Article

Patterns

Evolving reservoir computers reveal bidirectional coupling between predictive power and emergent dynamics

Highlights

- Optimizing for predictive accuracy increases emergence, and vice versa
- Emergence tends to be highly sufficient and necessary for accurate prediction
- Optimizing for emergence in one task environment can enhance performance in others

Authors

Hanna M. Tolle, Andrea I. Luppi, Anil K. Seth, Pedro A.M. Mediano

Correspondence

h.tolle23@imperial.ac.uk

In brief

Biological and artificial neural networks perform complex computations that single neurons cannot. How this collective ability arises remains unclear. Here, the authors study this using bio-inspired reservoir computers and quantify emergence—when a system is more than the sum of its parts. The results reveal a direct, bidirectional link between a network's predictive power and its emergent dynamics, suggesting emergence as a core mechanism for computation.

Tolle et al., 2026, Patterns 7, 101457
March 13, 2026
https://doi.org/10.1016/j.patter.2025.101457

CellPress

Patterns

CellPress
OPEN ACCESS

Article

Evolving reservoir computers reveal bidirectional coupling between predictive power and emergent dynamics

Hanna M. Tolle,[1,8,*] Andrea I. Luppi,[2,3,4] Anil K. Seth,[5] and Pedro A.M. Mediano[1,6,7]
[1]Department of Computing, Imperial College London, London SW7 2AZ, UK
[2]Department of Psychiatry, Centre for Eudaimonia and Human Flourishing, University of Oxford, Oxford OX3 7JX, UK
[3]St. John's College, University of Cambridge, Cambridge CB2 1TP, UK
[4]Division of Information Engineering, University of Cambridge, Cambridge CB2 1PZ, UK
[5]Department of Informatics, Sussex AI and Sussex Centre for Consciousness Science, University of Sussex, Brighton BN1 9QJ, UK
[6]Division of Psychology and Language Sciences, University College London, London WC1H 0AP, UK
[7]Senior author
[8]Lead contact
*Correspondence: h.tolle23@imperial.ac.uk
https://doi.org/10.1016/j.patter.2025.101457

THE BIGGER PICTURE A long-held tenet of neuroscience posits that intelligence is a collective phenomenon: the brain is more than a mere collection of individual cells because it performs complex predictions that no single neuron can achieve alone. This aligns with the concept of "emergence," describing systems that exhibit capabilities absent in their constituent parts. While emergence has historically remained a philosophical notion, recent advances provide formal frameworks, grounded in information theory, to quantify it in data. These tools allow us to mathematically capture when a network's computation generates information that is irreducible to its individual components. Our study leverages these frameworks to determine the extent to which emergent dynamics drive prediction in neural networks. By modeling environmental prediction tasks in bio-inspired networks, we found that performance is tightly coupled with—and often dependent upon—the capacity for emergence. These findings support the hypothesis that emergence plays a functional role in prediction. Crucially, this underscores the utility of modern information-theoretic tools in transforming the view of the brain as a "collective" system from an illustrative idea into a rigorous, analytical approach.

SUMMARY

Biological neural networks perform complex computations to predict their environment, far exceeding the capabilities of individual neurons. Here, we argue that understanding these computations requires considering *emergent* dynamics—dynamics that make the whole system "more than the sum of its parts." We examine the relationship between prediction performance and emergence by leveraging quantitative metrics of emergence and modeling environmental time-series prediction within a bio-inspired computational framework called reservoir computing. Notably, three key results reveal a robust bidirectional coupling between prediction performance and emergence: (1) optimizing hyperparameters for performance enhances emergent dynamics, and vice versa; (2) emergent dynamics serve as a highly sufficient and often also necessary condition for prediction success in most environments; and (3) training with larger datasets results in stronger emergent dynamics, encoding task-relevant information. These findings emphasize the importance of emergence-based approaches for studying neural networks—biological or artificial—as they enable network-level insights, complementing traditional single-neuron-based analyses.

INTRODUCTION

Biological nervous systems may be regarded as collections of weak learners: individual neurons, each with limited predictive power over the system's environment. Yet, as a whole, these networks perform highly non-trivial computations that support the prediction of environmental dynamics necessary for survival.

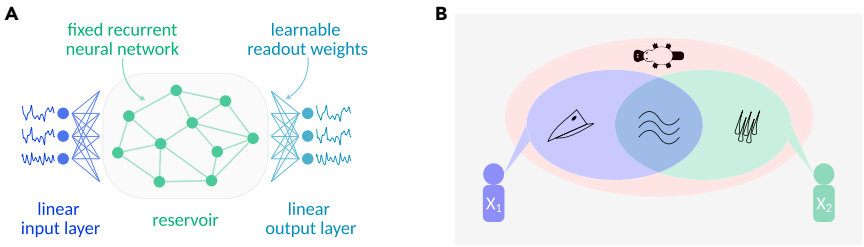

Figure 1. Schematic illustration of central methodological concepts

(A) The basic components of a reservoir computer. The reservoir, a recurrent neural network with fixed internal weights, serves as the core processing unit. Inputs are fed to the reservoir via a linear weight matrix. Outputs are computed by another linear weight matrix, the so-called readout weights, which are learned.

(B) Illustrative example of the partial information decomposition (PID) of Shannon's mutual information into unique, redundant, and synergistic information. Two witnesses (X_1 and X_2) are interrogated to identify the type of animal (Y) that seems to have consumed large parts of the seasonal harvest. X_1 reports that the animal had a beak and that when they approached, it fled to hide in the pond. X_2 confirms that they also observed the animal swimming in the pond and that, additionally, they saw that the animal was furry. Taken together, both sources X_1 and X_2 provided redundant information (that the animal can swim) and unique information (that the animal has a beak and is furry) about the target Y. However, the identity of Y was only revealed after the information from both X_1 and X_2 was combined (synergistic information) since the only animal with a beak and fur is a platypus!

From a reductionist perspective, a complete understanding of a system's computations can, in theory, be achieved by describing the behavior of each of its components and their interactions. Yet in practice, it is overwhelmingly the case that biological neural networks are partially observable systems, rendering a full description of components and their interactions intractable.

In this context, machine learning frameworks for modeling computations in biological neural networks offer an attractive avenue, enabling full observability and controlled interventions. One popular such framework is reservoir computing,[1,2] which was originally introduced as a model of spatiotemporal sensory information processing in the brain.[3] Reminiscent of biological neural networks, the core processing unit of a reservoir computer (RC) constitutes a fixed recurrent neural network (RNN), called the reservoir (Figure 1A). The reservoir receives input from a linear input layer and projects it to a high-dimensional non-linear space, facilitating linear separability.[4] Subsequently, the reservoir state is mapped to a desired output via a linear readout layer. Importantly, RC training involves optimizing only the readout weights, which allows for fast and efficient training and the flexibility to use any arbitrary RNN as the reservoir. Notably, there has been a recent trend of using RCs with bio-inspired reservoir topologies, informed by empirical brain connectivity data, to relate structural properties of brain networks to computational functions.[5–7]

Despite full observability and the possibility to directly assess the contribution of individual artificial neurons to a computation through methods such as neuron deactivation (dropout), elucidating how artificial neural networks (ANNs) perform complex computations remains challenging. One reason for this is that relevant information for these computations may not be encoded at the level of single neurons but instead at the level of groups of neurons or even the whole network. Such macroscale phenomena that seem irreducible to their constituent parts are characteristic of both artificial and biological neural networks.[8] For instance, in the human brain, information from various sensory modalities is processed in different cortical areas, yet we perceive objects as whole, unified entities rather than disjointed collections of sensory inputs. Exactly how the brain achieves this integration of distributed information streams poses a major question in neuroscience, commonly referred to as the "binding problem."[9]

The binding problem exemplifies how the macroscale can possess qualities that are absent at the microscale. This perhaps somewhat paradoxical condition is captured by the concept of emergence,[10] with emergent phenomena being in some sense "greater than the sum of their parts."[11,12] Emergence offers a promising conceptual approach to addressing the challenges faced by traditional reductionist methods in explaining complex computations in neural networks. However, quantifying the role of emergence in computation has long been limited by the lack of effective measures.

Recently, Rosas et al.[13] developed a framework that offers a formal definition of causal emergence alongside practical tools for measuring emergence in empirical data. Drawing on the theory of partial information decomposition (PID),[14] the framework demonstrates how macroscale phenomena can have excess causal power over the evolution of the system, beyond what can be accounted for by considering a microscale description of the system (see Mediano et al.[15] for a comprehensive review). Please note that in this context, "causal power" means predictive power, consistent with the Granger interpretation of causality.[16]

PID, which was introduced as an extension of Shannon's information theory, proposes conceptualizing the information that multiple sources provide about a target as a composition of distinct information atoms: (1) redundant information that is shared between sources, (2) unique information that is exclusive to each source, and (3) synergistic information that can only be accessed by integrating the information from multiple sources[14] (Figure 1B). Building on this approach, Rosas et al.[13] posit that an emergent feature must encode synergistic information about the future of the system.

To illustrate the rationale behind this approach, note that synergistic information only exists at the macroscale—it is not contained in any subset of system parts—yet, it is disclosed when all parts are known and considered together. In other words, synergy elegantly accommodates excess causal (i.e., predictive) powers at the macroscale without violating reductionist principles. Hence, the causal emergence framework is entirely compatible with ontological reductionism while having great potential to provide informative insights into part-whole relationships.

In this study, we explore the hypothesis that complex computations supporting the prediction of environmental variables in biological neuronal networks rely on emergent dynamics. More

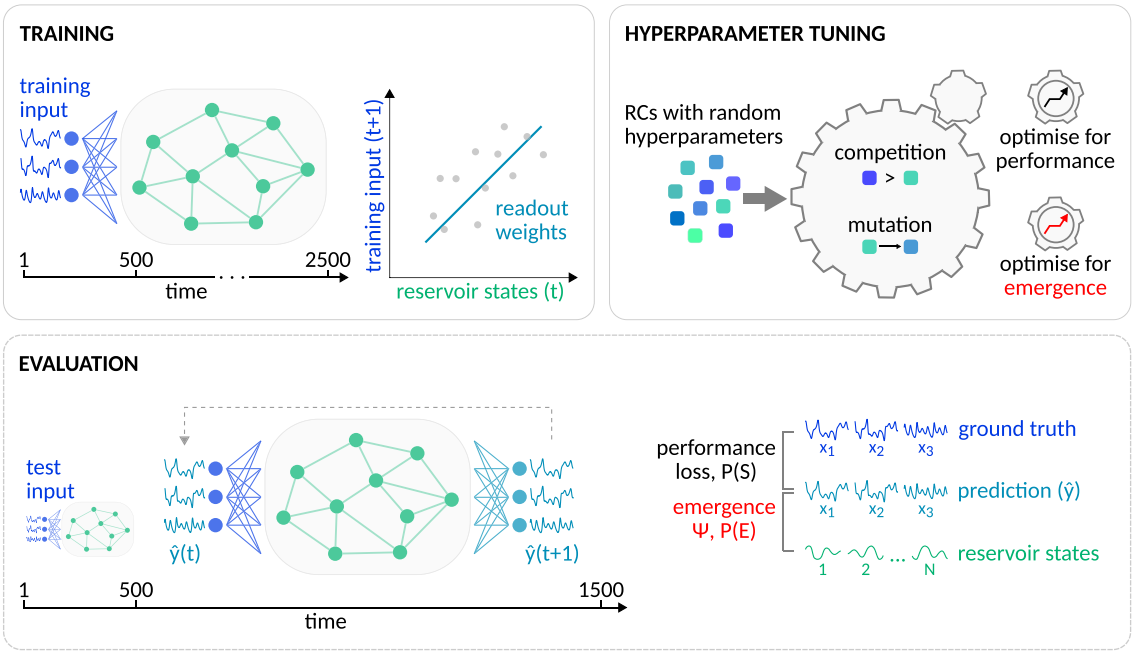

Figure 2. Schematic overview of the approach

Our approach comprises three main operations: training, evaluation, and hyperparameter tuning. Training involves recording the reservoir state trajectory as the RC is fed with some training input and then computing the readout weights with closed-form linear regression (Tikhonov-regularized ridge regression) to predict the future training input at time t + 1 from the current reservoir states at time t. Evaluating the prediction performance and emergent dynamics of a trained RC requires generating predictions by employing an iterative one-step-ahead prediction approach after the RC was fed with the initial 500 time steps of the test input. Prediction performance is assessed by comparing the ground-truth environmental dynamics with the forecast, measuring prediction loss and the probability of a successful prediction P(S) with S: = loss < 1. Emergent dynamics are evaluated by comparing the forecast with the underlying reservoir state trajectory, measuring ψ, i.e., the amount of self-predictive information of the forecast that is not contained in individual reservoir neurons. Additionally, we also estimated the probability of emergence P(E) with E: = ψ > 0. Finally, hyperparameter tuning is performed using an evolutionary algorithm to optimize specific RC hyperparameters. We employed different objective functions, optimizing for either maximal prediction performance or emergent dynamics.

specifically, we leverage reservoir computing to model the prediction of environmental dynamics in a bio-inspired computational architecture, and we capitalize on the causal emergence framework to examine the relationship between emergence and prediction performance. Notably, we find evidence for a bidirectional coupling between emergence in the system's dynamics and prediction performance. Overall, our results reveal clear empirical benefits of considering emergent dynamics for the study of complex computations in neural networks. In turn, these *in silico* results may provide insights into computation in biological neural networks.

RESULTS

To examine the role of emergence in facilitating predictions of environmental dynamics in biological neural networks, we optimized bio-inspired RCs to forecast the trajectories of various chaotic dynamical systems, representing the RCs' environments. Our set of task environments included the well-known Lorenz attractor[17] and five Sprott chaotic flow systems.[18] The reservoir connections of our RCs were determined by a human connectome representation describing the anatomical connections between 100 cortical brain regions,[19] which was derived from empirical data of 100 healthy subjects of the Human Connectome Project (HCP).[20]

Our approach for investigating the relationship between prediction performance and emergence in reservoir computing comprised three main operations: training, evaluation, and hyperparameter tuning (Figure 2).

During training, the RC is first "driven" with the training input. That is, the reservoir state is iteratively updated as an environmental time series is fed to the network via the linear input weights. Subsequently, the readout weights are analytically solved using closed-form linear regression with Tikhonov regularization, also known as (a.k.a.) ridge regression, where the future training inputs at time t + 1 are regressed against the generated reservoir states at time t. Unless stated otherwise, our training inputs were 2,500 time steps long, and the first 500 time steps were not included in the readout weight fitting.

The evaluation step involves measuring the prediction performance and emergent dynamics of a trained RC. Predictions were initialized by driving the RC with the first 500 time steps of the test input. Then, the actual prediction (forecast) was generated by turning the RC into an autonomous system that drives itself with its own prediction output in an iterative one-step-ahead prediction approach for 1,000 time steps. For each evaluation of one RC, multiple (100, unless otherwise specified) forecasts were generated by initializing the RC with different test inputs, and the RC's performance was measured by computing its average prediction loss (see methods for a precise definition)

across all generated forecasts. Additionally, we also derived the probability of a successful prediction P(*S*) with *S*: = loss < 1 as a less outlier-sensitive measure of performance.

The tendency of an RC to exhibit emergent dynamics was assessed using ψ, which provides a lower bound on causal emergence[13] (see methods). In this study, ψ measures the predictive information of the forecast $\hat{y}(t)$ about its immediate future time point $\hat{y}(t+1)$, minus the total information provided by each individual reservoir neuron considered separately. Note that the forecast, being a linear combination of all reservoir neuron states, depends on the full microstate of the system at time t and therefore constitutes a valid macroscopic feature under the causal emergence framework.[13] $\psi > 0$ provides a strict and sufficient criterion for emergence. However, when reservoir neurons share redundant information, this redundancy is subtracted multiple times, which can drive ψ below zero even in the presence of emergence. Hence, $\psi \leq 0$ does not imply the absence of emergence but rather reflects an inconclusive result. To capture a reservoir's overall tendency to generate emergent predictions, we therefore introduced an additional metric, P(*E*), defined as the probability that a reservoir produces predictions with *E*: = $\psi > 0$. By default, all metrics of performance (loss, P(*S*)) and emergence (ψ, P(*E*)) were estimated over 100 test inputs, each being 1,000 time steps long (not including the initial 500 time steps of the test input that are fed to the reservoir for initialization).

Finally, hyperparameter tuning involves using an evolutionary algorithm[21] to optimize specific RC hyperparameters, such as topological properties of the reservoir network, that are key for effective training. Mimicking biological evolution, in evolutionary optimization, a population of individuals (here, RCs) with randomly initialized hyperparameter configurations is evolved toward a specific optimization objective over the course of many generations of competition between individuals and subsequent mutation of the hyperparameter configurations employed by inferior individuals (Figure 1). In the present study, we evolved RCs with two different objective functions: (1) to maximize prediction performance (optimization objective I) and (2) to maximize causal emergence (optimization objective II).

Overall, this approach allowed us to manipulate performance during hyperparameter tuning and training and evaluate how these manipulations affect emergence, and vice versa. Specifically, we assessed the effect of hyperparameter optimization for optimization objective I and optimization objective II on emergent dynamics and prediction performance. We estimated the correlations between performance and emergence across the entire hyperparameter space without optimization. By varying the input sample size during training, we targeted prediction loss and measured the resulting changes in emergent dynamics. Finally, we analyzed whether hyperparameter optimization for emergence enhances RC performance in a range of unfamiliar (non-optimized) environments and whether a human connectome-based reservoir topology promotes emergent dynamics and performance.

Prediction performance is strongly coupled with emergent dynamics

To examine the relationship between loss and ψ during hyperparameter tuning, we evolved 10 populations, each comprising 100 RCs, for 3,000 generations to the Lorenz task environment, optimizing for low loss (optimization objective I). As expected, the average population loss decreased over the course of artificial evolution (Figure 3A), and the best individual across all populations in the final generation was verified to produce a good forecast (Figure S1), indicating successful optimization. Notably, the decrease in population loss was consistently paralleled by an increase in population ψ (Figure 3A). The increase in ψ was due to both increased self-predictive information of the macroscale feature (i.e., the forecast) and reduced predictive information held by single-reservoir neurons (Figure S2). Indeed, we found significant negative correlations ($p < 0.0001$; see the correlation coefficients in the bottom image of Figure 3A) between the evolutionary trajectories of loss and ψ in all populations. These findings replicated across all tested task environments (Figure S3).

Further analysis uncovered that the inverse relationship between loss and ψ is bidirectional: evolving 10 additional populations, which were initialized identically to those optimized for minimal loss, with the objective to maximize ψ (optimization objective II), led not only to an increase in ψ (as we should expect) but also to a simultaneous decrease in population loss over the course of artificial evolution (Figure 3B). Notably, optimizing for high ψ occasionally resulted in transient spikes in the loss curve, suggesting that ψ maximization may introduce some instability to the reservoir dynamics, leading to attenuated prediction performance. However, the spikes in loss were short-lived, implying that ψ optimization effectively pruned hyperparameter configurations associated with poor performance.

To ensure that the observed relationship between loss and ψ was not primarily driven by RCs with particularly poor performance and highly synchronized (i.e., redundant) reservoir neurons, we examined the coupling between emergent ($\psi > 0$) and successful (loss < 1) predictions across the hyperparameter search space. To this end, we sampled 4,000 RCs with random hyperparameter configurations and estimated their marginal and joint probabilities of producing emergent dynamics and successful predictions for each of the six task environments.

Remarkably, emergent predictions were found to be significantly more likely to be successful, as indicated by positive pointwise mutual information between success and emergence (PMI(*S*,*E*) > 0; Figure 4A). The precise form of this relationship varied by environment: in some cases, we observed strongly emergent reservoirs (high P(*E*)) with limited success (low P(*S*)), while in others, there were reservoirs that succeeded without emergence (low P(*E*) and high P(*S*)) (Figure S5). To further characterize these differences, we adopted graded, observational definitions of sufficiency and necessity from Comolatti and Hoel,[22] with sufficiency defined as P(*S*|*E*) and necessity as 1 – P(*S*|¬*E*). We found that sufficiency values were consistently high across environments, with median P(*S*|*E*) exceeding 0.8 (Figure 4B). Furthermore, in Sprott B, G, K, and R, the absence of emergence was highly predictive of unsuccessful prediction, resulting in high necessity values (Figure 4C).

Given that $\psi > 0$ represents a sufficient criterion for emergence, these results complement our previous findings with the supporting evidence that prediction performance is indeed linked to emergence and not solely to the reduced correlation of reservoir neuronal activity, leading to higher but possibly still negative ψ values.

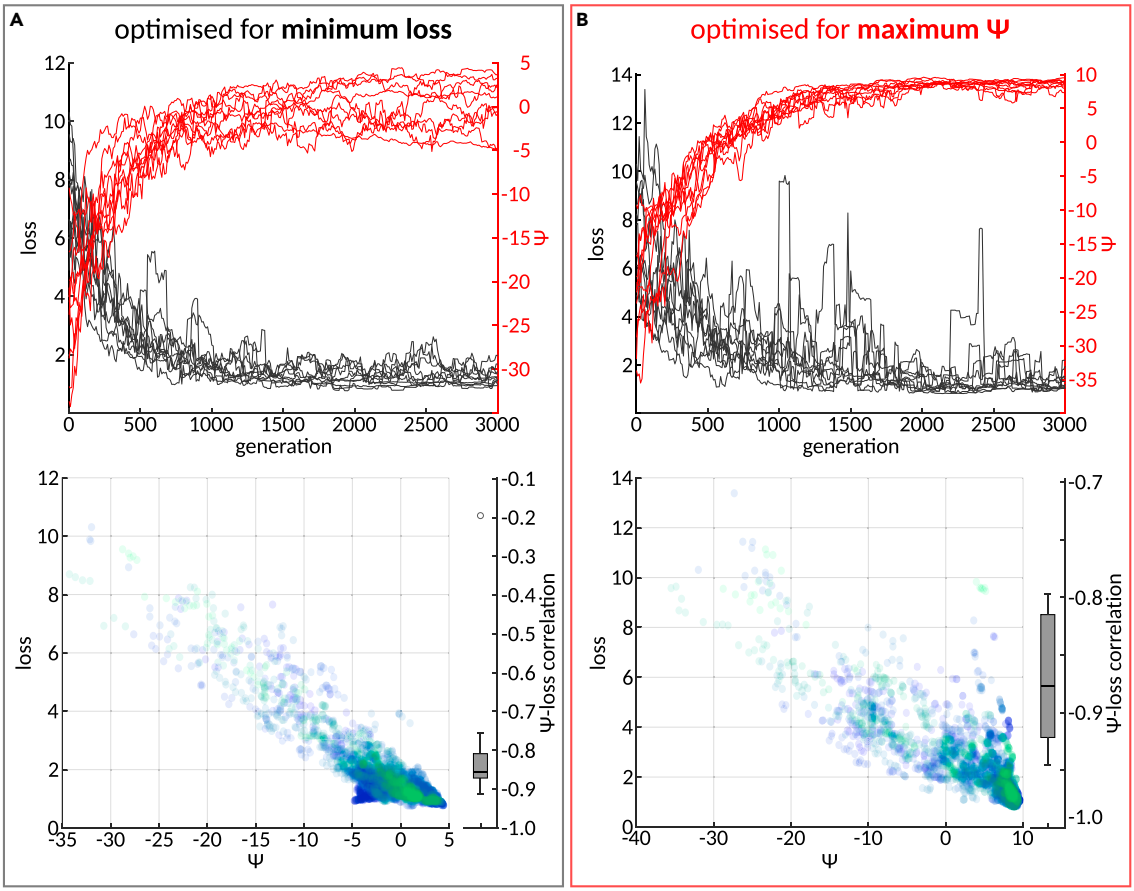

Figure 3. Inverse relationship between loss and ψ during evolutionary optimization

(A) Top: trajectories of average loss (black) and ψ (red) of 10 populations over the course of evolutionary optimization with the objective to minimize loss (optimization objective I). Bottom: average loss plotted against average ψ across generations. Each population is plotted in a different color. The corresponding Spearman correlation coefficients of all populations are summarized in the boxplots (all correlations were significant; $p < 0.0001$).

(B) Analogous to (A) but for evolutionary optimization with the objective to maximize ψ (optimization objective II).

Our previous analyses focused on the relationship between emergence and prediction performance across different hyperparameter configurations. Next, we aimed to explore the relationship between emergence and prediction performance within a single RC while varying the size of the training dataset. Specifically, we trained an RC employing the hyperparameter configuration that yielded the lowest loss across all 10 loss-optimized populations. The RC was trained on 100 different training inputs from the Lorenz environment, and each training was evaluated on the same set of 10 test time series. Crucially, instead of training the RC on the entire length of the input data (2,000 time steps), we trained it on truncated versions with varying sample sizes, i.e., numbers of time steps (n = {10, 20, …,2,000}).

Figure 4E displays the average log(loss) and P(E) across all evaluations for each training sample size n. As expected, loss decreases with increasing training sample size. Interestingly, larger training sample sizes not only led to lower loss but also to a higher probability of emergence, suggesting that fitting the output weights with more data results in enhanced readout of emergent dynamics.

A further analysis revealed that those emergent dynamics are indeed likely to encode task-relevant information (Figure 4F): We trained and subsequently evaluated the same loss-optimal RC 100 times, each time on a new set of 1 training and 1 test time series from the Lorenz environment. Additionally, during each repetition, we also computed the loss and ψ with respect to a random readout of the RC, which was obtained by randomly permuting the trained readout weights, while preserving the weight distributions of the mappings from reservoir neurons to each of the target variables. Unsurprisingly, random readouts yielded significantly greater loss than the predictions generated with trained readout weights ($g = -1.70$, $t = -12.33$, $p < 0.0001$, degrees of freedom [df] = 99). More interestingly, random readouts were also associated with significantly and substantially lower ψ values ($g = 3.47$, $t = 37.63$, $p < 0.0001$, df = 99). Of note, this trend was found to be even more pronounced when repeating the analysis with an RC that employed ψ-optimized, rather than loss-optimized, hyperparameters, where trained readouts nearly always achieved $\psi > 0$, whereas random readouts never did (Figure 4G) (loss comparison: $g = -1.00$, $t = -7.36$, $p < 0.0001$, df = 99; ψ comparison: $g = 5.96$, $t = 51.37$, $p < 0.0001$, df = 99).

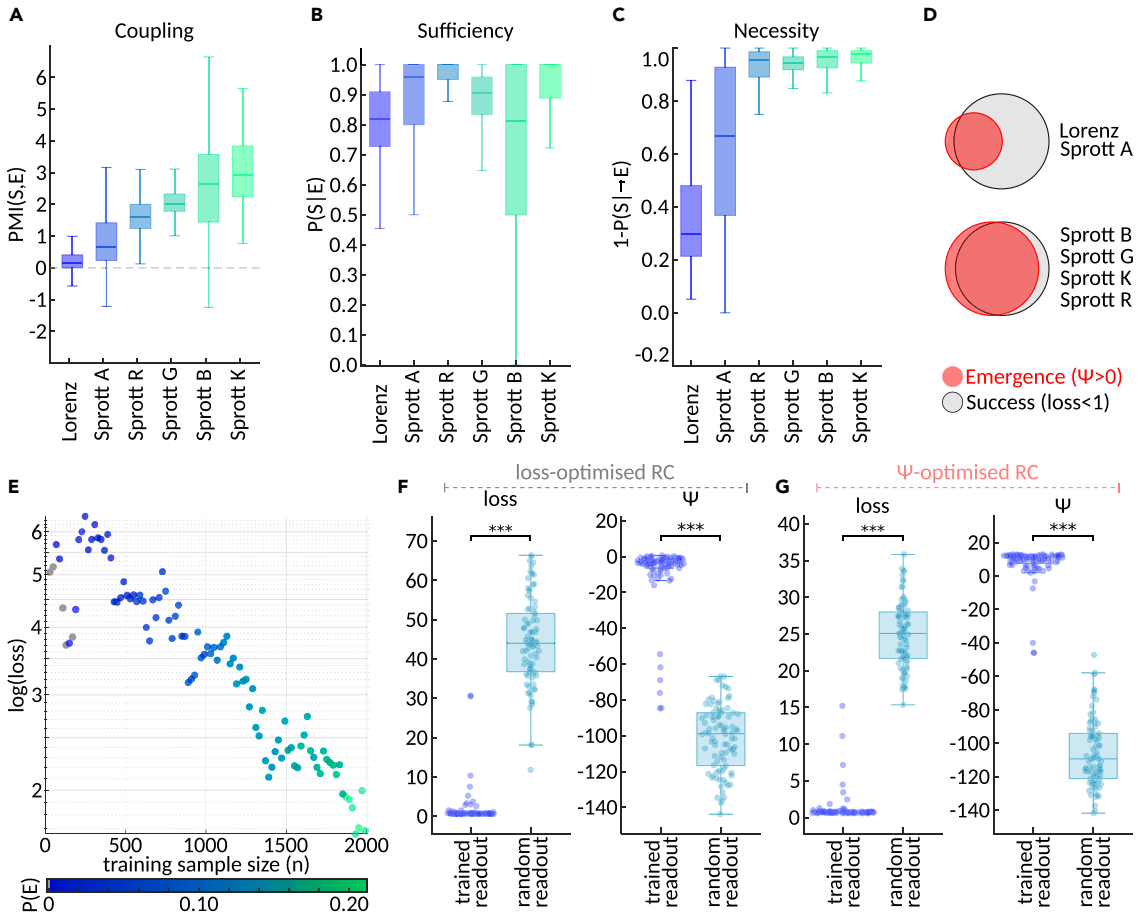

Figure 4. Relationship between emergence and prediction performance across hyperparameter configurations and training sample sizes

(A) 4,000 RCs with randomly sampled hyperparameter configurations were evaluated 100 times on each of the six task environments. The boxplots summarize pointwise mutual information (PMI) between emergence ($\psi > 0$) and success (loss < 1) for each environment. In all cases, PMI values were significantly greater than zero (one-sample t test, $p < 0.0001$ after FDR correction).

(B) Sufficiency, quantified as the probability of success given emergence P($S|E$).

(C) Necessity, quantified as 1 – P($S|\neg E$).

(D) Schematic Venn diagrams of emergence and prediction success in the different task environments, inferred from the results in (B) and (C).

(E) Average log(loss) and probability of emergence achieved by the loss-optimized RC when trained with different training sample sizes (n).

(F) Comparison of loss and ψ values of predictions produced by the loss-optimized RC employing trained versus randomized readout weights.

(G) Analogous to (F), but the predictions were produced by the ψ-optimized RC. Asterisks (***) denote statistically significant differences after FDR correction ($p < 0.0001$).

Emergence-based optimization can facilitate transfer learning to unfamiliar environments

We next investigated whether hyperparameter optimization to maximize emergence may offer advantages over performance-focused optimization in the challenging context of transfer learning. Specifically, based on previous work showing that synergistic information processing facilitates general-purpose learning in ANNs,[23] we wondered whether hyperparameter tuning toward a higher tendency to exhibit emergence—which, by definition, relies on synergistic information[13]—would prime RCs to perform well on a range of unfamiliar environments (that they were not evolved to).

To test this hypothesis, we evolved RCs to maximize various weighted combinations of P(S) and P(E). More precisely, our optimization objective was defined as κ*P(S) + (1 – κ)*P(E), where κ ranged from 0.0 to 1.0 in increments of 0.25. Hence, $\kappa = 0.0$ corresponds to 100% optimization for P(E) and 0% for P(S), and vice versa in the case of $\kappa = 1.0$. For each κ value, we evolved 10 populations for 3,000 generations, each comprising 100 randomly initialized RCs, to a given task environment. The analysis was conducted with four different task environments, the Lorenz attractor and selected Sprott systems (A, B, and R), which were chosen based on the effectiveness of the hyperparameter optimization for these environments in the previous analysis. Finally, the best hyperparameter configuration from each population, which was evolved with a specific κ value and to a specific environment i, was evaluated once on a new set of 100 test time series from the optimized environment i and additionally on a set of 100 test time series from each of the 3 other environments $j \neq i$.

Figure 5 summarizes the test performance (P(S)) in the optimized environment (top row) and the average test performance across all non-optimized environments (bottom row) of the fittest

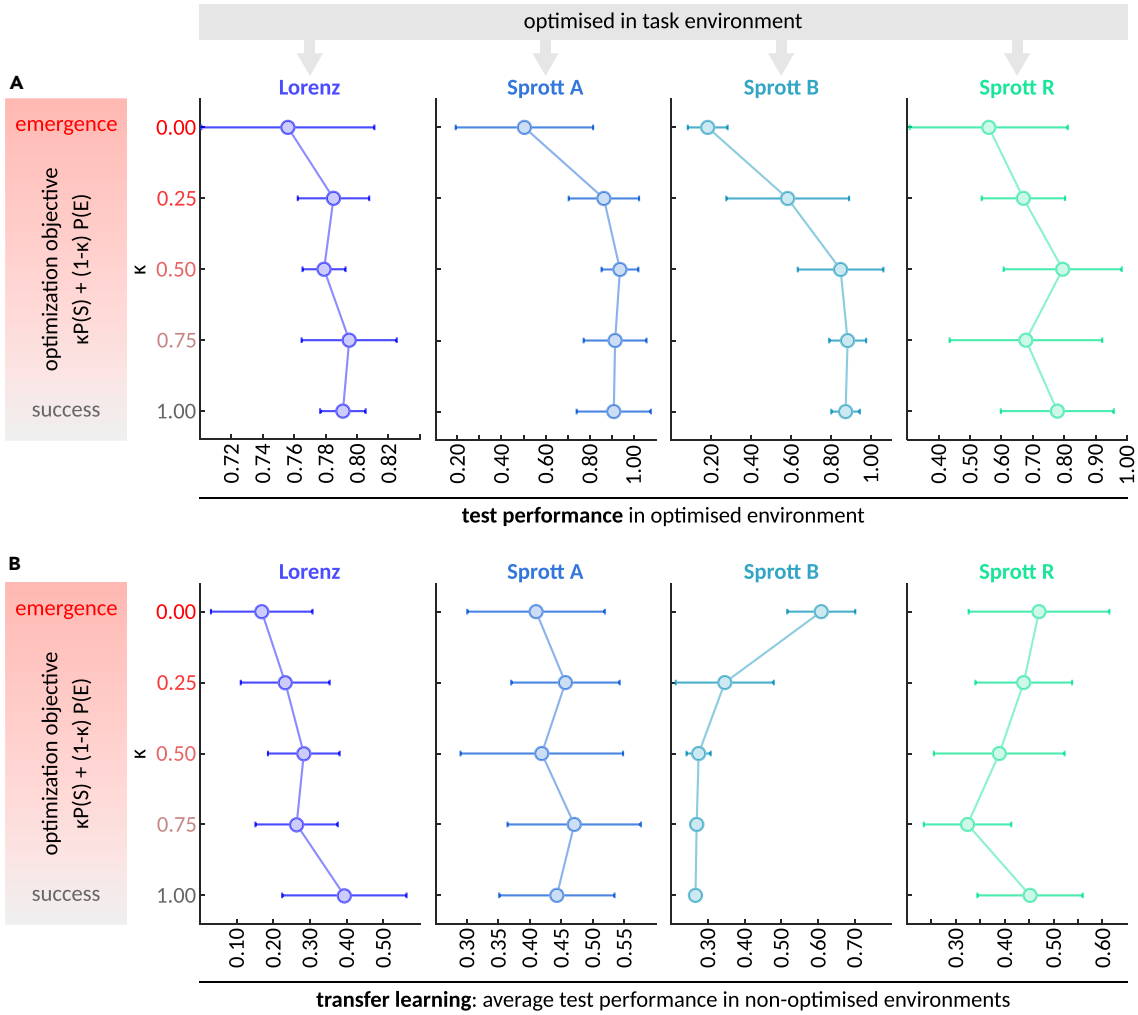

Figure 5. Test performance of reservoir computers, evolved to maximize performance or emergence, across optimized and non-optimized environments

(A) Means ± standard deviations of the test performances (P(*S*)) of the fittest RCs that were evolved to maximize κP(*S*) + (1 − κ)P(*E*) in a particular environment, indexed by the column and the line plot color and evaluated on the same environment. Each dot in the line plot summarizes the performance of the fittest RC from each of the 10 populations that were evolved for the given environment (column) and κ value (*y* axis).

(B) Analogous to (A), but the *x* axis displays the average test performance of the RCs across all other (non-optimized) environments.

RC from each evolved population. As expected, we found that optimizing for P(*S*) alone (i.e., κ = 1.0) tends to yield the best test performance in the optimized task environment. When the RC is evaluated on a *different* task environment, optimizing for P(*E*) may be advantageous in some cases—although not all. While further investigation is needed, the fact that optimization for emergence alone yielded solutions with competitive, and occasionally superior, test performance in non-optimized environments hints at the potential value for future research into the conditions where emergence optimization may be beneficial.

Bio-inspired human connectome topologies offer no performance or emergence advantage over random networks

In standard reservoir computing, reservoir neurons are typically randomly connected such that every neuron is connected to all other neurons with a fixed probability *p*.[1,2] Recently, there has been a trend toward leveraging RCs with bio-inspired reservoir topologies, where the reservoir neurons are connected like brain regions in empirically derived structural brain networks.[5–7] This approach has been proposed for relating structural properties of brain networks to computational functions.

The RCs analyzed above were all constructed using empirical brain network data. To test whether this mattered, we compared the prediction performance and emergence of randomly connected RCs with those of our bio-inspired RCs. More precisely, for each reservoir type (human connectome or random network) and for each of the six task environments, we evolved 10 populations of 100 RCs for 3,000 generations, optimizing for minimal prediction loss. To improve comparability, we ensured that for each population of bio-inspired RCs in a given environment, there was a sister population of randomly connected RCs in

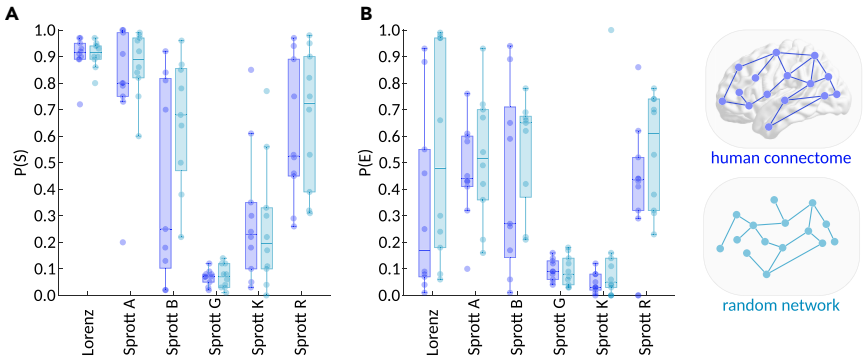

Figure 6. Comparing prediction performance and emergent dynamics of bio-inspired and randomly connected reservoir computers

The analysis involved evolving 10 populations for each task environment and reservoir network type.

(A) Performance (P(*S*)) of the fittest RC from each evolved population of RCs with bio-inspired reservoir topology, informed by the human connectome (dark blue) and RCs with randomly connected reservoirs (light blue).

(B) Emergence (P(*E*)) of the same RCs as in (A). There were no significant differences in performance or emergence between bio-inspired and randomly connected RCs.

the same environment, which was identically initialized at generation 0.

Evaluating the best solution from each evolved population on 100 test time series from the optimized environments revealed no significant differences between bio-inspired versus randomly connected RCs in success or emergence probability (Figure 6). In other words, we found no evidence for systematic differences in prediction performance (P(*S*)) or emergence (P(*E*)) between bio-inspired and random reservoirs. Thus, the coupling between performance and emergence is not specific to human connectome topology but rather appears to be a more general property of reservoir computing networks.

Previous studies comparing RCs with bio-inspired, human connectome-based reservoirs to those with randomly connected reservoirs have yielded seemingly conflicting results: some found worse performance for bio-inspired RCs or equal performance when a slight weight randomization is employed,[7] while others reported better performance for bio-inspired RCs.[5] In this context, our results emphasize the need for further research to determine when reservoir computing represents a suitable model for computation in the human brain. Which aspects are captured, and which ones are not? Furthermore, the mentioned studies[5,7] and our study all differ in their methods, underscoring the importance of identifying which approaches are appropriate for specific types of questions.

Besides pointing out the crucial need for those methodological clarifications, our result suggests that the relationship between emergence and prediction performance observed in previous analyses is not specific to the bio-inspired RC architecture employed in our study, but rather, it may represent a more general principle. Indeed, we confirm that the strong inverse correlations between ψ and loss during hyperparameter optimization are recapitulated in randomly connected RCs (Figure S4).

DISCUSSION

Our study reveals a bidirectional relationship between prediction performance and emergence in reservoir computing, which is consistent across task environments and reservoir topologies. Optimizing hyperparameters for performance enhances emergent dynamics, and vice versa. Emergent dynamics are highly sufficient for prediction success in all environments and necessary in most. Training RCs with larger sample sizes increases the readout of emergent dynamics, indicating that task-relevant information is encoded in synergistic neuronal interactions. Additionally, our results suggest that emergence-based hyperparameter optimization in one environment may promote performance in other, non-optimized environments.

Weak learners collectively become strong through emergence

Our findings highlight the importance of considering emergent dynamics when studying computation in weak learner systems such as RCs. But what exactly do we gain from an emergence-based approach? Emergence-based approaches typically provide tools for identifying optimal representations, or "coarse grainings," of the system that enable better prediction and/or control of its behavior—with different definitions of emergence offering complementary insights. For instance, according to Barnett et al.'s dynamical independence framework,[12] emergence occurs when a coarse graining is optimally predicted based on its own history alone, without requiring further knowledge about the underlying microscale dynamics. Furthermore, Hoel et al. introduced the notion of causal (i.e., interventionist) emergence,[24,25] where the system's response to interventions can be predicted more easily in the macroscale than in the microscale. This approach reveals coarse grainings that allow for more effective control over the system's behavior.

In our study, we used the causal emergence framework by Rosas et al.,[13] where the macroscale representation has excess self-causing powers due to synergistic interactions at the microscale. Unlike Hoel et al.'s framework,[24,25] which is based on interventionist causality,[26] the notion of causal emergence in our study is grounded in Granger causality.[16] Hence, causality ought to be understood in the predictive sense, i.e., the macroscale possesses self-predictive powers. Yet, unlike the dynamical independence framework,[12] Rosas et al.'s approach[13] requires emergent features to have self-predictive powers that *exceed* what is encoded in the microscale dynamics. This makes it particularly suitable for studying time-series prediction in neural networks because it allows for the detection of higher-order interactions between neurons that carry task-relevant information.

To understand why synergy-based causal emergence may be key to enabling environmental predictions in reservoir computing, it is useful to conceptualize each reservoir neuron as a weak learner with limited predictive information about the environment. In the worst-case scenario, all neurons are

perfectly synchronized (fully redundant), rendering the predictive power of the reservoir reducible to that of a single neuron. If instead some neurons provide unique information about the environment, the reservoir can achieve improved predictive power by aggregating the redundant and unique information from each neuron. Additionally, task-relevant information may also be encoded synergistically in any possible combination of neurons. Thus, exploiting synergistic information dynamics allows for greatly amplified representational power, effectively uniting weak learners to form a powerful prediction machine with greater computational capacity than the sum of its parts.

Based on these considerations, one might expect RCs to rely more heavily on emergent dynamics for more challenging tasks. Indeed, some prior work is in line with this idea: one study, evolving ANNs to task environments of varying complexity, demonstrated that integrated information, a metric related to synergy,[27,28] increases over the course of artificial evolution, with the most significant increases occurring in more complex environments[29] (see also Seth[30] for an earlier example, showing increased [Granger] causal connectivity in networks evolved in complex relative to simple environments). However, our exploratory analyses relating sufficiency and necessity to sample entropy and Lyapunov exponents did not reveal a consistent pattern across environments (Figure S6). Future work should examine in more detail how different notions of task complexity shape the reliance of RCs on emergent dynamics.

Our results may seem to suggest that maximizing emergence, and thereby maximizing synergy over redundancy, always leads to improved performance. However, previous studies indicate a trade-off between synergy and redundancy regarding prediction performance.[8,23,31] Greater synergy enhances the capacity to integrate information but can lead to unstable, chaotic dynamics.[23,31] Conversely, greater redundancy provides stable dynamics and robustness to perturbations but results in poor information integration.[8,23,31] This suggests that the relationship between emergence and performance may in fact be non-linear. Indeed, although maximizing ψ reduces overall loss, we observed transient spikes in the evolutionary loss trajectories, suggesting that certain hyperparameter configurations yield poor performance despite emergent dynamics. Notably, these spikes were especially prominent when ψ exceeded the levels achieved by RCs optimized for performance (low loss) rather than emergence (high ψ). Overall, we conclude that optimal performance requires (1) a balance between synergy and redundancy and (2) significantly more synergistic dynamics than randomly initialized RCs.

Possible implications for biological neural networks

Our findings allow for several speculations about the relationship between emergence and prediction performance in biological neural networks.

First, we found that learning a mapping from neural activity to environmental dynamics leads to an increased readout of task-relevant synergistic information and that emergent dynamics may be especially important for solving challenging tasks. Given these results, we should expect to find emergent dynamics, particularly in brain areas that are implicated in learning and performing complex tasks, such as the higher-level association cortices. In support of this idea, recent work revealed a gradient of increasing synergistic compared to redundant dynamics from lower-level sensory to higher-level association cortices.[32,33] Future studies could investigate the levels of synergistic or emergent dynamics in empirical neuroimaging data from subjects performing cognitive tasks with varying degrees of difficulty.

Second, the robust correlation between emergence and prediction performance during evolutionary hyperparameter optimization raises the possibility that human brain evolution may have been influenced by emergence-favoring selection pressure. Notably, a recent study showed that the proportion of synergistic interactions between brain regions is significantly higher in humans compared to macaques, while the proportion of redundant interactions remains preserved.[32] The same study also found that brain regions with highly synergistic dynamics have undergone the greatest expansion during human evolution and display the highest synaptic density and plasticity. Furthermore, dendrites of human pyramidal neurons are capable of purely synergistic computations such as separating linearly non-separable inputs—a feature that has not been shown in non-human neurons.[8,34,35]

Interestingly, a recent theoretical account proposed a potential mechanism for these empirical observations: greater synaptic turnover, and ultimately higher synergy, may have been facilitated by metabolites downstream of a non-oxidative glycolysis pathway that brain regions increasingly relied on as their evolutionary expansion outpaced the oxygen delivery capacity of the vascular system.[36] Given these findings, it would be interesting to compare the synergistic or emergent dynamics in RCs with reservoir topologies of different species, perhaps following the approaches of Damicelli et al.[7] or using the toolbox of Suárez et al.[6]

If the human brain's structure was indeed shaped by evolutionary pressure to support emergent dynamics, what are the precise structural properties that promote these dynamics? The architecture of the human brain exhibits a number of features, including modular organization and small-world character, that clearly distinguish it from random networks.[37] Yet, our results did not show a significant difference in emergence or performance between RCs employing the human connectome versus Erdös-Réyni-type random networks as reservoirs. However, Suárez et al., comparing different randomized versions of the human connectome, did find that preserving the modular organization and topology of the human connectome was essential for optimal RC performance on a memory task.[5] Additionally, two independent whole-brain modeling studies demonstrated that changes in emergent or synergistic dynamics in healthy aging volunteers[38] and coma patients[11] could be explained by underlying structural changes in their connectomes. These findings suggest that the human brain's structure promotes both computational performance and emergent dynamics that our method may have failed to pick up on. Replicating Suárez et al.'s study[5] while measuring both performance and emergence could provide further insights.

Finally, our analysis provokes the question: was the development of emergence-promoting structural properties in biological brains accelerated by evolutionary pressure to maximize predictive power over diverse environments? While our evidence for emergence enhancing performance in unfamiliar environments is preliminary, it aligns with the findings from Proca et al., who

trained ANNs on a diverse set of tasks, demonstrating the critical role of synergy in facilitating multi-purpose learning.[23] Exploring the conditions under which optimizing for emergence enhances transfer learning abilities to unfamiliar task environments promises to yield important insights not only into brain evolution but also for developing improved artificial intelligence.

Limitations and future work

While computational models offer full observability (in contrast to the drastic partial observability of *in vivo* biological networks), their major drawback lies in their inherent abstraction from the real-world system of interest. Our finding that bio-inspired RCs with human connectome topology fail to outperform their randomly connected counterparts implies that critical aspects of brain computation are not captured by our model—at least not at the resolution of macroscale brain networks, where reservoir neurons represent brain regions. Notably, reservoir computing theory indicates that RCs rely on rich internal representations, which partly depend on some randomness in the reservoir's connections.[39] This may explain why the specific topology of the human connectome did not improve prediction performance in our study. However, we do note that others have arrived at opposite conclusions,[5] highlighting the need for further research to resolve these conflicting results. In the meantime, replicating our analysis in an alternative bio-inspired computational framework, such as spiking neural networks,[40] could fruitfully complement our findings. Additionally, invasive neuroimaging techniques such as optogenetics[41] and two-photon calcium imaging[42] can be employed to enhance the observability of biological neural networks and measure emergent dynamics during task performance in animal studies.

Another limitation of our study is that it does not prove a causal relationship between emergence and prediction performance. The interventionist approach to causality, widely considered the gold standard, establishes causation if a controlled intervention on X elicits a response in Y.[26] While our computational model theoretically allowed for controlled interventions, emergence is a property of the system's dynamics, which depend on various factors, rendering a specific perturbation of emergence infeasible. This raises the possibility that we optimized a hidden variable, affecting both performance and emergence without a direct causal link between the two.

However, two arguments support a causal relationship. Firstly, emergence and performance were bidirectionally coupled not only during evolutionary hyperparameter optimization but also when sampling across the entire hyperparameter space, where emergence was highly sufficient and necessary for prediction success in most environments. Secondly, randomly permuting the readout weights of a trained RC, which dramatically impairs prediction performance, resulted in highly non-emergent dynamics. This suggests that the observed emergence is not merely due to autocorrelation effects in the RC's forecast but that it encodes task-relevant information. In essence, when we detect emergence in our study, it means that the RC utilizes synergistic information to compute the next prediction output. We show that this coincides with good performance and that the synergistic information is likely task relevant. Note, however, that these results may depend on the chosen tasks and temporal scale. In our analyses, emergence was computed between consecutive time steps (t and $t + 1$), which is appropriate for the first-order Markovian systems considered here. Whether similar relationships hold for non-Markovian or qualitatively different systems (e.g., discrete) remains an open question for future work.

The lack of consensus on defining the partition of information into redundant, unique, and synergistic atoms poses a limitation on all studies using PID.[43,44] However, although causal emergence in our study is defined based on PID,[13] the metric we used to measure emergence, ψ, does not require the computation of PID information atoms. Instead, it relies on the well-established Shannon mutual information.[13]

On the downside, however, ψ only provides a lower bound on emergence. That is, while $\psi > 0$ is sufficient for emergence, it is not necessary. This may have masked part of the effect in our analysis, suggesting that the true relationship between prediction success and emergent dynamics could be even stronger than measured. Future research should aim to develop more exact measures of emergence that are equally efficient to compute.

Furthermore, ψ does not provide insights into the precise causal architecture underlying the observed emergent dynamics. Tantalizingly, recent work has introduced a framework, utilizing ϵ machines to map emergent mechanisms across multiple scales.[45] Although estimating ϵ machines from empirical data remains challenging in practice, a promising approach employing kernel methods has been proposed.[46] Applying this method to our reservoir computing paradigm could elucidate emergent mechanisms during environmental prediction.

Conclusion

We have uncovered a (causal) bidirectional relationship between causal emergence and prediction performance in a bio-inspired reservoir computing model of forecasting environmental dynamics. This emphasizes the need for leveraging tools capable of capturing synergistic higher-order interactions, such as the ones used here, when studying computation in biological neural networks. Future research expanding on our work should focus on identifying characteristic features of biological neural networks that promote emergent dynamics. Additionally, understanding the conditions under which optimizing for emergence enhances transfer learning to unfamiliar environments remains an important open question. Addressing these questions not only could provide insights into brain evolution but also hold potential for advancing artificial intelligence.

METHODS

Causal emergence

Emergence was measured using the framework of Rosas et al.[13] The framework posits that a macroscale feature, V_t, is an emergent feature of a multivariate system, X_t, if (1) it is supervenient on X_t and (2) it predicts the future of X_t better than any subset of system parts. The condition of supervenience requires V_t to be fully determined by X_t at any time point t, such that there exists a possibly noisy function $F(X_t) = V_t + \epsilon$, where ϵ is an independent noise term.

More formally, this notion of causal emergence can be expressed in the language of PID[14]: by decomposing the

information that the sources V_t and all subsets X_t^α of the N-dimensional system X_t with $\alpha \subseteqq A$ and $A = \{1, 2, \ldots, N\}$ at time t provide about the target $X_{t'}$ with $t' > t$, we obtain redundant, unique, and synergistic information. Causal emergence can then be quantified as the unique information of the source V_t about X_t that is not contained in any X_t^α for all α. If this quantity is greater than zero, then V_t is said to be emergent. V_t is emergent if

$$\mathrm{Un}\left(V_t; X_t \middle| X_t^\alpha\right) > 0 \ \forall.$$

Given the super-exponential growth of the number of possible subsets X_t^α with increasing system size, Rosas et al.[13] introduced the notion of k-th order causal emergence, which occurs when V_t encodes information about the future of X_t beyond what is contained in any set of k or less parts of the system with $k < N$. Note that it is generally more difficult for a feature to meet the requirements for causal emergence of higher orders. In this work, we restrict our analysis to first-order causal emergence, i.e., $\mathrm{Un}(1)\left(V_t; X_t \middle| X_t^i\right) > 0$ for all $i \in \{1, 2, \ldots, N\}$. Furthermore, we capitalize on a computationally efficient proxy measure for causal emergence, which was defined by Rosas et al.[13] as follows (here, it is only shown for $k = 1$).

$$\psi_{t,t'}(V) = \mathrm{I}(V_t; V_{t'}) - \sum_j \mathrm{I}\left(X_t^i; V_{t'}\right)$$

As can be seen, ψ aims to capture the information that a supervenient feature V_t provides about its own future, beyond what is already known after considering each of the system's parts X_t^i separately. Importantly, $\psi > 0$ constitutes a sufficient criterion for V_t to exhibit causal emergence. Note, however, that $\psi \leq 0$ does not necessarily imply the absence of causal emergence. This is due to the fact that the system parts X_t^i may share redundant information about the future of V_t, which is then subtracted multiple times. In fact, for this reason, it is generally harder to detect causal emergence using ψ in systems with high redundancy and many parts.

It is worth pointing out that ψ is always calculated for a specific supervenient feature V_t and a time interval τ that determines the future time point $t' = t + \tau$ of the prediction target $V_{t'}$. In this study, V_t represents the computational output of an RC. Furthermore, we chose $\tau = 1$, which is the most meaningful choice in the context of one-step-ahead prediction as performed by the RCs. Finally, ψ was computed using the software accompanying the theoretical work[13] (available at https://github.com/pmediano/ReconcilingEmergences).

As every mutual information estimate on a finite sample is subject to bias, we performed a bias analysis to confirm that our results are not affected by biased ψ estimates (Figure S7). Note that ψ depends on both reservoir dynamics and the statistical structure of each input trajectory. Thus, we computed ψ separately for each test time series and averaged across time series to estimate the extent to which a reservoir relies on emergent dynamics on average.

Human connectome data

To construct bio-inspired RCs with human connectome topology, we used diffusion tensor imaging (DTI) data from the cohort of 100 unrelated, healthy subjects (46 males; 22–35 years) of the HCP.[20] The DTI scans were acquired in a 3T Siemens Skyra[47] (acquisition scheme = monopolar gradient echo planar imaging [EPI]; voxel size = 1.25 mm isotropic; repetition time [TR] = 5,500 ms; echo time [TE] = 89.50 ms; b values = 1,000, 2,000, and 3,000 s/mm^2; sampling directions = 90 directions per shell + 6 b0 images; flip angle = 2 apodized sinc radio frequency (RF) pulses: 78° and 160°; 111 slices without a gap).

A minimally preprocessed version[48] of the data with corrections for eddy current distortions, susceptibility, and motion artifacts was downloaded from the open-access HCP database. Subsequent preprocessing steps, as detailed by Luppi and Stamatakis,[49] included transformation to Montreal Neurological Institute (MNI)-152 standard space using the q-space diffeomorphic reconstruction tool, implemented in diffusion spectrum imaging (DSI) Studio,[50] and a non-linear registration algorithm from the statistical parametric mapping software (diffusion sampling length ratio = 2.5; output resolution = 1 mm). Finally, white matter fibers were reconstructed with fiber assignment by continuous tracking (FACT),[51] a deterministic tractography algorithm (angular cutoff = 55°; step size = 1.0 mm; minimum length = 10 mm; maximum length = 400 mm; spin density function smoothing = 0.0; the quantitative anisotropy threshold was determined by the signal in the cerebrospinal fluid). A white matter mask was generated by thresholding the resulting images at a quantitative anisotropy of 0.6, and reconstructed fibers that terminated outside of this mask were removed. The tractography algorithm was iterated until the maximum number of 1 million reconstructed streamlines was reached. The streamlines of neighboring voxels were aggregated into 100 parcels representing brain regions according to the Schaefer parcellation,[19] resulting in a symmetric 100×100 matrix per subject encoding the number of axonal fibers between each pair of brain regions. Such a matrix is commonly termed "connectome."[52]

Finally, we derived one group-level consensus connectome from the 100 subject-individual connectomes by averaging the streamline counts between each pair of brain regions across subjects and then setting all matrix entries, which were zero in more than half of the subjects, to zero. Bio-inspired RCs were constructed using the consensus human connectome.

Reservoir computing

Initialization

The basic architecture of an RC in this study comprises (1) a 3-dimensional input layer; (2) the reservoir, a network of 100 recurrently connected neurons; and (3) a 3-dimensional output layer. Input and output layers are fully and linearly connected to the reservoir via the weight matrices W_{in} and W_{out}, respectively. The recurrent connections of the reservoir neurons are given by a symmetric matrix C, which is either a human connectome or a symmetric, random matrix with uniform weights from 0 to 1. Upon initialization of each RC, the density of the reservoir network C (ρ) was adjusted by setting the weakest weights in C to zero. We chose to discard the weakest weights because tractography-based reconstructions of brain structural connectivity are prone to producing false positives, and stronger weights are more likely to represent true axonal fibers.[53] After adjusting the network density, the weights of C were scaled to a range from 0 to 1 and subsequently multiplied by the desired spectral radius

Table 1. RC hyperparameters

Name	Description	Initialization
N	number of reservoir neurons	100
C	reservoir connection weights	human connectome or random weights $c_{i,j} \in [0;1]$ with $c_{i,j} = c_{j,i}$ and $c_{i,i} = 0$
α^*	spectral radius of C	$\alpha = \{0.1, 0.2, \ldots, 2.0\}$
β^*	Tikhonov regularization parameter	$\beta = \{1.0, 1.5, \ldots, 1.0\} \times 10^{-8}$
ρ^*	connection density of C, defined as $2E/(N(N-1))$, where E are the number of non-zero weights in C	$\rho = \{0.01, 0.02, \ldots, 0.15\}$
σ^*	maximum absolute input weight; weights in W_{in} were uniformly sampled from $w_{in} \in [-\sigma; \sigma]$	$\sigma = \{0.01, 0.02, \ldots, 0.1\}$
θ^*	input bias	$\theta = \{0.1, 0.3, \ldots, 1.9\}$

Hyperparameters marked with "*" were optimized.

(α). Finally, all diagonal elements in C were set to zero to improve the comparability between RCs with human connectome and random network topology, as the human connectome lacks self-connections.

Workflow

Following Platt et al.,[54] our reservoir computing paradigm involves three key operations: drive, train, and forecast. Firstly, during the drive operation, the reservoir dynamics are synchronized with those of the task environment by feeding an environmental input time series to the RC while updating the reservoir states r_t^{Nx1} according to

$$r_{t+1} = h\tanh(Cr_t + W_{in}u_t + \theta) + (1 - h)r_t, \quad \text{(Equation 1)}$$

where u_t encodes the input at time t, θ is the input bias, tanh() denotes the hyperbolic tangent, and $h = 0.005$ is the forward-Euler step size. Secondly, the train operation implements the computation of the RC output weights W_{out} via Tikhonov ridge regression.

$$W_{out} = ur^{\mathrm{T}}\left(rr^{\mathrm{T}} + \beta\right)^{-1} \quad \text{(Equation 2)}$$

Here, β^{NxN} is a diagonal matrix with the Tikhonov regularization parameter on its diagonal elements, and u and r, respectively, denote the input and reservoir-state time series that were generated during a preceding drive operation.

Thirdly, the forecast operation produces the forecast of the environmental dynamics in an iterative one-step-ahead prediction approach, where u_t in Equation 1 is replaced with $W_{out}r_t$.

$$r_{t+1} = h\tanh(Cr_t + W_{in}W_{out}r_t + \theta) + (1 - h)r_t \quad \text{(Equation 3)}$$

Building on these three operations, RCs were trained and evaluated as follows. First, the RC was driven with the initial 500 time points of the train input. This short drive operation, a.k.a. spin-up,[54] serves to synchronize the RC with the environmental dynamics. Reservoir states generated during spin-up are not used to compute W_{out}. Following spin-up, the RC was driven with the remaining training input, and the reservoir states generated during this longer drive operation, which, by default, lasted 2,000 time steps, were subsequently used to compute W_{out}. Next, the now-trained RC was spun up (i.e., driven) with the first 500 time steps of the test input. This step may be seen as passing the initial condition of the prediction target to the RC. Finally, the environmental dynamics were forecast for 1,000 time steps, and the prediction performance was evaluated.

Evaluation

Inspired by Platt et al.,[54] we defined prediction loss as follows. Let $\epsilon_i(t)$ be the standardized error of predicting the dynamics of the i-th environmental variable:

$$\epsilon_i(t) := |\hat{y}_i(t) - y_i(t)|\frac{1}{\sigma_i}, \quad \text{(Equation 4)}$$

where $|.|$ denotes the absolute difference; $\hat{y}_i(t)$ and $y_i(t)$ are the forecast and ground-truth trajectories of the environmental variable i at time point t, respectively; and σ_i is the standard deviation of $y_i(t)$ across time. Then, we define

$$\text{loss} := \frac{1}{TD}\sum_{t=1}^{T}\sum_{i=1}^{D}\epsilon_i(t)e^{\frac{-t}{T}}, \quad \text{(Equation 5)}$$

where $T = 1{,}000$ is the length of the forecast and $D = 3$ is the number of environmental variables. Later error terms are given exponentially less weight to account for the exponential growth of prediction error due to the chaotic nature of the environmental systems. We also estimated the emergence of each RC test output with respect to the reservoir states using ψ. Additionally, by thresholding loss and ψ, we derived two metrics with improved scalability and less sensitivity to outliers: the probability of emergence P(E), where E denotes the event of $\psi > 0$, and the probability of prediction success P(S), with S denoting the event loss < 1. By default, loss and ψ values are reported as averages across 1 train and 100 test time series, and P(E) and P(S) are estimated based on the same 100 test time series.

Hyperparameter optimization

We optimized five RC hyperparameters: the spectral radius of the reservoir adjacency matrix C (α), the density of C (ρ), the Tikhonov ridge regularization parameter (β), input strength (σ), and input bias (θ) (Table 1). Hyperparameters were tuned for each type of reservoir network (human connectome or random) and for each task environment separately using a microbial genetic algorithm[21] and optimizing either for performance or for emergence. When optimizing for performance, the utility to be maximized, a.k.a. "fitness," was given by negative loss or P(S). When optimizing for emergence, we maximized ψ or P(E) instead.

Table 2. ODEs of each task environment

Environment	ODEs	Initialization
Lorenz	$dx/dt = 10(y - x)$ $dy/dt = x(28 - z)$ $dz/dt = xy - \frac{8}{3}z$	$x \in [40; -20]$ $y \in [50; -25]$ $z \in [50; 0]$
Sprott A	$dx/dt = y$ $dy/dt = -x + yz$ $dz/dt = 1 - y^2$	$x \in [10; -5]$ $y \in [10; -5]$ $z \in [10; -5]$
Sprott B	$dx/dt = yz$ $dy/dt = x - y$ $dz/dt = 1 - xy$	$x \in [10; -5]$ $y \in [10; -5]$ $z \in [10; -5]$
Sprott G	$dx/dt = 0.4x + z$ $dy/dt = xz - y$ $dz/dt = -x + y$	$x \in [5; -3]$ $y \in [4; -3]$ $z \in [6; -3]$
Sprott K	$dx/dt = xy - z$ $dy/dt = x - y$ $dz/dt = x + 0.3z$	$x \in [6.6; -4.7]$ $y \in [4; -2.5]$ $z \in [6.7; -0.8]$
Sprott R	$dx/dt = 0.9 - y$ $dy/dt = 0.4 + z$ $dz/dt = xy - z$	$x \in [7; -5]$ $y \in [7.5; -2.5]$ $z \in [10; -9]$

Genetic algorithms work by "evolving" a population of solutions with different hyperparameter combinations, encoded in so-called genotypes, toward high fitness through random mutations and competition between genotypes. We evolved populations of 100 RCs with different genotypes toward maximal prediction performance or emergence. Upon initialization, the hyperparameters of each RC in the population were sampled from a hyperparameter-specific discrete search space (Table 1). The search spaces for β, σ, and θ were based on the respective hyperparameter choices in a similar study.[54] The search space for α was based on the widely accepted rule of thumb that RCs tend to transition from a stable to a chaotic regime at a spectral radius of about 1.[5,55] Lastly, the search space for ρ was constrained so as not to exceed the density of the consensus human connectome.

Prior to each evolutionary optimization, we generated 1 training and 100 test inputs from the given task environment, which provided the input data throughout the optimization. Subsequently, populations evolved for 3,000 generations: first, the fitness of two randomly selected RCs in the population was compared. Next, each hyperparameter of the inferior genotype had a 20% chance of being "mutated," i.e., shifted to the next higher or lower value in the corresponding search space. The direction of shift was chosen randomly if possible under the constraint that the new value remained within the search space. Following the mutation, each hyperparameter of the inferior genotype was set to the corresponding hyperparameter value of the superior genotype with a probability of 0.2.

Task environments

We implemented six 3-dimensional chaotic dynamical systems, which represented the task environments of the RCs: the Lorenz attractor[17] and five Sprott chaotic flow systems.[18] Sprott identified 19 algebraically simple systems with complex chaotic dynamics. From these, we selected systems A, B, G, K, and R due to their tendency to maintain values within manageable ranges. The ordinary differential equations (ODEs) describing the dynamics of each environment are listed in Table 2. The environmental systems were simulated using forward Euler with a step size $h = 0.005$ and $h = 0.05$ for the Lorenz and Sprott systems, respectively. The initial condition of each generated time series was sampled by first drawing from a uniform distribution over the approximate value range of each system variable (Table 2) and then iterating the system for a random number of time steps between 1 and $\frac{1}{h}$. The resulting position of the system was used as the initial condition of the input time series to be generated.

Statistical analysis

Statistical tests were performed using a permutation-based t test and bootstrapping 10,000 times. Effect sizes were measured with Hedge's g. p values were false discovery rate (FDR) corrected for multiple comparisons.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Hanna M. Tolle (h.tolle23@imperial.ac.uk).

Materials availability

This study did not generate new materials.

Data and code availability

- All code and data for reproducing the results are publicly available at GitHub (https://github.com/Imperial-MIND-lab/emergentReservoirs) and Zenodo (DOI: https://doi.org/10.5281/zenodo.17501267).[56] The accession number for the code and pre-computed data reported in this paper is Zenodo: https://doi.org/10.5281/zenodo.17501267.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

H.M.T. was supported by the Jan Metzger Scholarship in Artificial Intelligence and Adaptive Systems MSc. and the Doctoral Teaching Scholarship from the Department of Computing, Imperial College London. A.K.S. is supported by the European Research Council Advanced Investigator Grant 101019254. A.I.L. acknowledges support of a Wellcome Early Career Award (grant number 226924/Z/23/Z) and St John's College, Cambridge.

AUTHOR CONTRIBUTIONS

Conceptualization, H.M.T. and P.A.M.M.; data curation, A.I.L.; formal analysis, H.M.T.; investigation, H.M.T.; methodology, H.M.T.; project administration, H.M.T.; software, H.M.T.; supervision, P.A.M.M., A.I.L., and A.K.S.; visualization, H.M.T.; writing – original draft, H.M.T.; writing – review & editing, H.M.T., P.A.M.M., A.I.L., and A.K.S.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the write-up of this work, the authors used ChatGPT in some sections for language checks and polishing. No scientific or philosophical content was generated with AI. After using this tool or service, the authors reviewed and

edited the content as needed and take full responsibility for the content of the publication.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.patter.2025.101457.

Received: June 13, 2025
Revised: September 23, 2025
Accepted: November 27, 2025
Published: February 6, 2026

REFERENCES

1. Jaeger, H. The" echo state" approach to analysing and training recurrent neural networks-with an erratum note'. Tech. Rep. 148 German National Research Center for Information Technology GMD Technical Report Bonn, Germany (2001). doi: https://doi.org/10.24406/publica-fhg-291111.
2. Maass, W., Natschläger, T., and Markram, H. (2002). Real-time computing without stable states: A new framework for neural computation based on perturbations. Neural Comput. *14*, 2531–2560. https://doi.org/10.1162/089976602760407955.
3. Buonomano, D.V., and Maass, W. (2009). State-dependent computations: spatiotemporal processing in cortical networks. Nat. Rev. Neurosci. *10*, 113–125. https://doi.org/10.1038/nrn2558.
4. Yan, M., Huang, C., Bienstman, P., Tino, P., Lin, W., and Sun, J. (2024). Emerging opportunities and challenges for the future of reservoir computing. Nat. Commun. *15*, 2056. https://doi.org/10.1038/s41467-024-45187-1.
5. Suárez, L.E., Richards, B.A., Lajoie, G., and Mišić, B. (2021). Learning function from structure in neuromorphic networks. Nat. Mach. Intell. *3*, 771–786. https://doi.org/10.1038/s42256-021-00376-1.
6. Suárez, L.E., Mihalik, Á., Milisav, F., Marshall, K., Li, M., Vértes, P.E., Lajoie, G., and Mišić, B. (2024). Connectome-based reservoir computing with the conn2res toolbox. Nat. Commun. *15*, 656. https://doi.org/10.1038/s41467-024-44900-4.
7. Damicelli, F., Hilgetag, C.C., and Goulas, A. (2022). Brain connectivity meets reservoir computing. PLoS Comput. Biol. *18*, e1010639. https://doi.org/10.1371/journal.pcbi.1010639.
8. Luppi, A.I., Rosas, F.E., Mediano, P.A.M., Menon, D.K., and Stamatakis, E.A. (2024). Information decomposition and the informational architecture of the brain. Trends Cogn. Sci. *28*, 352–368. https://doi.org/10.1016/j.tics.2023.11.005.
9. Feldman, J. (2013). The neural binding problem(s). Cogn. Neurodyn. *7*, 1–11. https://doi.org/10.1007/s11571-012-9219-8.
10. Bedau, M.A. (2008). Downward causation and autonomy in weak emergence. In Emergence: Contemporary Readings in Philosophy and Science, pp. 505–512. https://doi.org/10.7551/mitpress/9780262026215.001.0001.
11. Luppi, A.I., Mediano, P.A.M., Rosas, F.E., Allanson, J., Pickard, J.D., Williams, G.B., Craig, M.M., Finoia, P., Peattie, A.R.D., Coppola, P., et al. (2023). Reduced emergent character of neural dynamics in patients with a disrupted connectome. Neuroimage *269*, 119926. https://doi.org/10.1016/j.neuroimage.2023.119926.
12. Barnett, L., and Seth, A.K. (2023). Dynamical independence: Discovering emergent macroscopic processes in complex dynamical systems. Phys. Rev. E *108*, 014304. https://doi.org/10.1103/PhysRevE.108.014304.
13. Rosas, F.E., Mediano, P.A.M., Jensen, H.J., Seth, A.K., Barrett, A.B., Carhart-Harris, R.L., and Bor, D. (2020). Reconciling emergences: An information-theoretic approach to identify causal emergence in multivariate data. PLoS Comput. Biol. *16*, e1008289. https://doi.org/10.1371/journal.pcbi.1008289.
14. Williams, P.L., and Beer, R.D. (2010). Nonnegative decomposition of multivariate information. Preprint at arXiv. https://doi.org/10.48550/arXiv.1004.2515.
15. Mediano, P.A.M., Rosas, F.E., Luppi, A.I., Jensen, H.J., Seth, A.K., Barrett, A.B., Carhart-Harris, R.L., and Bor, D. (2022). Greater than the parts: a review of the information decomposition approach to causal emergence. Philos. Trans. A Math. Phys. Eng. Sci. *380*, 20210246. https://doi.org/10.1098/rsta.2021.0246.
16. Bressler, S.L., and Seth, A.K. (2011). Wiener–granger causality: A well established methodology. Neuroimage *58*, 323–329. https://doi.org/10.1016/j.neuroimage.2010.02.059.
17. Lorenz, E.N. (1963). Deterministic nonperiodic flow. J. Atmos. Sci. *20*, 130–141. https://doi.org/10.1175/1520-0469.
18. Sprott, J.C. (1994). Some simple chaotic flows. Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics *50*, R647–R650. https://doi.org/10.1103/PHYSREVE.50.R647.
19. Schaefer, A., Kong, R., Gordon, E.M., Laumann, T.O., Zuo, X.N., Holmes, A.J., Eickhoff, S.B., and Yeo, B.T.T. (2018). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri. Cereb. Cortex *28*, 3095–3114. https://doi.org/10.1093/cercor/bhx179.
20. Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E.J., Yacoub, E., and Uğurbil, K.; WU-Minn HCP Consortium (2013). The wu-minn human connectome project: An overview. Neuroimage *80*, 62–79. https://doi.org/10.1016/j.neuroimage.2013.05.041.
21. Harvey, I. (2011). The microbial genetic algorithm. In Advances in Artificial Life. Darwin Meets von Neumann (Springer Berlin Heidelberg), pp. 126–133. https://doi.org/10.1007/978-3-642-21314-4_16.
22. Comolatti, R., and Hoel, E. (2025). Consilience in causation: Causal emergence is found across measures of causation. Entropy *27*, 825. https://doi.org/10.3390/e27080825.
23. Proca, A.M., Rosas, F.E., Luppi, A.I., Bor, D., Crosby, M., and Mediano, P.A.M. (2024). Synergistic information supports modality integration and flexible learning in neural networks solving multiple tasks. PLoS Comput. Biol. *20*, e1012178. https://doi.org/10.1371/journal.pcbi.1012178.
24. Hoel, E.P., Albantakis, L., and Tononi, G. (2013). Quantifying causal emergence shows that macro can beat micro. Proc. Natl. Acad. Sci. USA *110*, 19790–19795. https://doi.org/10.1073/pnas.1314922110.
25. Hoel, E.P., Albantakis, L., Marshall, W., and Tononi, G. (2016). Can the macro beat the micro? integrated information across spatiotemporal scales. Neurosci. Conscious. *2016*, niw012. https://doi.org/10.1093/nc/niw012.
26. Pearl, J., and Mackenzie, D. (2018). The Book of Why: The New Science of Cause and Effect (Penguin Books). https://books.google.co.uk/books?id=a3eFtAEACAAJ.
27. Mediano, P.A.M., Rosas, F.E., Luppi, A.I., Carhart-Harris, R.L., Bor, D., Seth, A.K., and Barrett, A.B. (2025). Toward a unified taxonomy of information dynamics via integrated information decomposition. Proc. Natl. Acad. Sci. USA *122*, e2423297122. https://doi.org/10.1073/pnas.2423297122.
28. Luppi, A.I., Mediano, P.A.M., Rosas, F.E., Harrison, D.J., Carhart-Harris, R.L., Bor, D., and Stamatakis, E.A. (2021). What it is like to be a bit: an integrated information decomposition account of emergent mental phenomena. Neurosci. Conscious. *2021*, niab027. https://doi.org/10.1093/nc/niab027.
29. Albantakis, L., Hintze, A., Koch, C., Adami, C., and Tononi, G. (2014). Evolution of integrated causal structures in animats exposed to environments of increasing complexity. PLoS Comput. Biol. *10*, e1003966. https://doi.org/10.1371/journal.pcbi.1003966.
30. Seth, A.K. (2005). Causal connectivity of evolved neural networks during behavior. Network *16*, 35–54. https://doi.org/10.1080/09548980500238756.
31. Varley, T.F., and Bongard, J. (2024). Evolving higher-order synergies reveals a trade-off between stability and information integration capacity in complex systems. Chaos *34*, 063127. https://doi.org/10.1063/5.0200425.

32. Luppi, A.I., Mediano, P.A.M., Rosas, F.E., Holland, N., Fryer, T.D., O'Brien, J.T., Rowe, J.B., Menon, D.K., Bor, D., and Stamatakis, E.A. (2022). A synergistic core for human brain evolution and cognition. Nat. Neurosci. *25*, 771–782. https://doi.org/10.1038/s41593-022-01070-0.

33. Varley, T.F., Pope, M., Faskowitz, J., and Sporns, O. (2023). Multivariate information theory uncovers synergistic subsystems of the human cerebral cortex. Commun. Biol. *6*, 451. https://doi.org/10.1038/s42003-023-04843-w.

34. Gidon, A., Zolnik, T.A., Fidzinski, P., Bolduan, F., Papoutsi, A., Poirazi, P., Holtkamp, M., Vida, I., and Larkum, M.E. (2020). Dendritic action potentials and computation in human layer 2/3 cortical neurons. Science *367*, 83–87. https://doi.org/10.1126/science.aax6239.

35. Rosas, F.E., Mediano, P.A.M., Rassouli, B., and Barrett, A.B. (2020). An operational information decomposition via synergistic disclosure. J. Phys. A: Math. Theor. *53*, 485001. https://doi.org/10.1088/1751-8121.

36. Luppi, A.I., Rosas, F.E., Noonan, M.P., Mediano, P.A.M., Kringelbach, M.L., Carhart-Harris, R.L., Stamatakis, E.A., Vernon, A.C., and Turkheimer, F.E. (2024). Oxygen and the spark of human brain evolution: Complex interactions of metabolism and cortical expansion across development and evolution. Neuroscientist *30*, 173–198. https://doi.org/10.1177/10738584221138032.

37. Bullmore, E., and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. Nat. Rev. Neurosci. *10*, 186–198. https://doi.org/10.1038/nrn2575.

38. Gatica, M., E Rosas, F., A M Mediano, P., Díez, I., P Swinnen, S., Orio, P., Cofré, R., and M Cortes, J. (2022). High-order functional redundancy in ageing explained via alterations in the connectome in a whole-brain model. PLoS Comput. Biol. *18*, e1010431. https://doi.org/10.1371/journal.pcbi.1010431.

39. Farrell, M., Recanatesi, S., and Shea-Brown, E. (2023). From lazy to rich to exclusive task representations in neural networks and neural codes. Curr. Opin. Neurobiol. *83*, 102780. https://doi.org/10.1016/j.conb.2023.102780.

40. Tan, C., Šarlija, M., and Kasabov, N. (2020). Spiking neural networks: Background, recent development and the neucube architecture. Neural Process. Lett. *52*, 1675–1701. https://doi.org/10.1007/s11063-020-10322-8.

41. Joshi, J., Rubart, M., and Zhu, W. (2019). Optogenetics: Background, methodological advances and potential applications for cardiovascular research and medicine. Front. Bioeng. Biotechnol. *7*, 466. https://doi.org/10.3389/fbioe.2019.00466.

42. Grienberger, C., Giovannucci, A., Zeiger, W., and Portera-Cailliau, C. (2022). Two-photon calcium imaging of neuronal activity. Nat. Rev. Methods Primers *2*, 67. https://doi.org/10.1038/s43586-022-00147-1.

43. Gutknecht, A.J., Makkeh, A., and Wibral, M. (2025). From babel to boole: The logical organization of information decompositions. Proc. R. Soc. A A. *481*, 20240174. https://doi.org/10.1098/rspa.2024.0174.

44. Kay, J.W., Schulz, J.M., and Phillips, W.A. (2022). A comparison of partial information decompositions using data from real and simulated layer 5b pyramidal cells. Entropy *24*, 1021. https://doi.org/10.3390/e24081021.

45. Rosas, F.E., Geiger, B.C., Luppi, A.I., Seth, A.K., Polani, D., Gastpar, M., and Mediano, P.A.M. (2024). Software in the natural world: A computational approach to emergence in complex multi-level systems. Preprint at arXiv. https://doi.org/10.1016/j.conb.2023.102780.

46. Brodu, N., and Crutchfield, J.P. (2022). Discovering causal structure with reproducing-kernel hilbert space ϵ-machines. Chaos *32*, 023103. https://doi.org/10.1063/5.0062829.

47. Uğurbil, K., Xu, J., Auerbach, E.J., Moeller, S., Vu, A.T., Duarte-Carvajalino, J.M., Lenglet, C., Wu, X., Schmitter, S., van de Moortele, P.F., et al. (2013). Pushing spatial and temporal resolution for functional and diffusion mri in the human connectome project. Neuroimage *80*, 80–104. https://doi.org/10.1016/j.neuroimage.2013.05.012.

48. Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J.L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J.R., et al. (2013). The minimal preprocessing pipelines for the human connectome project. Neuroimage *80*, 105–124. https://doi.org/10.1016/j.neuroimage.2013.04.127.

49. Luppi, A.I., and Stamatakis, E.A. (2021). Combining network topology and information theory to construct representative brain networks. Netw. Neurosci. *5*, 96–124. https://doi.org/10.1162/netn_a_00170.

50. Yeh, F.C. (2021). Dsi studio. Zenodo. https://doi.org/10.5281/zenodo.4978980.

51. Yeh, F.C., Verstynen, T.D., Wang, Y., Fernández-Miranda, J.C., and Tseng, W.Y.I. (2013). Deterministic diffusion fiber tracking improved by quantitative anisotropy. PLoS One *8*, e80713. https://doi.org/10.1371/journal.pone.0080713.

52. Sporns, O., Tononi, G., and Kötter, R. (2005). The human connectome: A structural description of the human brain. PLoS Comput. Biol. *1*, e42. https://doi.org/10.1371/journal.pcbi.0010042.

53. Maier-Hein, K.H., Neher, P.F., Houde, J.C., Côté, M.A., Garyfallidis, E., Zhong, J., Chamberland, M., Yeh, F.C., Lin, Y.C., Ji, Q., et al. (2017). The challenge of mapping the human connectome based on diffusion tractography. Nat. Commun. *8*, 1349. https://doi.org/10.1038/s41467-017-01285-x.

54. Platt, J.A., Wong, A., Clark, R., Penny, S.G., and Abarbanel, H.D.I. (2021). Robust forecasting using predictive generalized synchronization in reservoir computing. Chaos *31*, 123118. https://doi.org/10.1063/5.0066013.

55. Yildiz, I.B., Jaeger, H., and Kiebel, S.J. (2012). Re-visiting the echo state property. Neural Netw. *35*, 1–9. https://doi.org/10.1016/j.neunet.2012.07.005.

56. Tolle, H. (2025). Code for the article "Evolving reservoir computers reveals bidirectional coupling between predictive power and emergent dynamics". Zenodo. https://doi.org/10.5281/zenodo.17501267.