions in simulated agents. *ALIFE 2019: The 2019 Conference on Artificial Life* (2019).
22. Mayner, W. G. P. et al. PyPhi: a toolbox for integrated information theory. *PLoS Comput. Biol.* 14, e1006343 (2018).
23. Marshall, W., Albantakis, L. & Tononi, G. Black-boxing and cause-effect power. *PLoS Comput. Biol.* 14, e1006114 (2018).
24. Dewan, E. M. Consciousness as an emergent causal agent in the context of control system theory. (eds Globus G., Maxwell G., Savodnik I.) (Plenum Press, 1976).
25. Davidson, E. H. et al. A genomic regulatory network for development. *Science* 295, 1669–1678 (2002).
26. Singh, A. J., Ramsey, S. A., Filtz, T. M. & Kioussi, C. Differential gene regulatory networks in development and disease. *Cell Mol. Life Sci.* 75, 1013–1025 (2018).
27. Gyurkó, D. M. et al. Adaptation and learning of molecular networks as a description of cancer development at the systems-level: potential use in anti-cancer therapies. *Semin. Cancer Biol.* 23, 262–269 (2013).
28. Csermely, P., Korcsmáros, T., Kiss, H. J., London, G. & Nussinov, R. Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. *Pharm. Ther.* 138, 333–408 (2013).
29. Ten Tusscher, K. H. & Hogeweg, P. Evolution of networks for body plan patterning; interplay of modularity, robustness and evolvability. *PLoS Comput. Biol.* 7, e1002208 (2011).
30. Peter, I. S. & Davidson, E. H. Evolution of gene regulatory networks controlling body plan development. *Cell* 144, 970–985 (2011).

31. Uller, T., Moczek, A. P., Watson, R. A., Brakefield, P. M. & Laland, K. N. Developmental bias and evolution: a regulatory network perspective. *Genetics* **209**, 949–966 (2018).

32. Velazquez, J. J., Su, E., Cahan, P. & Ebrahimkhani, M. R. Programming morphogenesis through systems and synthetic biology. *Trends Biotechnol.* **36**, 415–429 (2018).

33. Brophy, J. A. & Voigt, C. A. Principles of genetic circuit design. *Nat. Methods* **11**, 508–520 (2014).

34. Saltepe, B., Bozkurt, E. U., Güngen, M. A., Çiçek, A. E. & Şeker, U. O. S. Genetic circuits combined with machine learning provides fast responding living sensors. *Biosens. Bioelectron.* **178**, 113028 (2021).

35. Krzysztoń, R., Wan, Y., Petreczky, J. & Balázsi, G. Gene-circuit therapy on the horizon: synthetic biology tools for engineered therapeutics. *Acta Biochim. Pol.* **68**, 377–383 (2021).

36. Peng, W., Song, R. & Acar, M. Noise reduction facilitated by dosage compensation in gene networks. *Nat. Commun.* **7**, 12959 (2016).

37. Peng, W., Liu, P., Xue, Y. & Acar, M. Evolution of gene network activity by tuning the strength of negative-feedback regulation. *Nat. Commun.* **6**, 6226 (2015).

38. Guye, P., Li, Y., Wroblewska, L., Duportet, X. & Weiss, R. Rapid, modular and reliable construction of complex mammalian gene circuits. *Nucleic Acids Res.* **41**, e156 (2013).

39. Slusarczyk, A. L., Lin, A. & Weiss, R. Foundations for the design and implementation of synthetic genetic circuits. *Nat. Rev. Genet.* **13**, 406–420 (2012).

40. Beal, J., Lu, T. & Weiss, R. Automatic compilation from high-level biologically-oriented programming language to genetic regulatory networks. *PLoS ONE* **6**, e22490 (2011).

41. Kim, H. & Sayama, H. How criticality of gene regulatory networks affects the resulting morphogenesis under genetic perturbations. *Artif. Life* **24**, 85–105 (2018).

42. Solé, R. V., Fernández, P. & Kauffman, S. A. Adaptive walks in a gene network model of morphogenesis: insights into the Cambrian explosion. *Int. J. Dev. Biol.* **47**, 685–693 (2003).

43. Villani M., D'Addese G., Kauffman S. A., Serra R. Attractor-specific and common expression values in random boolean network models (with a preliminary look at single-cell data). *Entropy* **24**, 311(2022).

44. Beggs, J. M. The criticality hypothesis: how local cortical networks might optimize information processing. *Philos. Trans. A***366**, 329–343 (2008).

45. Graudenzi, A. et al. Dynamical properties of a boolean model of gene regulatory network with memory. *J. Comput. Biol.* **18**, 1291–1303 (2011).

46. Graudenzi, A., Serra, R., Villani, M., Colacci, A. & Kauffman, S. A. Robustness analysis of a Boolean model of gene regulatory network with memory. *J. Comput. Biol.* **18**, 559–577 (2011).

47. Serra, R., Villani, M., Barbieri, A., Kauffman, S. A. & Colacci, A. On the dynamics of random boolean networks subject to noise: attractors, ergodic sets and cell types. *J. Theor. Biol.* **265**, 185–193 (2010).

48. Damiani, C., Kauffman, S. A., Serra, R., Villani, M. & Colacci, A. Information transfer among coupled random boolean networks. *Lect. Notes Comput. Sci.* **6350**, 1 (2010).

49. Csermely, P. et al. Learning of signaling networks: molecular mechanisms. *Trends Biochem. Sci.* **45**, 284–294 (2020).

50. Perez-Lopez, A. R. et al. Targets of drugs are generally, and targets of drugs having side effects are specifically good spreaders of human interactome perturbations. *Sci. Rep.* **5**, 10182 (2015).

51. Kovács, I. A., Mizsei, R. & Csermely, P. A unified data representation theory for network visualization, ordering and coarse-graining. *Sci. Rep.* **5**, 13786 (2015).

52. Gyurkó, M. D., Steták, A., Sőti, C. & Csermely, P. Multitarget network strategies to influence memory and forgetting: the Ras/MAPK

53. Csermely, P. et al. Cancer stem cells display extremely large evolvability: alternating plastic and rigid networks as a potential Mechanism: network models, novel therapeutic target strategies, and the contributions of hypoxia, inflammation and cellular senescence. *Semin. Cancer Biol.* **30**, 42–51 (2015).

54. Schreier, H. I., Soen, Y. & Brenner, N. Exploratory adaptation in large random networks. *Nat. Commun.* **8**, 14826 (2017).

55. Soen, Y., Knafo, M. & Elgart, M. A principle of organization which facilitates broad Lamarckian-like adaptations by improvisation. *Biol. Direct.* **10**, 68 (2015).

56. Hoel, E. & Levin, M. Emergence of informative higher scales in biological systems: a computational toolkit for optimal prediction and control. *Commun. Integr. Biol.* **13**, 108–118 (2020).

57. Tononi, G., Boly, M., Massimini, M. & Koch, C. Integrated information theory: from consciousness to its physical substrate. *Nat. Rev. Neurosci.* **17**, 450–461 (2016).

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

59. Hoel, E. P., Albantakis, L., Marshall, W. & Tononi, G. Can the macro beat the micro? Integrated information across spatiotemporal scales. *Neurosci. Conscious* **2016**, niw012 (2016).

60. Rajpal, H. et al. Quantifying Hierarchical Selection. Preprint at https://arxiv.org/abs/2310.20386 (2023).

61. Cang, Z. & Nie, Q. Inferring spatial and signaling relationships between cells from single cell transcriptomic data. *Nat. Commun.* **11**, 2084 (2020).

62. Varley, T. F., Pope, M., Faskowitz, J. & Sporns, O. Multivariate information theory uncovers synergistic subsystems of the human cerebral cortex. *Commun. Biol.* **6**, 451 (2023).

63. Oizumi, M., Albantakis, L. & Tononi, G. From the phenomenology to the mechanisms of consciousness: Integrated Information Theory 3.0. *PLoS Comput. Biol.* **10**, e1003588 (2014).

64. Mediano, P. A. M., Rosas, F. E., Carhart-Harris, R. L., Seth, A. K. & Barrett, A. B. Beyond integrated information: a taxonomy of information dynamics phenomena. Preprint at https://arxiv.org/abs/1909.02297 (2019).

65. Mediano, P. A. M. et al. Greater than the parts: a review of the information decomposition approach to causal emergence. *Philos. Trans. A***380**, 20210246 (2022).

66. Rosas, F. E. et al. Reconciling emergences: an information-theoretic approach to identify causal emergence in multivariate data. *PLoS Comput. Biol.* **16**, e1008289 (2020).

67. Luppi, A. I. et al. A synergistic core for human brain evolution and cognition. *Nat. Neurosci.* **25**, 771–782 (2022).

68. Luppi, A. I. et al. A synergistic workspace for human consciousness revealed by Integrated Information Decomposition. *Elife* **12**, RP88173 (2024).

69. Luppi, A. I. et al. Reduced emergent character of neural dynamics in patients with a disrupted connectome. *Neuroimage* **269**, 119926 (2023).

70. Malik-Sheriff, R. S. et al. BioModels-15 years of sharing computational models in life science. *Nucleic Acids Res.* **48**, D407–D415 (2020).

71. Reinitz, J. & Sharp, D. H. Mechanism of eve stripe formation. *Mech. Dev.* **49**, 133–158 (1995).

72. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).

73. Hartigan, J. A. & Wong, M. A. Algorithm AS 136: a K-means clustering algorithm. *Appl. Stat.* **28**, 100–108 (1979).

74. van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).

75. Clawson, W. P. & Levin, M. Endless forms most beautiful 2.0: teleonomy and the bioengineering of chimaeric and synthetic organisms. *Biol. J. Linn. Soc.* **139**, 457–486 (2023).

76. Rosenblueth, A., Wiener, N. & Bigelow, J. Behavior, purpose, and teleology. *Philos. Sci.* **10**, 18–24 (1943).

77. Latora, V., Nicosia, V. & Russo, G. *Complex Methods: Principles, Methods and Applications* (Cambridge University Press, 2017).

78. Brin M., Stuck G. *Introduction to Dynamical Systems* (Cambridge University Press, 2002).

79. de Jong, H. Modeling and simulation of genetic regulatory systems: a literature review. *J. Comput. Biol.* **9**, 67–103 (2002).

80. Kim, D. et al. Gene regulatory network reconstruction: harnessing the power of single-cell multi-omic data. *NPJ Syst. Biol. Appl.* **9**, 51 (2023).

81. Schlitt, T. & Brazma, A. Current approaches to gene regulatory network modelling. *BMC Bioinform.* **8**, S9 (2007).

82. Ho, C. & Morsut, L. Novel synthetic biology approaches for developmental systems. *Stem Cell Rep.* **16**, 1051–1064 (2021).

83. Liu, Y. Y., Slotine, J. J. & Barabási, A. L. Controllability of complex networks. *Nature* **473**, 167–173 (2011).

84. Zañudo, J. G. T., Yang, G. & Albert, R. Structure-based control of complex networks with nonlinear dynamics. *Proc. Natl. Acad. Sci. USA* **114**, 7234–7239 (2017).

85. Gates, A. J. & Rocha, L. M. Control of complex networks requires both structure and dynamics. *Sci. Rep.* **6**, 24456 (2016).

86. Biswas, S., Manicka, S., Hoel, E. & Levin, M. Gene regulatory networks exhibit several kinds of memory: quantification of memory in biological and random transcriptional networks. *iScience* **24**, 102131 (2021).

87. Biswas, S., Clawson, W. & Levin, M. Learning in transcriptional network models: computational discovery of pathway-level memory and effective interventions. *Int. J. Mol. Sci.* **24**, 285 (2022).

88. Mathews, J., Chang, A. J., Devlin, L. & Levin, M. Cellular signaling pathways as plastic, proto-cognitive systems: Implications for biomedicine. *Patterns* **4**, 100737 (2023).

89. Baluška, F., Miller, W. B. & Reber, A. S. Cellular and evolutionary perspectives on organismal cognition: from unicellular to multicellular organisms. *Biol. J. Linn. Soc.* **139**, 503–513 (2023).

90. Reber, A. S. & Baluška, F. Cognition in some surprising places. *Biochem. Biophys. Res. Commun.* **564**, 150–157 (2021).

91. Lyon, P. Of what is "minimal cognition" the half-baked version? *Adapt. Behav.* **28**, 407–424 (2019).

92. Lyon, P. The biogenic approach to cognition. *Cogn. Process* **7**, 11–29 (2006).

93. Abramson, C. I. & Levin, M. Behaviorist approaches to investigating memory and learning: a primer for synthetic biology and bioengineering. *Commun. Integr. Biol.* **14**, 230–247 (2021).

94. Volkov, A. G., Nyasani, E. K., Blockmon, A. L. & Volkova, M. I. Memristors: memory elements in potato tubers. *Plant Signal Behav.* **10**, e1071750 (2015).

95. Kauffman S. A. *The Origins of Order: Self Organization and Selection in Evolution* (Oxford University Press, 1993).

96. Varley, T. F. & Bongard, J. Evolving higher-order synergies reveals a trade-off between stability and information-integration capacity in complex systems. *Chaos* **34**, 063127 (2024).

97. Gallimore, A. R. Restructuring consciousness -the psychedelic state in light of integrated information theory. *Front. Hum. Neurosci.* **9**, 346 (2015).

98. Bayne, T. & Carter, O. Dimensions of consciousness and the psychedelic state. *Neurosci. Conscious.* **2018**, niy008 (2018).

99. *Evolution and Diversification of Land Plants* (Springer 1997).

100. Gullan, P. J. & Cranston, P. S. *The Insects: an Outline of Entomology* 5th edn (Wiley Blackwell, 2014).

101. Eryomin, A. L. Biophysics of evolution of intellectual systems. *Biophysics* **67**, 320–326 (2022).

102. Vallverdú, J. et al. Slime mould: the fundamental mechanisms of biological cognition. *Biosystems* **165**, 57–70 (2018).

103. Pearson, G. et al. Mitogen-activated protein (MAP) kinase pathways: regulation and physiological functions. *Endocr. Rev.* **22**, 153–183 (2001).

104. Veres, T. et al. Cellular forgetting, desensitisation, stress and ageing in signalling networks. When do cells refuse to learn more?. *Cell Mol. Life Sci.* **81**, 97 (2024).

105. Shreesha, L. & Levin, M. Stress sharing as cognitive glue for collective intelligences: a computational model of stress as a coordinator for morphogenesis. *Biochem. Biophys. Res. Commun.* **731**, 150396 (2024).

106. Jiang, L., Chen, Y., Luo, L. & Peck, S. C. Central roles and regulatory mechanisms of dual-specificity MAPK phosphatases in developmental and stress signaling. *Front. Plant Sci.* **9**, 1697 (2018).

107. Barbuti, R., Gori, R., Milazzo, P. & Nasti, L. A survey of gene regulatory networks modelling methods: from differential equations, to Boolean and qualitative bioinspired models. *J. Membr. Comput.* **2**, 207–226 (2020).

108. Etcheverry, M., Moulin-Frier, C., Oudeyer, P.-Y. & Levin, M. AI-driven automated discovery tools reveal diverse behavioral competencies of biological networks. *Elife* **13**, RP92683 (2024).

109. Pietak, A. & Levin, M. Bioelectric gene and reaction networks: computational modelling of genetic, biochemical and bioelectrical dynamics in pattern regulation. *J. R. Soc. Interface* **14**, 20170425 (2017).

110. Pietak, A. & Levin, M. Exploring instructive physiological signaling with the bioelectric tissue simulation engine. *Front. Bioeng. Biotechnol.* **4**, 55 (2016).

111. Riol, A., Cervera, J., Levin, M. & Mafé, S. Cell systems bioelectricity: how different intercellular gap junctions could regionalize a multicellular aggregate. *Cancers* **13**, 5300 (2021).

112. Cervera, J., Levin, M. & Mafé, S. Bioelectrical coupling of single-cell states in multicellular systems. *J. Phys. Chem. Lett.* **11**, 3234–3241 (2020).

113. Cervera, J., Pai, V. P., Levin, M. & Mafé, S. From non-excitable single-cell to multicellular bioelectrical states supported by ion channels and gap junction proteins: Electrical potentials as distributed controllers. *Prog. Biophys. Mol. Biol.* **149**, 39–53 (2019).

114. Cervera, J., Manzanares, J. A., Mafé, S. & Levin, M. Synchronization of bioelectric oscillations in networks of nonexcitable cells: from single-cell to multicellular states. *J. Phys. Chem. B* **123**, 3924–3934 (2019).

115. Vanslambrouck, M. et al. Image-based force inference by biomechanical simulation. *PLoS Comput. Biol.* **20**, e1012629 (2024).

116. Delile, J., Herrmann, M., Peyriéras, N. & Doursat, R. A cell-based computational model of early embryogenesis coupling mechanical behaviour and gene regulation. *Nat. Commun.* **8**, 13929 (2017).

117. Beloussov, L. V. & Grabovsky, V. I. A Geometro-mechanical model for pulsatile morphogenesis. *Comput. Methods Biomech. Biomed. Eng.* **6**, 53–63 (2003).

118. Bugaj, L. J., O'Donoghue, G. P. & Lim, W. A. Interrogating cellular perception and decision making with optogenetic tools. *J. Cell Biol.* **216**, 25–28 (2017).

119. Friston, K. J. A free energy principle for biological systems. *Entropy* **14**, 2100–2121 (2012).

120. Safron, A. An integrated world modeling theory (IWMT) of consciousness: combining integrated information and global neuronal workspace theories with the free energy principle and active inference framework; toward solving the hard problem and characterizing agentic causation. *Front. Artif. Intell.* **3**, 30 (2020).

121. Lundbak Olesen, C., Waade, P. T., Albantakis, L. & Mathys, C. Phi fluctuates with surprisal: an empirical pre-study for the synthesis of the free energy principle and integrated information theory. *PLoS Comput. Biol.* **19**, e1011346 (2023).

122. van Duijn, M. Phylogenetic origins of biological cognition: convergent patterns in the early evolution of learning. *Interface Focus* **7**, 20160158 (2017).

123. Watson, R. A. & Szathmáry, E. How can evolution learn? *Trends Ecol. Evol.* **31**, 147–157 (2016).

124. Watson, R. A. et al. Evolutionary connectionism: algorithmic principles underlying the evolution of biological organisation in evo-devo, evo-eco and evolutionary transitions. *Evol. Biol.* **43**, 553–581 (2016).

125. Sznajder, B., Sabelis, M. W. & Egas, M. How adaptive learning affects evolution: reviewing theory on the Baldwin effect. *Evol. Biol.* **39**, 301–310 (2012).

126. Shreesha L., Levin M. Cellular competency during development alters evolutionary dynamics in an artificial embryogeny model. *Entropy* **25**, 131 (2023).

127. Pio-Lopez, L., Bischof, J., LaPalme, J. V. & Levin, M. The scaling of goals from cellular to anatomical homeostasis: an evolutionary simulation, experiment and analysis. *Interface Focus* **13**, 20220072 (2023).

128. Jablonka, E. & Ginsburg, S. Learning and the evolution of conscious agents. *Biosemiotics* **15**, 401–437 (2022).

129. Hinton, G. E. & Nowlan, J. How learning can guide evolution. *Complex Syst.* **1**, 495–502 (1987).

130. Bayne, T. On the axiomatic foundations of the integrated information theory of consciousness. *Neurosci. Conscious.* **2018**, niy007 (2018).

131. Tononi, G. Integrated information theory of consciousness: an updated account. *Arch. Ital. Biol.* **150**, 56–90 (2012).

132. Tononi, G. Consciousness as integrated information: a provisional manifesto. *Biol. Bull.* **215**, 216–242 (2008).

133. Baluška, F., Reber, A. S. & Miller, W. B. Jr Cellular sentience as the primary source of biological order and evolution. *Biosystems* **218**, 104694 (2022).

134. Tan, T. H. et al. Odd dynamics of living chiral crystals. *Nature* **607**, 287–293 (2022).

135. Zampetaki A. V., Liebchen B., Ivlev A. V., Löwen H. Collective self-optimization of communicating active particles. *Proc. Natl. Acad. Sci. USA* **118**, e2111142118 (2021).

136. Ozkan-Aydin, Y., Goldman, D. I. & Bhamla, M. S. Collective dynamics in entangled worm and robot blobs. *Proc. Natl. Acad. Sci. USA* **118**, e2010542118 (2021).

137. Stern, M., Pinson, M. B. & Murugan, A. Continual learning of multiple memories in mechanical networks. *Phys. Rev. X* **10**, 031044 (2020).

138. McGivern, P. Active materials: minimal models of cognition?. *Adapt. Behav.* **28**, 441–451 (2019).

139. Bernheim-Groswasser, A., Gov, N. S., Safran, S. A. & Tzlil, S. Living matter: mesoscopic active materials. *Adv. Mater.* **30**, e1707028 (2018).

140. Kaspar, C., Ravoo, B. J., van der Wiel, W. G., Wegner, S. V. & Pernice, W. H. P. The rise of intelligent matter. *Nature* **594**, 345–355 (2021).

141. Adamatzky, A., Chiolerio, A. & Szaciłowski, K. Liquid metal droplet solves maze. *Soft Matter* **16**, 1455–1462 (2020).

142. Safonov, A. A. Computing via natural erosion of sandstone. *Int. J. Parallel, Emerg. Distrib. Syst.* **33**, 742–751 (2018).

143. Mayne, R., Whiting, J. & Adamatzky, A. Toxicity and applications of internalised magnetite nanoparticles within live paramecium caudatum cells. *Bionanoscience* **8**, 90–94 (2018).

144. Adamatzky, A. Towards fungal computer. *Interface Focus* **8**, 20180029 (2018).

145. Čejková, J., Banno, T., Hanczyc, M. M. & Štěpánek, F. Droplets as liquid robots. *Artif. Life* **23**, 528–549 (2017).

146. Katz, E. Biocomputing—tools, aims, perspectives. *Curr. Opin. Biotechnol.* **34**, 202–208 (2015).

147. Etcheverry, M., Levin, M., Moulin-Frier, C. & Oudeyer P.-Y. SBMLtoODEjax: efficient simulation and optimization of biological network models in JAX. *NeurIPS 2023 AI for Science Workshop* (2023).

148. Blackiston, D. et al. Revealing non-trivial information structures in aneural biological tissues via functional connectivity. *PLOS Comput. Biol.* **21**, e1012149 (2024).

149. Liu, T. T., Nalci, A. & Falahpour, M. The global signal in fMRI: nuisance or information? *Neuroimage* **150**, 213–229 (2017).

150. Afyouni, S., Smith, S. M. & Nichols, T. E. Effective degrees of freedom of the Pearson's correlation coefficient under autocorrelation. *Neuroimage* **199**, 609–625 (2019).

151. Daube, C., Gross, J. & Ince, R. A. A. A whitening approach for transfer entropy permits the application to narrow-band signals. Preprint at https://arxiv.org/abs/2201.02461 (2022).

152. McMillen P., Walker S. I. & Levin M. Information theory as an experimental tool for integrating disparate biophysical signaling modules. *Int. J. Mol. Sci.* **23**, 9580 (2022).

153. Williams, P. L. & Beer, R. D. Nonnegative decomposition of multivariate information. Preprint at https://arxiv.org/abs/1004.2515 (2010).

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

155. Rosas, F. E., Mediano, P. A. M., Gastpar, M. & Jensen, H. J. Quantifying high-order interdependencies via multivariate extensions of the mutual information. *Phys. Rev. E* **100**, 032305 (2019).

156. Barrett, A. B. Exploration of synergistic and redundant information sharing in static and dynamical Gaussian systems. *Phys. Rev. E* **91**, 052802 (2015).

157. Friston, K. J. Functional and effective connectivity: a review. *Brain Connect* **1**, 13–36 (2011).

158. Jansma, A., Mediano, P. A. M. & Rosas, F. E. The Fast Möbius Transform: an algebraic approach to information decomposition. Preprint at https://arxiv.org/abs/2410.06224 (2024).

159. Kitazono, J., Kanai, R. & Oizumi, M. Efficient algorithms for searching the minimum information partition in integrated information theory. *Entropy* **20**, 173 (2018).

Author contributions

M.L. and F.P.—conceptualization of project goals, experimental design, data analysis and interpretation, manuscript preparation and editing. F.P.—coding, performing in silico experiments, data gathering. A.G.—interpretation, writing—review & editing. M.L.—funding acquisition.

Competing interests

M.L. receives funding in the form of a sponsored research agreement to Tufts University from Astonishing Labs, a company that seeks to advance biomedicine via tools that exploit the basal cognition of molecular networks. All other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-025-08411-2.

Correspondence and requests for materials should be addressed to Michael Levin.

Peer review information *Communications Biology* thanks Schahram Dustdar, Gordana Dodig-Crnkovic and Robert Prentner for their contribution to the peer review of this work. Primary Handling Editors: Aylin Bircan.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.