Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Functional equivalence without structural convergence: A Monte Carlo study of evolutionary and evolution-strategy search in small neural network architectures

Leon Sandler

sandler.leon@gmail.com

Independent Researcher

Research Article

Keywords: Degeneracy, Evolutionary search, Neural network architecture, Integrated information, Neural complexity, Functional equivalence

Posted Date: November 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11586095/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Functional equivalence without structural convergence: A Monte Carlo study of evolutionary and evolution-strategy search in small neural network architectures

Leon Sandler
Independent Researcher, Northbrook, Illinois, USA
Corresponding author: sandler.leon@gmail.com · ORCID 0009-0007-4584-808X

Abstract

Background. Neural systems can produce similar functions from different circuit parameters (degeneracy). Do independent optimization routes with comparable functions share one architecture?

Methods. We searched symmetric weighted networks (N = 8, 12, 16) for high neural complexity, a Phi-like measure and synergy dominance with a DNA-encoded genetic algorithm, a continuous genetic algorithm (GA), a (1+10) evolution strategy (ES) and random search at equal budgets under linear, AR(1) and nonlinear dynamics. The structure was compared with five descriptors and a simulated perturbation response, as a convergence ratio D (D < 1: routes closer than random networks).

Results. All routes beat random search (neural complexity, normalized: DNA 0.63, GA 0.75, ES 0.86, random 0.46). DNA and GA were not statistically distinguishable (+0.13, 95% CI −0.02 to +0.21), which does not establish equivalent encodings; ES exceeded the GA (+0.10, CI +0.07 to +0.22), so optimization operators contributed. The spectral descriptors did not converge in 9 of the 10 cells (D = 1.02 to 1.34; exception: nonlinear neural complexity, 0.85). The perturbation response showed at most modest convergence (D = 0.90, neural complexity; 1.30, Phi-like). The sampled neural complexity matched the exhaustive values (Spearman 0.997 or higher). The diversity of the high-scoring sets increased from 0.79–0.84 (N = 8) to 0.92–1.06 (N = 12, 16), an uncertain increase. Deleting 10% of edges cut scores from at least 0.8 to 0.34; repair restored 0.64.

Conclusions. A similar function is reached by architecturally diverse networks, and apparent convergence depends on the descriptor and the dynamics. The results concern proxy functionals, not consciousness.

Keywords: Degeneracy; Evolutionary search; Neural network architecture; Integrated information; Neural complexity; Functional equivalence

Code and data: https://github.com/sandlerleon/shapesearch-functional-equivalence (software archive, DOI 10.5281/zenodo.23862548); preprint DOI 10.5281/zenodo.23862546.

1 Introduction

Biological neural circuits can generate similar activity from widely different parameters (Prinz et al. 2004), and degeneracy, the capacity of structurally different elements to yield the same function, is regarded as a basic property of complex biological systems (Edelman and Gally 2001; Marder and Goaillard 2006). A natural computational question follows: When optimization processes with different representations and operators are

asked to maximize the same functional of network structure, do they find the same architecture, or many different ones?

We study this question in a deliberately small, fully controlled setting. We define functionals that reward a balance of integration and segregation, which have been used as proxies for the richness of neural dynamics (Tononi et al. 1994; Barrett and Seth 2011), and search the space of symmetric weighted networks with a DNA-like evolutionary route and with standard continuous optimizers. The structure of the paper has four research questions.

• **Q1 (functional performance).** Do the routes reach high functional values at equal evaluation cost relative to random search?

• **Q2 (structural diversity).** Do independent routes converge on similar architectures, measured by several descriptors, or does the high-scoring set contain many distinct networks, and does this depend on the dynamics, the network size and the measure of similarity?

• **Q3 (function-based equivalence).** Do networks that are structurally distant also differ in their dynamical response to a perturbation, measured with a linear response proxy and with the simulated nonlinear trajectories?

• **Q4 (maintenance and repair).** How gradually does performance degrade under mutation, and after damage, how far does repair restore function and the original structure?

The study is exploratory. It makes no claim about any biological system, and the functionals are proxies. A possible connection to theories that tie consciousness to network structure is discussed only as a future implication (Section 7).

2 Theoretical framework

Let W be a weighted connectivity matrix and $F(W)$ be a scalar functional. For a tolerance ε, we define the *functional equivalence set*

$$E_\varepsilon=\{W:\ s(W) \geq 1-\varepsilon\} \tag{1}$$

where s is a normalized score (Section 3.5). We call a set of routes *structurally convergent* for a descriptor φ if networks from different routes are closer in the φ-space than random networks are. Functional equivalence without structural convergence means that E_ε is populated by networks that are as dissimilar as, or more dissimilar than, random ones. A descriptor that is itself a function of the objective (for example, the integration profile with neural complexity) can show convergence for trivial reasons, and we flag this where it applies.

3 Methods

3.1 Networks and dynamics

A network is a symmetric weight matrix W with N nodes and $N(N-1)/2$ free edge weights in $[-1, 1]$. W is rescaled to a unit spectral radius, so only its relative structure matters. Three dynamics are used.

Linear equilibrium (baseline). The activity is the equilibrium response to unit-variance Gaussian noise ε, $x=(I-aW)^{-1}\varepsilon$, with a=0.8, and its covariance is

$$\Sigma=(I-aW)^{-1}(I-aW)^{-\mathsf{T}} \tag{2}$$

Because the spectral radius of aW is $a<1$, $I-aW$ is invertible. This is an instantaneous linear transformation of the noise, not the stationary solution of a recurrent update, and we state this explicitly.

Linear AR(1). Recurrent process

$$x_{t+1}=aW\,x_t+\varepsilon_t, \qquad \varepsilon_t \sim \mathrm{N}(0,I) \tag{3}$$

is stable because the spectral radius of aW is less than one. For symmetric $W=U\Lambda U^{\mathsf{T}}$, its stationary covariance is $\Sigma=U\,\mathrm{diag}(1/(1-a^2\lambda_i^2))U^{\mathsf{T}}$.

Nonlinear rate network.

$$x_{t+1}=(1-\alpha)\,x_t+\alpha\tanh(g\,Wx_t)+\sigma\,\xi_t \tag{4}$$

with α=0.3, g=1.2 and σ=0.3, run for 500 steps with a 100-step burn-in. The covariance is the sample covariance of the remaining 400 states. A single noise realization is shared by every candidate network in an experiment (common random numbers), so fitness is deterministic.

3.2 Functionals

All three are computed from the covariance Σ. They are proxies, and none measure consciousness.

Neural complexity (CN). With integration

$$I(S)=\frac{1}{2}\left[\sum_{i\in S}\ln\sigma_{ii}-\ln\det\Sigma_S\right] \tag{5}$$

neural complexity (Tononi et al. 1994) is

$$\mathrm{CN}=\sum_{k=1}^{N-1}\left[\frac{k}{N}\,I(X)-\langle I(X_k^j)\rangle_j\right] \tag{6}$$

where $\langle\cdot\rangle_j$ is the average over subsets of size k. When N=8, the average runs over all 2^8-1=255 nonempty subsets (the empty set is excluded by construction, and the $k=N$ term is zero). For N=12 and 16, 40 fixed random subsets per size are used (all subsets where there are fewer than 40); Section 3.7 validates this approximation against the exhaustive computation.

Phi-like measure (PHI). The minimum over bipartitions (A,B) of the Gaussian mutual information divided by $\min(|A|,|B|)$,

$$\mathrm{MI}(A;B)=\frac{1}{2}\left[\ln\det\Sigma_A+\ln\det\Sigma_B-\ln\det\Sigma\right], \qquad \mathrm{PHI}=\min_{(A,B)}\frac{\mathrm{MI}(A;B)}{\min(|A|,|B|)} \tag{7}$$

is a normalized minimum-information-partition measure in the spirit of the "whole minus sum" integration measures reviewed by Barrett and Seth (2011) and Balduzzi and Tononi (2008). It is not the Φ of integrated information theory, which requires a cause–effect structure, and we use the label "Phi-like" throughout. For N=8, all 127 bipartitions are used; for N=12 and 16, only 30 sampled bipartitions per part size are used, so the sampled minimum is an upper bound on the true minimum.

Synergy dominance (SYN). The negative Gaussian O-information (Rosas et al. 2019),

$$\Omega=\frac{1}{2}\left[(N-2)\ln\det\Sigma+\sum_i\ln\sigma_{ii}-\sum_i\ln\det\Sigma_{-i}\right], \qquad \mathrm{SYN}=-\Omega \tag{8}$$

where Σ_{-i} omits node *i*. Positive Ω indicates redundancy-dominated and negative Ω synergy-dominated interdependence. This replaces the spectral-balance control used in an earlier draft, which scored approximately 0.99 for single random networks and was therefore uninformative.

3.3 Search routes

All routes receive the same budget of function evaluations (2,000, or 1,000 for the nonlinear model).

• **DNA.** A genome of $3\times N(N-1)/2$ bases over {A, C, G, T}; each codon sets one edge weight with a positional code (first base large effect, third base small effect). Point mutation at a rate of $1/L$ per base (one expected mutation per genome per generation, $\rho=\mu L=1$), tournament selection ($k=3$), population 20, elitism.

• **GA (continuous).** The same population, tournament and elitism, but with continuous genes, $1/(N(N-1)/2)$ per-gene mutation probability and a Gaussian step of 0.25. Comparing DNA with this route isolates the encoding; comparing the GA with ES isolates the algorithm.

• **ES.** A (1+10) evolution strategy with multiplicative step-size adaptation (Beyer and Schwefel 2002).

• **RND.** The best of the budgeted random draws. A separate sample of 1,500 random networks ("single random"; 750 for the nonlinear model) provides the no-search reference.

3.4 Structural descriptors and convergence ratio

Five permutation-invariant descriptors are computed for each optimized network: (i) the sorted spectrum of the rescaled *W*; (ii) a signed triad census (fractions of triples that are fully connected and balanced, fully connected and unbalanced, or have two, one or zero edges, at $|w|>0.3$; cf. (Milo et al. 2002)); (iii) the integration–segregation profile, the mean subset integration as a function of subset size divided by total integration; (iv) the modularity *Q* of $|W|$ from a leading-eigenvector bipartition (Newman 2006); and (v) the distribution of the normalized linear impulse-response magnitudes of $(I-\alpha W)^{-1}$, summarized by 16 quantiles. Descriptor (v) is a linear-response proxy and is computed from *W* in the same way for all three dynamics.

Nonlinear perturbation response. Because the descriptor (v) is a linear proxy, the nonlinear network is also characterized by its own simulated trajectories. For each network and each node *i,* the rate network is run twice with the same noise, once undisturbed and once with a one-step pulse of size A_p added to node *i* at the end of the burn-in. With $x^{(i)}_{j,t}$ the perturbed and $x_{j,t}$ the undisturbed trajectory,

$$R_{ij}=\frac{1}{H}\sum_{t=1}^{H}\left|x_{j,t}^{(i)}-x_{j,t}\right| \tag{9}$$

with $H=20$ steps after the pulse. The permutation-invariant descriptors are the 16 quantiles of $R_{ij}/\max R$ ("nonlinear response") and the sorted vector of evoked magnitudes $\sum_j R_{ij}$ over stimulated nodes ("evoked magnitude"). Pulse sizes $A_p=0.2$, 0.5 and 1.0 are compared (noise standard deviation of 0.3).

Each feature is scaled by its standard deviation across 150 random networks. The convergence ratio is

$$D_\varphi=\frac{\langle d_\varphi(\mathrm{A,B})\rangle}{\langle d_\varphi(\mathrm{rand,rand})\rangle} \tag{10}$$

the mean cross-route distance over the mean distance between random networks. The bootstrap unit is the independent run, not the pair, so correlated pairwise distances are not treated as independent. Within-route ratios are reported as a reference for replicate-to-replicate diversity.

3.5 Normalized score and statistics

In each experiment and for each functional, the normalized score is

$$S = \frac{F - F_{\text{rand}}}{F_{\text{ref}} - F_{\text{rand}}} \tag{11}$$

where F_{rand} is the median of the single-random sample (1,500 networks drawn uniformly in $[-1,1]^{N(N-1)/2}$; 750 for the nonlinear model) and F_{ref} is the best value found by any route, including random search, in that experiment. A score of 0 is therefore a typical random network, and 1 is the best score. Because F_{ref} depends on the experiment, absolute normalized values are not comparable across experiments; the within-experiment contrasts carry the inference. Two different random references appear and are kept apart throughout: "single random" is one uniformly drawn network, whereas "best of 2,000 random" is the RND route, the best of a full evaluation budget of random draws, and its median over runs is a route result similar to the others.

Routes within an experiment are independent runs drawn sequentially from one random stream; they are not paired, and no seeds are shared between routes, so unpaired procedures are appropriate. Differences between routes are reported as differences in medians with bootstrap 95% confidence intervals (Efron 1979) across runs, as Cliff's δ (Cliff 1993) and as Mann–Whitney p values (Mann and Whitney 1947). The p values are not corrected for the 15 contrasts of Table 2, and a nonsignificant difference is not read as evidence of equivalence; the intervals carry the inference. A frequency of zero in a sample of n random networks is reported only as a one-sided 95% upper bound, $1-0.05^{1/n}$ (approximately $3/n$), under the sampling distribution used.

Automated tests in the repository compare CN and PHI at N=8 with an independent explicit enumeration of all subsets and bipartitions (agreement to 10^{-8}), check the closed-form covariances of the linear dynamics, and confirm that the sampled PHI is never below the exhaustive value.

3.6 Experiments

Baseline: 40 independent runs per route and functional (N=8, linear equilibrium). Dynamics: linear AR(1), 40 runs; nonlinear, 40 runs at a budget of 1,000. Scaling: N=12 and 16, 40 runs, CN and PHI. Maintenance: mutation–erosion sweep over $\rho \in [0.1,16]$ and tournament size $k \in \{2,3,5\}$ with 10 runs per cell (population 20, 60 generations); order I is the final population-mean normalized score. Damage and repair: 30 ES-optimized networks (normalized score ≥0.8); a random fraction of edges (0.1–0.6) is set to zero in two independent draws per network; and repair is a (1+10) ES with initial step 0.1 and 300 evaluations. The analysis unit is the network, whose two draws are averaged. Sensitivity: one-at-a-time sweep of a, ρ, k, budget and population (12 runs per setting). Uncertainty propagation: 40 draws from priors ($a \sim U[0.5,0.9]$, μ log-uniform in [0.003, 0.03], $k \in \{2,\ldots,5\}$, budget in [800, 3000]) with 3 replicates each. All random seeds are fixed.

3.7 Validation of the sampled functionals

At N=12 and 16, the CN and PHI use sampled subsets and bipartitions. To test this approximation, the final network of each of the 40 runs of every route (DNA, GA, ES) and 100 single random networks (220 per size and functional) were rescored with all 2^N-2 nonempty proper subsets and all bipartitions (4,082 subsets and 2,047 bipartitions at N=12; 65,518 and 32,767 at N=16). We report the rank correlation between exact and sampled scores, the relative error, and the median normalized scores of each route when both the random reference and the best value are computed with the exact functional.

3.8 Reproducibility and use of AI tools

The code (ShapeSearch_core.py, run_experiments.py, make_report.py) regenerates every table and figure from fixed seeds. A clean rerun of the version-2 code from the supplied scripts, in an empty directory (Python 3.11.9, NumPy 2.4.6, SciPy 1.17.1), was compared with the tables of the earlier draft before any change was made (Section 4.8). An AI assistant (Claude, Anthropic) was used under the author's direction to help write and test the simulation code, to run the analyses and to draft and revise the text; the author specified the questions, checked the outputs and the references and takes full responsibility for the content.

4 Results

4.1 Functional performance

Functional	DNA	GA (continuous)	ES (continuous)	Best of 2000 random
Neural complexity	0.63 [0.55, 0.73]	0.75 [0.63, 0.78]	0.86 [0.85, 0.86]	0.46 [0.44, 0.48]
Phi-like (min. bipartition MI)	0.60 [0.54, 0.65]	0.61 [0.57, 0.70]	0.90 [0.86, 0.92]	0.44 [0.43, 0.47]
Synergy dominance (-O-information)	0.90 [0.90, 0.91]	0.91 [0.91, 0.92]	0.93 [0.93, 0.93]	0.78 [0.77, 0.79]

Table 1. *Median normalized score (95% bootstrap CI) by route and functional (N = 8, linear equilibrium, 40 runs per route; 0 = median single random network, 1 = best found; "best of 2,000 random" is the RND route).*

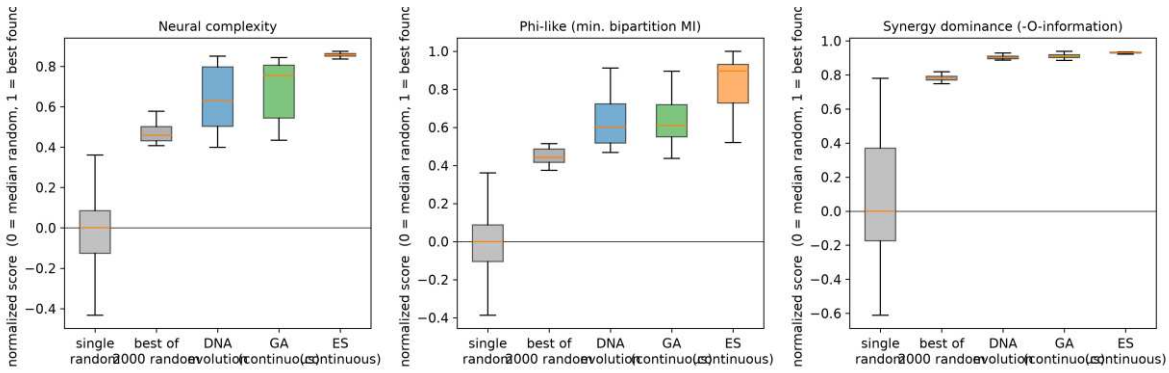


Fig. 1. *Normalized scores for a single random network, the best of 2,000 random networks, and the three optimization routes (N = 8, linear equilibrium dynamics, 40 runs per route).*

Every route beat random search for all three functionals (Table 1, Fig. 1). For neural complexity, the median normalized score was 0.63 for DNA evolution, 0.75 for the continuous GA, 0.86 for ES and 0.46 for the best of 2,000 random networks. Table 2 presents the contrasts. DNA beat random search by +0.17 (CI +0.09 to +0.27; Cliff's δ = 0.72) and ES by +0.40 (δ = 0.97). The Phi-like measure was +0.16 for DNA (δ = 0.91) and +0.45 for ES (δ = 1.00). The synergy functional was easy: the best of the 2,000 random networks already reached 0.78, and the routes increased from +0.12 to +0.15.

Functional	Contrast	Difference in median [95% CI]	Cliff's delta	Mann–Whitney p
CN	DNA vs RND	+0.17 [+0.09, +0.27]	+0.72	<0.001
CN	GA vs RND	+0.29 [+0.17, +0.33]	+0.76	<0.001
CN	ES vs RND	+0.40 [+0.38, +0.42]	+0.97	<0.001
CN	GA vs DNA	+0.13 [-0.02, +0.21]	+0.13	0.310
CN	ES vs GA	+0.10 [+0.07, +0.22]	+0.93	<0.001
PHI	DNA vs RND	+0.16 [+0.09, +0.22]	+0.91	<0.001
PHI	GA vs RND	+0.17 [+0.12, +0.26]	+0.92	<0.001
PHI	ES vs RND	+0.45 [+0.41, +0.48]	+1.00	<0.001

Functional	Contrast	Difference in median [95% CI]	Cliff's delta	Mann–Whitney p
PHI	GA vs DNA	+0.01 [-0.06, +0.11]	+0.08	0.541
PHI	ES vs GA	+0.29 [+0.19, +0.33]	+0.72	<0.001
SYN	DNA vs RND	+0.12 [+0.12, +0.13]	+0.99	<0.001
SYN	GA vs RND	+0.13 [+0.12, +0.14]	+1.00	<0.001
SYN	ES vs RND	+0.15 [+0.15, +0.16]	+1.00	<0.001
SYN	GA vs DNA	+0.01 [+0.00, +0.02]	+0.34	0.008
SYN	ES vs GA	+0.02 [+0.01, +0.03]	+0.80	<0.001

Table 2. *Contrasts between routes: difference in median normalized score [95% bootstrap CI], Cliff's δ and Mann–Whitney p (unpaired, uncorrected for the 15 contrasts).*

The matched-algorithm contrast addresses the encoding. Under the conditions tested, the difference between the continuous GA and the DNA route was not statistically distinguishable from zero for neural complexity (+0.13, CI −0.02 to +0.21, p = 0.310) or for the Phi-like measure (+0.01, CI −0.06 to +0.11, p = 0.541). This is an absence of evidence for a difference, not evidence of equivalence: the interval for neural complexity still admits a difference of approximately 0.21 in favor of the GA, and for the synergy functional, the continuous GA was slightly but detectably better (+0.01, p = 0.008). ES clearly outperforms the continuous GA (+0.10 for neural complexity, +0.29 for the Phi-like measure). The lower DNA scores relative to ES therefore indicate that the optimization operators, such as the population structure, selection and step-size adaptation of ES, contribute to performance at this budget; because ES differs from the GA in several respects at once, the present design cannot say which of them matters, and it cannot rule out a modest effect of the four-letter encoding.

4.2 Structural convergence

Descriptor	D cross-route (DNA vs ES), CN	D cross-route (DNA vs ES), PHI	D cross-route (DNA vs ES), SYN	D within ES runs, CN	D within DNA runs, CN
spectral	1.19 [1.04, 1.37]	1.18 [1.06, 1.30]	0.91 [0.71, 1.15]	0.91	1.14
motif	0.94 [0.82, 1.08]	1.12 [0.93, 1.32]	0.88 [0.75, 1.04]	0.88	1.01
profile	0.61 [0.48, 0.79]	0.68 [0.58, 0.78]	0.52 [0.47, 0.58]	0.29	0.58
modularity	1.01 [0.85, 1.16]	0.79 [0.68, 0.91]	1.44 [1.17, 1.74]	1.06	0.99
response	0.88 [0.82, 0.94]	0.86 [0.79, 0.95]	0.40 [0.35, 0.45]	0.83	0.85

Table 3. *Convergence ratio D (95% CI over runs) between DNA- and ES-evolved networks for each descriptor (D < 1: closer than random networks), with within-route ratios for neural complexity.*

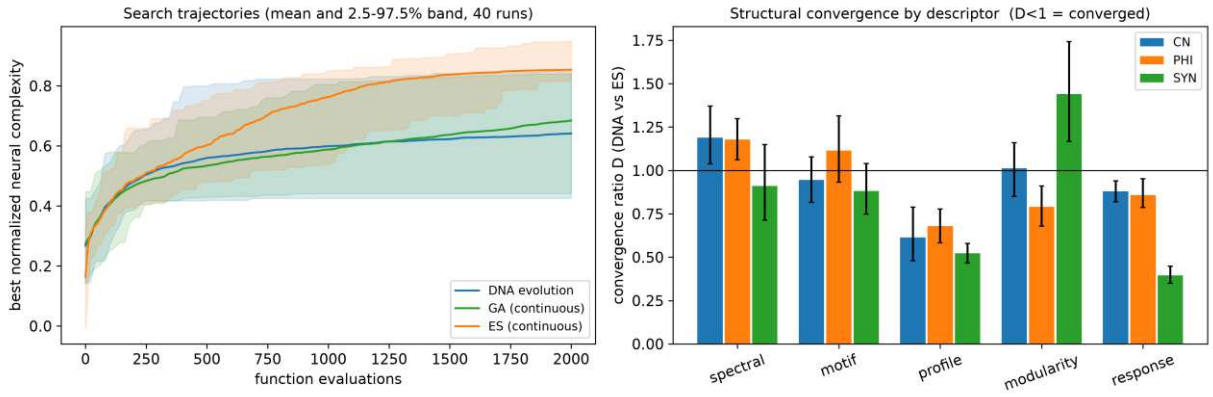


Fig. 2. *Left: search trajectories for neural complexity (mean and 2.5–97.5% band over 40 runs). Right: convergence ratio D between DNA and ES for each descriptor and functional (error bars: bootstrap 95% CI over runs; D < 1 indicates convergence).*

Convergence depended strongly on the descriptor (Table 3, Fig. 2). For the spectrum, D was 1.19 (1.04 to 1.37) for neural complexity and 1.18 (1.06 to 1.30) for the Phi-like measure; thus, the DNA-evolved and ES-evolved networks were no closer and were slightly farther apart than the random networks were, and the synergy D was 0.91 (0.71 to 1.15). Triad motifs gave D = 0.94 (0.82 to 1.08) and 1.12 (0.93 to 1.32). The modularity ranged from 0.79 to 1.44 depending on the functional. The integration–segregation profile showed the lowest ratios (0.61, 0.68 and 0.52), but this descriptor is a function of the same quantity that neural complexity optimizes, so this convergence is partly built in. The linear impulse–response distributions were D = 0.88 (0.82 to 0.94), 0.86 (0.79 to 0.95), and 0.40 (0.35 to 0.45) for the easy synergy functional.

Within-route ratios provide a reference for ordinary replicate diversity. For the spectrum, they were 0.91 within ES runs and 1.14 within DNA runs, comparable to the cross-route value of 1.19; for the profile, ES runs were much more alike (0.29) than DNA–ES pairs were (0.61). The spectral nonconvergence is therefore mostly general diversity among independent optima, not a route-specific signature. The three-way comparison (Table 4) yields similar values of approximately 1.08 to 1.17 for all route pairs for neural complexity and the Phi-like measure; thus, the spectral comparison does not separate the encoding from the algorithm.

Functional	DNA vs ES (encoding + algorithm differ)	GA vs ES (algorithm differs)	DNA vs GA (encoding differs)
CN	1.19	1.12	1.14
PHI	1.18	1.17	1.08
SYN	0.91	0.96	0.72

Table 4. *Spectral convergence ratio D for all route pairs.*

4.3 Functional equivalence set

Taking $\varepsilon = 0.2$ for neural complexity, 60 networks from the three optimization routes (39 ES, 11 GA, and 10 DNA) belonged to the equivalence set. Their mean pairwise spectral distance was 0.79 times that of random networks (bootstrap CI 0.66 to 0.94), and their linear response-distribution distance was 0.86 times that of random networks; thus, the set is somewhat more concentrated than random but is far from collapsed (Fig. 3). Across pairs, spectral distance was moderately associated with the difference in score (Spearman $\rho = 0.32$) and essentially unrelated to the distance between linear response distributions ($\rho = 0.03$). These coefficients are descriptive; pairs share networks and are not independent, so no p values are given. Structurally distant high-scoring networks are therefore not systematically more different in their linear response profile; Section 4.4 repeats the question with the nonlinear trajectories.

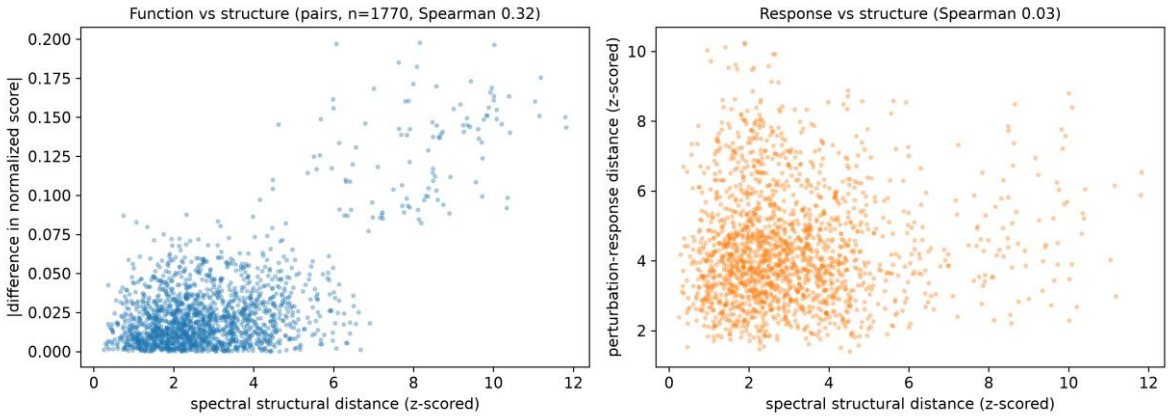


Fig. 3. *Pairs of high-scoring networks (normalized score ≥ 0.8). Left: score difference against spectral distance. Right: linear response-distribution distance against spectral distance.*

4.4 Robustness to the dynamics and the nonlinear perturbation response

Dynamics	Functional	Runs	DNA	GA	ES	Best of random
linear static (baseline)	CN	40	0.63	0.75	0.86	0.46
linear static (baseline)	PHI	40	0.60	0.61	0.90	0.44
linear AR(1)	CN	40	0.62	0.63	0.74	0.53
linear AR(1)	PHI	40	0.59	0.63	0.81	0.47
nonlinear rate network	CN	40	0.82	0.77	0.93	0.66
nonlinear rate network	PHI	40	0.75	0.70	0.83	0.56

Table 5. *Robustness to dynamics: median normalized scores by route (40 runs per cell; nonlinear model at a budget of 1,000).*

Dynamics	Functional	D spectral	D motif	D response
linear static (baseline)	CN	1.19 [1.04, 1.37]	0.94 [0.82, 1.08]	0.88 [0.82, 0.94]
linear static (baseline)	PHI	1.18 [1.06, 1.30]	1.12 [0.93, 1.32]	0.86 [0.79, 0.95]
linear AR(1)	CN	1.34 [1.27, 1.43]	1.12 [0.98, 1.26]	1.27 [1.10, 1.47]
linear AR(1)	PHI	1.02 [0.96, 1.08]	0.98 [0.87, 1.10]	1.11 [1.00, 1.22]
nonlinear rate network	CN	0.85 [0.74, 0.96]	0.99 [0.84, 1.13]	1.02 [0.88, 1.17]
nonlinear rate network	PHI	1.12 [1.00, 1.25]	0.89 [0.77, 1.01]	1.12 [0.95, 1.31]

Table 6. *Robustness to dynamics: DNA–ES convergence ratios D (95% CI) for the spectral, motif and linear-response descriptors.*

The functional enrichment persisted under AR(1) and nonlinear dynamics (Table 5): ES reached 0.74 and 0.93 for neural complexity and 0.62 and 0.82 for DNA, respectively, whereas it was 0.53 and 0.66 for the best of 2,000 random networks (1,000 random draws for the nonlinear model). Spectral convergence was absent in the AR(1) model (D = 1.34 [1.27, 1.43] for neural complexity, 1.02 [0.96, 1.08] for the Phi-like measure; Table 6), but the nonlinear model gave D = 0.85 [0.74, 0.96] for neural complexity, the only spectral cell in which the interval lies below 1, and 1.12 [1.00, 1.25] for the Phi-like measure (Table 6). The partial convergence of the profile and response descriptors seen in the baseline model did not carry over: the profile ratio was 1.72 and 1.15 under AR(1) and, for the nonlinear network, 0.49 for neural complexity but 0.87 for the Phi-like measure. We therefore do not claim that any descriptor converges robustly.

Nonlinear perturbation response. Because the linear-response descriptor is a proxy, the nonlinear networks were also compared with the perturbation response computed from their own trajectories (Section 3.4, Table 7). For neural complexity, the DNA–ES ratios for the nonlinear response were 0.90 (0.81 to 1.01), 0.90 (0.81 to 0.99) and 0.90 (0.82 to 0.99) at pulse sizes of 0.2, 0.5 and 1.0, and 0.89 (0.81 to 0.98), 0.88 (0.79 to 0.97) and 0.87 (0.79

to 0.96) for the evoked magnitude, respectively. The linear proxy on the same networks was 1.02 (0.91 to 1.17). The trajectory-based descriptors therefore detect modest convergence (approximately 10% closer than random pairs, intervals touching or just below 1) that the linear proxy does not. For the Phi-like measure, the nonlinear response was D = 1.31, 1.30 and 1.23 (above 1 for all three pulse sizes), whereas the evoked magnitude was 0.81, 0.77 and 0.80 (below 1). The two functionals thus behave differently; the results do not depend on pulse size, and the DNA–GA ratios (0.97 to 0.99 for neural complexity) show that the two genetic algorithms are approximately as far apart as random networks. In short, perturbation-based descriptors from simulated trajectories give mixed and at most modest evidence of functional similarity between routes, not a clean collapse onto one response profile.

Functional	Pulse amplitude	D nonlinear response, DNA-ES	D nonlinear response, GA-ES	D nonlinear response, DNA-GA	D evoked magnitude, DNA-ES	D linear-response proxy, DNA-ES
CN	0.2	0.90 [0.81, 1.01]	0.91 [0.81, 1.01]	0.99 [0.88, 1.11]	0.89 [0.81, 0.98]	1.02 [0.90, 1.14]
CN	0.5	0.90 [0.81, 0.99]	0.89 [0.80, 0.98]	0.97 [0.87, 1.08]	0.88 [0.79, 0.97]	1.02 [0.91, 1.17]
CN	1.0	0.90 [0.82, 0.99]	0.88 [0.78, 0.98]	0.97 [0.86, 1.08]	0.87 [0.79, 0.96]	1.02 [0.91, 1.14]
PHI	0.2	1.31 [1.11, 1.56]	1.19 [0.97, 1.40]	1.19 [1.01, 1.39]	0.81 [0.74, 0.91]	1.12 [0.95, 1.34]
PHI	0.5	1.30 [1.11, 1.51]	1.20 [0.97, 1.43]	1.20 [1.01, 1.36]	0.77 [0.68, 0.86]	1.12 [0.95, 1.31]
PHI	1.0	1.23 [1.02, 1.45]	1.16 [0.96, 1.33]	1.15 [0.98, 1.36]	0.80 [0.72, 0.88]	1.12 [0.96, 1.31]

Table 7. *Nonlinear perturbation response: convergence ratio D (95% CI) between routes (nonlinear rate network, 40 runs per route). "Nonlinear response" and "evoked magnitude" are computed from simulated trajectories; the last column is the linear-response proxy for comparison.*

4.5 Scaling with network size and validation of the sampled functionals

N	Functional	DNA	GA	ES	Best of random	Random structures with score >= 0.5
8	CN	0.63	0.75	0.86	0.46	0.000
8	PHI	0.60	0.61	0.90	0.44	0.000
12	CN	0.67	0.67	0.79	0.43	0.000
12	PHI	0.67	0.66	0.79	0.40	0.000
16	CN	0.62	0.60	0.81	0.39	0.000
16	PHI	0.65	0.62	0.82	0.38	0.000

Table 8. *Scaling with network size: median normalized scores by route (40 runs per cell) and the fraction of single random networks with a normalized score of at least 0.5 (of 1,500; N = 12 and 16 use sampled subsets).*

N	Functional	D spectral (DNA-ES)	D response (DNA-ES)	Diversity of high-scoring set/random [95% CI] (n networks)
8	CN	1.19 [1.04, 1.37]	0.88 [0.82, 0.94]	0.79 [0.66, 0.94] (n=60)
8	PHI	1.18 [1.06, 1.30]	0.86 [0.79, 0.95]	0.84 [0.71, 0.99] (n=40)
12	CN	1.20 [1.10, 1.31]	1.01 [0.90, 1.14]	0.92 [0.79, 1.08] (n=26)
12	PHI	1.15 [1.05, 1.25]	1.02 [0.91, 1.13]	1.06 [0.93, 1.18] (n=27)
16	CN	1.22 [1.11, 1.34]	0.97 [0.88, 1.06]	1.03 [0.89, 1.16] (n=28)
16	PHI	1.12 [1.03, 1.20]	1.11 [1.00, 1.22]	1.01 [0.89, 1.13] (n=23)

Table 9. *Scaling with network size: DNA–ES convergence ratios D [95% CI] and diversity of the high-scoring set relative to random networks [95% bootstrap CI] (n = number of networks with score ≥ 0.8).*

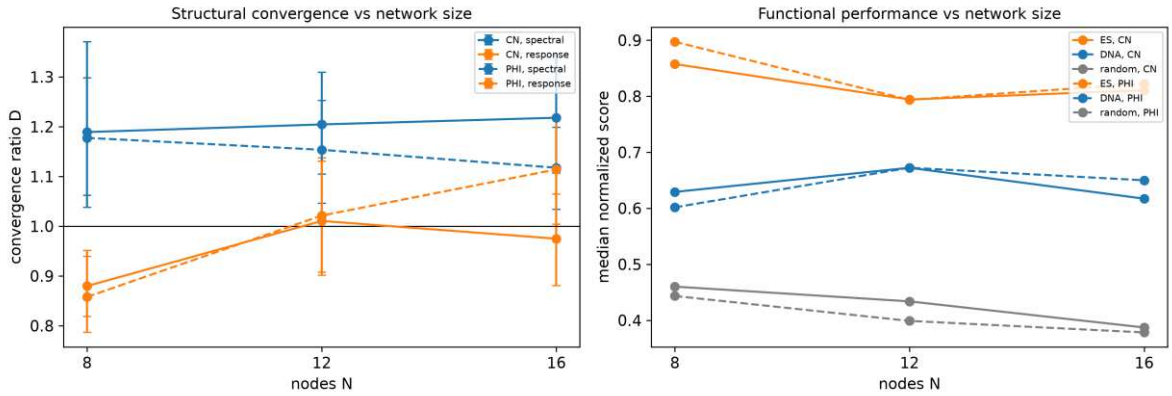


Fig. 4. *Left: convergence ratio D against N for the spectral and response descriptors. Right: median normalized score against N (error bars: bootstrap 95% CI over runs).*

The functional advantage over random search persisted at N = 12 and 16 (ES 0.79 and 0.81 for neural complexity; best of 2,000 random 0.43 and 0.39; Table 8, Fig. 4). The spectral ratio ranged from 1.12 to 1.22 at N = 12 and 16, and every interval excluded 1; thus, the DNA-evolved and ES-evolved networks were slightly farther apart than the random pairs were, as at N = 8. The linear response ratio was close to 1 (0.97 to 1.11).

How rare are high-scoring networks? Two different random references must be kept apart. None of the 1,500 single random networks reached a normalized score of 0.5 for neural complexity or the Phi-like measure at N = 8; with zero events in 1,500 draws, the 95% upper bound on the frequency of such networks, under uniform sampling of the weights, is 0.20%. The best of 2,000 random draws (the RND route), in contrast, had a median of 0.46 and reached 0.5 in 25% of its 40 runs. The two statements are consistent: a frequency of approximately 1.4e-04 per network gives a probability of 25% that a 2,000-draw search finds at least one such network, and an expected 0.2 of them among 1,500 single draws (probability of observing none: 0.81). A zero count in a sample of this size therefore bounds the frequency but does not show that such networks are absent, and it says nothing about other sampling distributions. For N = 16, no RND run reached 0.5 (0 of 40 for both functionals), whereas for the synergy functional, 14% of single random networks already did, so we do not call that functional rare.

Does the diversity of the high-scoring set increase with size? The diversity of the set relative to random networks was 0.79 for neural complexity and 0.84 for the Phi-like measure at N = 8, 0.92 and 1.06 at N = 12 and 1.03 and 1.01 at N = 16 (Table 9). Differences from N = 8 were +0.13 (CI −0.07 to +0.33) at N = 12 and +0.24 (CI +0.06 to +0.42) at N = 16 for neural complexity and +0.22 (CI +0.04 to +0.41) and +0.17 (CI −0.01 to +0.35) for the Phi-like measure. Two of the four intervals exclude zero, the N = 12 and 16 values are indistinguishable from each other and from 1 (random-level diversity), and the number of networks qualifying at each size is only 23 to 28. The data are therefore consistent with a somewhat higher diversity at N ≥ 12 than at N = 8, but they do not establish that degeneracy increases with network size, and no trend beyond N = 12 is visible.

N	Functional	Networks	Spearman (exact, sampled)	Mean relative error of sampled	Max. abs. relative error	Median normalized score DNA/GA/ES, sampled	Median normalized score DNA/GA/ES, exact
12	CN	220	0.997	+0.0%	4.0%	0.67/0.67/0.79	0.67/0.66/0.80
12	PHI	220	0.965	+13.5%	48.8%	0.68/0.66/0.80	0.66/0.64/0.78
16	CN	220	0.999	+0.1%	3.9%	0.61/0.59/0.81	0.61/0.60/0.81
16	PHI	220	0.981	+15.1%	34.8%	0.65/0.62/0.82	0.65/0.62/0.78

Table 10. *Validation of the sampled functionals against the exhaustive computation at N = 12 and 16 (final networks of 40 runs per route and 100 single random networks).*

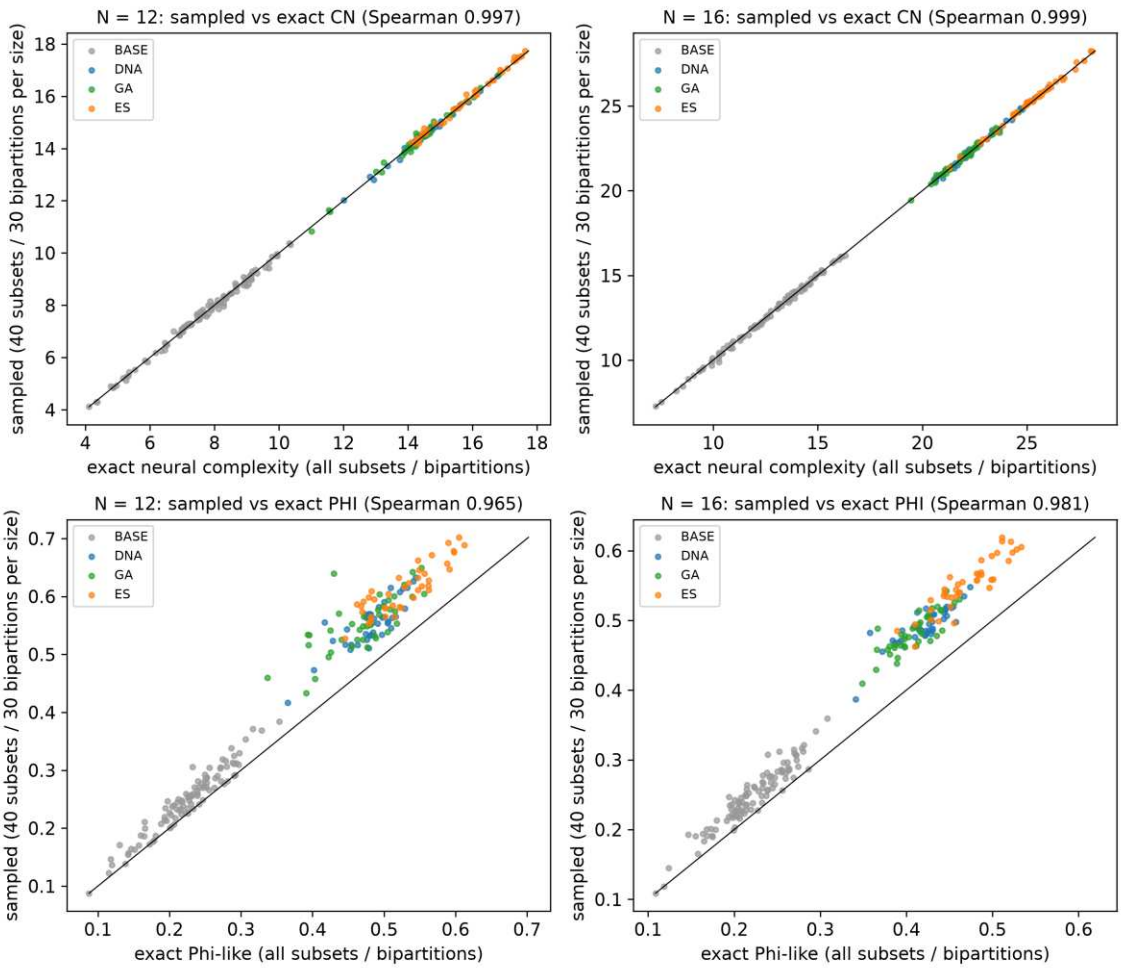


Fig. 5. *Exact against sampled scores at N = 12 and 16 for neural complexity and the Phi-like measure. Each point is one network.*

Validation of the sampling approximation. With respect to neural complexity, the sampled values agreed closely with the exhaustive values (Spearman 0.997 at N = 12 and 0.999 at N = 16, mean relative errors of 0.0% and 0.1%, and maximum values of 4.0% and 3.9%, respectively; Table 10, Fig. 5). The sampled Phi-like measure was biased upward, as expected for a minimum over a subset of bipartitions: the mean relative error was 13.5% at N = 12 and 15.1% at N = 16 (maximum 49% and 35%), but ranked the networks nearly identically (Spearman 0.965 and 0.981). Route medians computed with the exhaustive functional preserved the order ES > DNA ≈ GA; for the Phi-like measure at N = 16, the median ES score fell from 0.82 to 0.78 when the exhaustive functional was used, a sign that ES partly exploits the sampled bipartitions. The N = 12 and 16 results for the Phi-like measure should therefore be read as upper-bound approximations, and those for neural complexity should be read as effectively exact.

4.6 Maintenance, damage and repair

k	rho=0.1	rho=0.25	rho=0.5	rho=1	rho=2	rho=4	rho=8	rho=16
2	0.54	0.50	0.49	0.46	0.36	0.20	0.11	0.05
3	0.47	0.46	0.51	0.46	0.44	0.26	0.15	0.06
5	0.48	0.46	0.48	0.46	0.41	0.32	0.18	0.07

Table 11. *Maintained order I (mean over 10 runs) against erosion ρ = μL for tournament sizes k.*

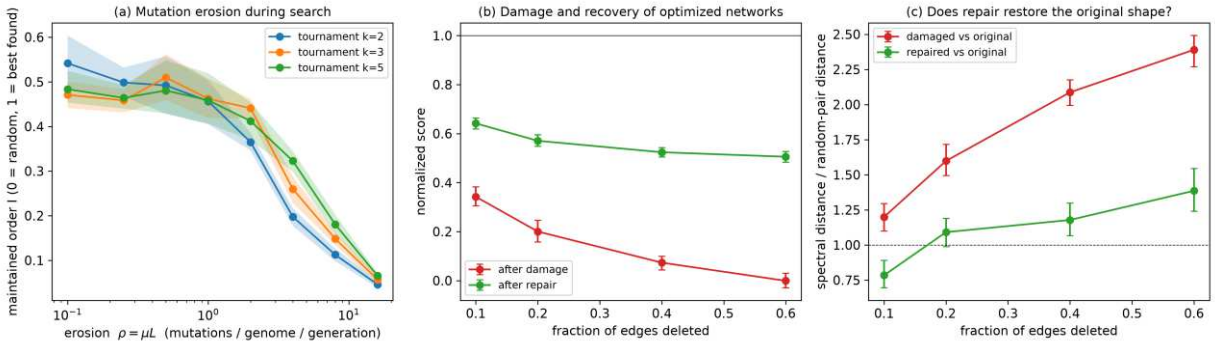


Fig. 6. *(a) Maintained order against mutational erosion ρ = μL for three tournament sizes (bands: bootstrap 95% CI over runs). (b) Normalized score of optimized networks after edge deletion and after repair. (c) Spectral distance to the original network, in units of the typical distance between random networks. Error bars in (b) and (c): bootstrap 95% CI over 30 networks.*

The order maintained during search fell smoothly with erosion (Table 11, Fig. 6). It ranged from approximately 0.46 to 0.54 for $\rho \leq 1$ and decreased to 0.20–0.32 at $\rho = 4$ and to 0.05–0.07 at $\rho = 16$, which is near the level of a random network. We found no sharp error threshold at this resolution. Stronger selection gave a modestly higher order at $\rho = 4$–16 (0.32 for k = 5 against 0.20 for k = 2 at $\rho = 4$). The plateau remained well below 1 because each run used a limited budget of 1,200 evaluations.

Damage to an already optimized network is a different process from mutation during search. Optimized networks were fragile: deleting 10% of the edges (approximately three of 28) lowered the normalized score from at least 0.8 to 0.34, and 40% deletion removed nearly all the advantages (Table 12, Fig. 6). Repair with 300 evaluations restored the score only partially (0.64 after a 10% deletion, 0.51 to 0.52 after 40–60%). The repaired networks did not return to the original architecture: their spectral distance to the original was 0.78 times the typical random-pair distance after 10% deletion and 1.09 to 1.39 times after larger deletions (Fig. 6). Functions can thus be partly recovered through architectures that differ from the original ones.

Edges deleted	n	Score after damage	Score after repair	Distance damaged-original	Distance repaired-original
0.1	30	0.34	0.64	1.20	0.78
0.2	30	0.20	0.57	1.60	1.09
0.4	30	0.07	0.52	2.09	1.18
0.6	30	0.00	0.51	2.39	1.39

Table 12. *Damage and repair of ES-optimized networks (30 networks; two damage draws averaged per network). Distances are spectral, in units of the typical random-pair distance.*

4.7 Sensitivity and uncertainty propagation

Across 17 one-at-a-time settings (12 runs each), the enrichment E = median(DNA)/median(random) for raw neural complexity ranged from 1.01 to 1.25, and the spectral ratio D ranged from 0.78 to 1.35, straddling 1 (Fig. 7). In the Monte Carlo propagation over parameter priors (40 draws, 3 replicates each), E for DNA had a median of 1.16 (95% interval 0.99 to 1.27) and exceeded 1 in 38 of 40 draws; E for ES had a median of 1.29 (1.11 to 1.36) and exceeded 1 in 40 of 40 (Fig. 8). A paired Wilcoxon test of DNA against random search over the 40 parameter draws gave p = 1e-11; because draws, not independent datasets, are the unit, we use it as a descriptive check and not as the principal inference. D had a median of 1.06 (0.74 to 1.65) and was less than 1 in 14 of 40 draws. The enrichment is therefore reliable but modest; the structural ratio is not distinguishable from 1.

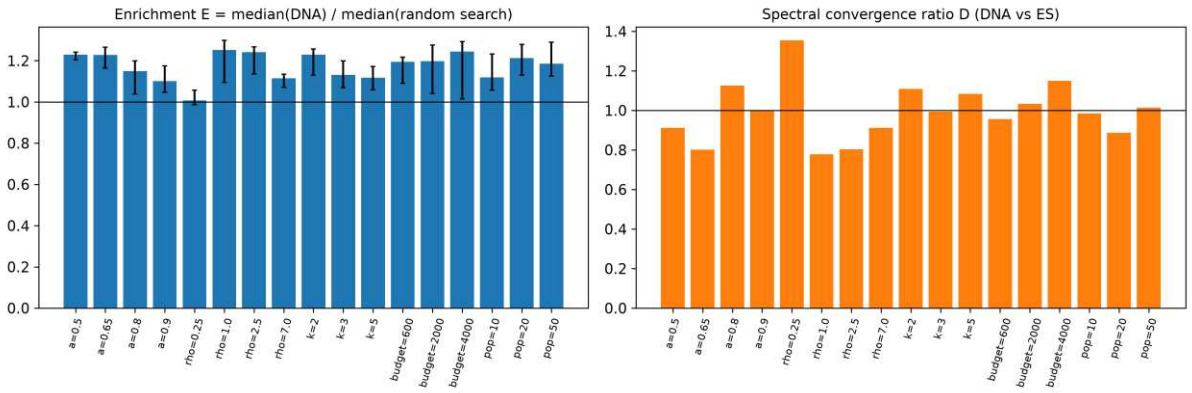


Fig. 7. *One-at-a-time sensitivity: enrichment E (left, with bootstrap 95% CI over DNA runs) and spectral convergence ratio D (right).*

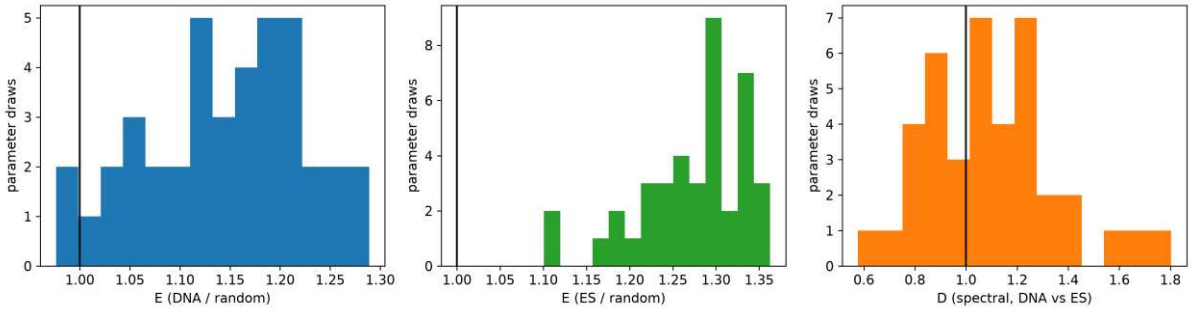


Fig. 8. *Monte Carlo propagation over parameter priors (40 draws): enrichment of DNA and ES over random search, and spectral convergence ratio D.*

4.8 Reproduction of the earlier draft

Before any change, the version-2 scripts supplied with the earlier draft were run unchanged in an empty directory. They regenerated 346 of the 362 summary numbers and 7 of the 8 tables identically, including the performance, contrast, convergence, dynamics, scaling and maintenance results. The remaining 16 numbers, all in the damage-and-repair experiment, differed from the shipped values by up to 0.04 in normalized scores and 0.15 in distance ratios (for example, the score after 10% deletion was 0.34 against 0.32, and the repaired score was 0.64 against 0.63). We did not trace the cause, which lies in the random stream of that stage; none of the conclusions depend on it. The values reported in this paper are those of the version-3 code, which also reproduces the clean rerun.

5 Discussion

The central result is a dissociation. Under a fixed budget, independent routes reach similar high values of structural-information functionals, yet the networks they find do not share one architecture by the spectral, motif and modularity descriptors, and the apparent convergence of the profile and response descriptors is neither robust to the dynamics nor free of circularity. This finding is in line with the degeneracy reported in biological circuits (Prinz et al. 2004) and suggests that, in this model, the functional landscape contains many distinct near-optimal solutions. The evidence that the number of such solutions increases with network size is weak (Section 4.5).

On the question of encoding versus algorithm, the data support a narrower statement than the earlier draft made. The DNA route and the continuous GA were not statistically distinguishable under the tested conditions, but with 40 runs, the interval for neural complexity still allows a difference on the order of 0.2; thus, the encodings are not shown to be equivalent. The finding that ES outperforms the continuous GA indicates that the optimization operators contribute to the gap between DNA and ES, without isolating which operator. A design that varies one operator at a time or tunes each route is required to separate them.

Four further points qualify the conclusion. First, a descriptor must be sensitive enough to detect differences and robust enough not to overdetect them. The spectrum is very sensitive: deleting three of 28 edges moved a network 1.20 times the typical random-pair distance. Nonconvergence in the spectrum therefore does not by itself show that networks implement different mechanisms. The perturbation responses computed from the simulated trajectories are closer to the function; they showed at most modest convergence and a different sign for the two functionals, so the same caution applies. Second, the optima are narrow: single random networks essentially never reach half of the best normalized score for neural complexity or the Phi-like measure, although a search of 2,000 draws finds such networks in a quarter of its runs, and optimized networks lose most of their advantage after small edge deletions. This is in tension with the robustness often attributed to biological circuits, and it suggests that selection on a single functional with no robustness pressure yields brittle solutions such that adding such pressure is a natural extension (Wagner 2005). Third, normalized scores depend on the best value found in each experiment; only within-experiment contrasts are interpreted. Fourth, the Phi-like values at $N \geq 12$ rest on sampled bipartitions that overestimate the minimum by approximately 13–15% on average, although they rank networks nearly identically to the exhaustive computation.

6 Limitations and future work

• The functionals are proxies for integration and differentiation. The Phi-like measure is not integrated information, and at $N \geq 12$, the sampled subsets and bipartitions make it approximate (Section 4.5).

• Symmetric weights, $N \leq 16$, and low-dimensional Gaussian or tanh dynamics are far from real directed, embodied, developing neural circuits.

• The nonlinear model uses a single fixed noise realization per experiment (common random numbers) and a smaller budget (1,000). The perturbation response is computed from the same realization, with three pulse sizes.

• The spectrum-based descriptors are sensitive to small edge changes, and the integration profile is partly built into neural complexity; the linear-response descriptor is a proxy for the dynamics.

• The high-scoring sets at $N = 12$ and 16 contain 23–28 networks, so the diversity ratios have wide intervals; the 15 route contrasts are not corrected for multiplicity.

• Hyperparameters were not tuned per route, so the comparison between routes is at a fixed, untuned configuration. The damage-repair experiment uses a single repair method and budget.

• The code was developed with AI assistance. It was reproduced from the supplied scripts (Section 4.8) but has not been independently audited.

Future work should (i) add robustness pressure (noise or damage during fitness evaluation) and test whether it narrows the equivalence set; (ii) use directed weights and spiking or conductance-based dynamics with perturbational measures computed from simulated stimulation; (iii) test whether the dissociation holds for

empirical connectomes as starting points; (iv) replace the spectral descriptor with graph-matching or learned embeddings; and (v) vary one operator at a time between the DNA, GA and ES routes.

7 Possible implications for theories of consciousness

Theories that tie consciousness to network structure, such as integrated information theory (Tononi 2004) and perturbational complexity indices (Casali et al. 2013), predict that systems with the right structure should share the relevant property. Our results are compatible with a family of architectures sharing a functional value and do not support a single privileged shape; they cannot say whether any such value corresponds to consciousness because that assumption is not tested here. The Phi-like measure and neural complexity are proxies, and the perturbation response used in Section 4.4 is a simple simulated analog, not a perturbational complexity index. Empirical work on real systems is needed.

8 Conclusion

In this controlled Monte Carlo study, independent evolutionary and evolution-strategy routes revealed networks with high values of three information-theoretic functionals without converging on a common architecture, and the degree of apparent convergence depended on the descriptor, the functional and the dynamics. DNA and continuous-GA routes were not statistically distinguishable under the tested conditions, which does not establish equivalent encodings; ES outperformed the continuous GA, so the optimization operators contributed. The perturbation responses from the simulated nonlinear trajectories gave mixed and at most modest convergence. Optimized networks were brittle under damage and recovered function only partially through architectures that differed from the originals. This study establishes functional equivalence without structural convergence only for the proxies and models tested.

Statements and Declarations

Funding. The author received no specific funding for this work.

Competing interests. The author has no competing interests to declare that are relevant to the content of this article.

Author contributions. Leon Sandler is the sole author. He conceived the study; specified the models, functionals and experiments; directed the analyses; checked the results and the references; and reviewed, edited and approved the final manuscript.

Ethics approval and consent to participate. Not applicable: This is a computational study with no human participants, animals or personal data.

Use of AI tools. An AI assistant (Claude, Anthropic) was used under the author's direction to help write and test the simulation code, to run the analyses and to draft and revise the text (Section 3.8). The author reviewed the outputs and takes full responsibility for the content.

Data and code availability. All code, configuration, raw result summaries (JSON and CSV), figures and the manuscript source are available at https://github.com/sandlerleon/shapesearch-functional-equivalence and archived on Zenodo (software, DOI 10.5281/zenodo.23862548; preprint, DOI 10.5281/zenodo.23862546). Fixed random seeds regenerate every table and figure; the full v3 run takes about half an hour on a single CPU core

(Python 3.11.9, NumPy 2.4.6, SciPy 1.17.1, Matplotlib 3.11.1). Raw pickled result objects are not distributed; run_experiments.py regenerates them.

References

Balduzzi D, Tononi G (2008) Integrated information in discrete dynamical systems: motivation and theoretical framework. PLoS Comput Biol 4:e1000091. https://doi.org/10.1371/journal.pcbi.1000091

Barrett AB, Seth AK (2011) Practical measures of integrated information for time-series data. PLoS Comput Biol 7:e1001052. https://doi.org/10.1371/journal.pcbi.1001052

Beyer H-G, Schwefel H-P (2002) Evolution strategies – a comprehensive introduction. Nat Comput 1:3–52. https://doi.org/10.1023/A:1015059928466

Casali AG, Gosseries O, Rosanova M, Boly M, Sarasso S, Casali KR, Casarotto S, Bruno M-A, Laureys S, Tononi G, Massimini M (2013) A theoretically based index of consciousness independent of sensory processing and behavior. Sci Transl Med 5:198ra105. https://doi.org/10.1126/scitranslmed.3006294

Cliff N (1993) Dominance statistics: ordinal analyses to answer ordinal questions. Psychol Bull 114:494–509. https://doi.org/10.1037/0033-2909.114.3.494

Edelman GM, Gally JA (2001) Degeneracy and complexity in biological systems. Proc Natl Acad Sci USA 98:13763–13768. https://doi.org/10.1073/pnas.231499798

Efron B (1979) Bootstrap methods: another look at the jackknife. Ann Stat 7:1–26. https://doi.org/10.1214/aos/1176344552

Mann HB, Whitney DR (1947) On a test of whether one of two random variables is stochastically larger than the other. Ann Math Stat 18:50–60. https://doi.org/10.1214/aoms/1177730491

Marder E, Goaillard J-M (2006) Variability, compensation and homeostasis in neuron and network function. Nat Rev Neurosci 7:563–574. https://doi.org/10.1038/nrn1949

Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U (2002) Network motifs: simple building blocks of complex networks. Science 298:824–827. https://doi.org/10.1126/science.298.5594.824

Newman MEJ (2006) Modularity and community structure in networks. Proc Natl Acad Sci USA 103:8577–8582. https://doi.org/10.1073/pnas.0601602103

Prinz AA, Bucher D, Marder E (2004) Similar network activity from disparate circuit parameters. Nat Neurosci 7:1345–1352. https://doi.org/10.1038/nn1352

Rosas FE, Mediano PAM, Gastpar M, Jensen HJ (2019) Quantifying high-order interdependencies via multivariate extensions of the mutual information. Phys Rev E 100:032305. https://doi.org/10.1103/PhysRevE.100.032305

Tononi G, Sporns O, Edelman GM (1994) A measure for brain complexity: relating functional segregation and integration in the nervous system. Proc Natl Acad Sci USA 91:5033–5037. https://doi.org/10.1073/pnas.91.11.5033

Tononi G (2004) An information integration theory of consciousness. BMC Neurosci 5:42. https://doi.org/10.1186/1471-2202-5-42

Wagner A (2005) Robustness and evolvability in living systems. Princeton University Press, Princeton