Evolution With Purpose:
Hierarchy-Informed Optimization of Whole-Brain Models

Hormoz Shahrzad
hormoz@cs.utexas.edu
The University of Texas at Austin;
Stanford University; Cognizant AI Lab

Niharika Gajawell
gajawell@stanford.edu
Stanford University

Kaitlin Maile*
kmaile@google.com
IRIT, University of Toulouse

Manish Saggar
saggar@stanford.edu
Stanford University

Risto Miikkulainen
risto@cs.utexas.edu
The University of Texas at Austin;
Cognizant AI Lab

Abstract

Evolutionary search is well suited for large-scale biophysical brain modeling, where many parameters with nonlinear interactions and no tractable gradients need to be optimized. Standard evolutionary approaches achieve an excellent fit to MRI data; however, among many possible such solutions, it finds ones that overfit to individual subjects and provide limited predictive power. This paper investigates whether guiding evolution with biological knowledge can help. Focusing on whole-brain Dynamic Mean Field (DMF) models, a baseline where 20 parameters were shared across the brain was compared against a heterogeneous formulation where different sets of 20 parameters were used for the seven canonical brain regions. The heterogeneous model was optimized using four strategies: optimizing all parameters at once, a curricular approach following the hierarchy of brain networks (HICO), a reversed curricular approach, and a randomly shuffled curricular approach. While all heterogeneous strategies fit the data well, only curricular approaches generalized to new subjects. Most importantly, only HICO made it possible to use the parameter sets to predict the subjects' behavioral abilities as well. Thus, by guiding evolution with biological knowledge about the hierarchy of brain regions, HICO demonstrated how domain knowledge can be harnessed to serve the purpose of optimization in real-world domains.

CCS Concepts

• **Computing methodologies** → **Genetic algorithms**; *Bio-inspired approaches*; *Distributed algorithms*; • **Mathematics of computing** → **Optimization algorithms**.

Keywords

Evolutionary Computation, Curriculum Learning, Dynamic Mean Field Models, Whole-Brain Modeling, Cross-Subject Generalization, Lyapunov Stability, Human Connectome Project

1 Introduction

Large-scale functional brain dynamics are commonly modeled as emergent phenomena arising from the interaction between anatomical connectivity and local neural physiology. Models such as neural mass, neural field, and Dynamic Mean Field (DMF) successfully reproduced empirical functional connectivity and dynamics at the level of resting-state networks (RSN)s [3, 6, 15, 19]. These models depend on many interacting hyperparameters that must be carefully tuned to match individual neuroimaging data. Evolutionary algorithms are well suited for this purpose because they explore complex, non-differentiable fitness landscapes without gradient information [12, 22, 25]. Prior work applying evolutionary strategies to DMF models demonstrated improved fits to resting-state data [19], but also highlighted that such fits often fail to generalize across subjects and may not predict behavior well. Similar issues of deception and overfitting arise in other real-world search domains. Although evolutionary computation (EC) techniques are well suited for large, high-dimensional, and deceptive search spaces—even at extreme scales involving thousands of states and billions of variables [5, 11, 22, 28, 29]—they can still converge to brittle or uninformative solutions in the absence of appropriate inductive biases.

This work investigates a previously underexplored factor in evolutionary optimization of dynamical systems: the temporal order in which parameters are introduced during search. Drawing on cortical hierarchies, established in humans and primates as a principal gradient of cortical organization [20], and incremental optimization insights [11], heterogeneous optimization was compared with hierarchy-informed curriculum (HICO), reverse-phased curriculum, and shuffled curriculum. Although introducing heterogeneity substantially improved in-sample fitness, only flat (non-curricular) heterogeneous, HICO, and reverse-phased strategies significantly outperformed the homogeneous baseline ($p \approx 10^{-23}$–10^{-25}). Marked differences emerged, however, under leave-one-out cross-subject evaluation using cohort-averaged parameters: flat heterogeneous and homogeneous baselines failed to generalize, whereas curriculum-based strategies remained robust. Among these, HICO yielded the most consistent and well-structured parameter solutions. Moreover, models based on HICO provided the strongest and most reliable predictions of cognitive and psychopathology targets spanning multiple domains, including fluid reasoning ability (as measured using the Penn Matrix Test [23]), behavioral problems related to internalization and externalization, as measured using the Achenbach Adult Self Report [2]. These results indicate that incorporating domain-informed structure through evolutionary curricula fundamentally alters search dynamics, providing a principled mechanism for scaling high-dimensional brain

---
*Now at Google, Paradigms of Intelligence Team

models while yielding more stable, generalizable, and behaviorally informative solutions than flat optimization.

2 Background

The following background situates the proposed HICO within prior work on whole-brain modeling and evolutionary search.

2.1 Whole-Brain Biophysical Models

A central goal of biophysical brain modeling is to explain how distributed functional dynamics emerge from the coupling of structural connectivity and local neural processes [3, 6]. Popular frameworks include neural mass, neural field, and Dynamic Mean Field (DMF) approaches. The resulting models can generate resting-state activity that reproduces key empirical signatures, including functional connectivity, dynamic functional connectivity, and metastability [4, 7].

A canonical study by Honey et al. [15] showed that structural connectivity shapes—but does not fully determine—functional connectivity, motivating the use of computational models to explore this mapping. Subsequent work has characterized how factors such as excitation–inhibition balance, global coupling strength, conduction delays, and noise influence emergent functional patterns [3, 6]. The DMF model used in this paper is one such framework, balancing biological plausibility with computational tractability [8].

2.2 Challenges in Parameter Optimization

Increasing biological realism in large-scale brain models comes at the cost of sharply increased number of parameters. Strong nonlinear interactions between parameters give rise to complex, multimodal, and highly non-convex optimization landscapes. As a result, exhaustive search is infeasible, and gradient-based methods are often unsuitable due to non-differentiability, stochasticity, and numerical instability.

A central challenge is *degeneracy*: many distinct parameter configurations may yield similar emergent dynamics [1]. While such degeneracy reflects biological robustness, it complicates inference and increases the risk of overfitting idiosyncratic subject-level patterns. Steering optimization toward solutions that generalize across subjects and remain dynamically stable is therefore an important open research question.

2.3 Cortical Hierarchy and Training Curricula

The brain is typically parcellated into canonical RSNs) Yeo et al. [35]. More recently, Margulies et al. [20] identified a dominant gradient ranging from unimodal (sensorimotor) to transmodal (default mode) RSNs, thereby providing a macroscale gradient of cortical organization.

We hypothesize that this organization of RSNs provide a natural basis for structured optimization. Thus, suggesting that parameters governing sensory-motor dynamics may be more foundational than those governing higher-order associative systems, or vice-versa. Motivated by this hierarchy, it is possible to construct phased and reverse-phased optimization curricula that introduce RSN-specific parameters in different temporal orders. This design makes it possible to test whether the order of parameter introduction acts as an inductive bias that shapes generalization and cross-subject stability.

Overall, HICO can be positioned as a domain-structured curriculum for EC that is distinct from (i) complexity-growth curricula (e.g., NEAT-style complexification), and (ii) permanent variable decomposition (e.g., cooperative coevolution). Its novelty is to use a biologically grounded hierarchy to define a *phase ordering* over pre-existing parameter blocks in a whole-brain dynamical model, in order to generalize across subjects better and predict behavior more accurately, rather than simply fit to the data.

2.4 Curricula and Staged Search in Evolutionary Computation

The use of *curricula* in EC often refers to deliberately shaping the search process over time so that later stages become feasible only after earlier structure has been established. One prominent line is *incremental evolution* or *complexification*, where problem difficulty or model complexity is increased gradually, e.g. by adding degrees of freedom, structure, or task demands [31]. The idea is to help evolution avoid brittle local optima early on and discover scalable representations. Related ideas appear in staged or progressive training regimes in evolutionary robotics and neuroevolution, where simpler components are learned before integrating higher-level behaviors, effectively regularizing the search trajectory through time [11].

HICO differs from such approaches in an important way: rather than increasing task difficulty or expanding representation size, it imposes a *domain-derived ordering* over parameter blocks that already exist in the model. In whole-brain DMF models, parameters are naturally grouped by RSNs and linked to known cortical gradients; HICO leverages this structure by prioritizing higher-level organizational degrees of freedom before specializing RSN-specific parameters. This makes HICO closer to *structured staged optimization* than to classic "easy-to-hard" curricula.

2.5 Decomposition, Cooperative Coevolution, and Multiobjective Optimization

Another major family of EC metaheuristics addresses high dimensional search by decomposing variables into subcomponents and optimizing them with limited coordination, most notably *cooperative coevolution* [26]. Cooperative coevolution and its many variants explicitly model the fact that different subsets of parameters may interact at different strengths, and they often rely on collaboration schemes to assemble partial solutions into a full candidate. Modern surveys emphasize decomposition design and variable-interaction structure learning as central to performance in high-dimensional domains [18].

HICO is conceptually aligned with this literature in that it treats the parameter vector as *structured*, but it differs operationally: rather than permanently decomposing the problem into concurrently evolving subpopulations, HICO introduces a *temporal decomposition* (a curriculum) that controls when each RSN-specific block is allowed to vary. The curriculum preserves a single coherent model throughout optimization while still exploiting modular structure to reduce destructive interference between strongly coupled parameter groups.

One way to obtain solutions that not only are a good fit with data but also generalize and predict performance would be to use

measurements of those qualities as secondary objectives. However, those objectives may not be known ahead of time. The model may be used later to diagnose disorders, predict measures of cognitive abilities, or propensity for psychopathology. The goal of HICO is to guide evolution towards useful models regardless of what other metrics may later be of interest. The rationale for not adopting a multiobjective formulation in this setting is discussed in Appendix A.5, where the distinction between model fitting and downstream evaluation is clarified.

2.6 Evolutionary Algorithms in Neuroscience

Evolutionary algorithms (EAs) maintain a population of candidate solutions that evolve under selection, mutation, and recombination [14]. They are especially effective for high-dimensional, non-differentiable problems with complex fitness landscapes. Early applications demonstrated their value in fitting single-neuron conductance parameters [17, 33]. Achard and De Schutter [1] used evolutionary optimization to fit detailed Purkinje cell models, revealing high-dimensional manifolds of viable solutions. Druckmann et al. [9] applied Multi-objective evolutionary algorithms to match multiple electrophysiological features simultaneously.

At larger scales, population-based evolutionary optimization has been used to tune RSNs and whole-brain models [19, 32]. While covariance-based strategies such as CMA-ES can be effective in moderate-dimensional settings [12], they scale poorly as the number of parameters grows and impose strong assumptions on parameter correlations. In contrast, genetic-algorithm–style evolutionary computation offers greater scalability and architectural flexibility, making it well suited for high-dimensional whole-brain models and curriculum-guided parameter schedules. These insights motivate the present focus on introducing structured inductive biases into evolutionary optimization.

3 Method

This section describes the data, modeling framework, optimization objectives, and evolutionary strategies used to compare homogeneous and heterogeneous whole-brain DMF models under different optimization curricula.
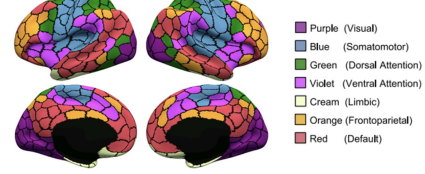
3.1 Data and Parcellation

Resting-state functional Magnetic Resonance Imaging (rs-fMRI) and diffusion-weighted imaging (DWI) data were obtained from 100 unrelated subjects from the Human Connectome Project (HCP). The cerebral cortex was parcellated into 400 regions using the Schaefer et al. [27] atlas, with parcels assigned to seven canonical RSNs according to the Yeo et al. [35] parcellation (Fig. 1). Subject-specific structural connectivity (SC) matrices were derived from tractography of the DWI data, and empirical functional connectivity (FC) matrices were computed as Pearson correlations of regional Blood-Oxygen-Level–Dependent (BOLD) rs-fMRI time series

3.2 Dynamic Mean Field Model

Each cortical region was modeled as a coupled excitatory–inhibitory neural mass using the Dynamic Mean Field (DMF) framework [3, 6, 7]. The DMF model provides a mesoscale approximation to large populations of spiking neurons; it was shown to reproduce

[a] Traditional parcellation of the brain regions into RSNs

Purple (Visual)
Blue (Somatomotor)
Green (Dorsal Attention)
Violet (Ventral Attention)
Cream (Limbic)
Orange (Frontoparietal)
Red (Default)

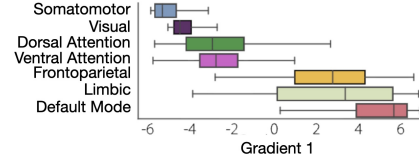[b] Macroscale gradient of cortical organization

Figure 1: Large-scale cortical organization used to define RSN-specific parameter blocks. (a) Surface renderings of a 400-region cortical parcellation mapped to the seven canonical RSNs of Yeo et al. [35], shown for lateral and medial views of both hemispheres. Colors denote Visual, Somatomotor, Dorsal Attention, Ventral Attention, Limbic, Frontoparietal Control, and Default Mode RSNs. (b) Distribution of parcels along the principal macroscale gradient of cortical organization (Gradient 1), ordered from unimodal sensory–motor systems to transmodal association cortex. This gradient defines the hierarchical ordering used to construct curriculum phases in the proposed optimization framework.

key properties of resting-state dynamics when driven by subject-specific structural connectivity [7, 24].

Let $S_i(t)$ denote the excitatory synaptic gating variable for region i. The excitatory population dynamics are governed by

$$\tau_E \frac{dS_i}{dt} = -S_i + (1 - S_i)\,\gamma\,H(I_i(t)), \tag{1}$$

where τ_E is the excitatory time constant, γ controls synaptic gain, and $H(\cdot)$ is a sigmoidal firing-rate function. Long-range coupling between regions is defined as

$$I_i(t) = I_b + w_E S_i(t) - w_I S_i^I(t) + G \sum_j \mathrm{SC}_{ij} S_j(t - d_{ij}), \tag{2}$$

where I_b is a background current, w_E and w_I denote local excitatory and inhibitory coupling strengths, $S_i^I(t)$ is the inhibitory gating variable, G is the global coupling parameter, SC_{ij} is the structural connectivity weight from region j to i, and d_{ij} represents conduction delays. Canonical DMF constants were adopted from prior work [7], except for the parameters optimized in this study. See Appendix A.1 for the full specification of DMF variables and constants. The homogeneous model comprised 20 shared parameters applied uniformly across all regions. The heterogeneous formulation comprised 140 parameters, corresponding to seven RSNs with independent copies of the same 20 parameters. See Appendix A.2 for the full parameter specification. In both cases, the DMF equations define a stochastic dynamical system that generates region-wise neural activity time series whose statistical structure depends on the parameter vector Θ.

3.3 Simulation and Objective Function

For a candidate parameter vector Θ, the DMF system defined above was simulated for five minutes of biological time using an Euler–Maruyama integration scheme [13] with a step size of 0.1 ms. The resulting excitatory synaptic activity $S_i(t)$ was converted into BOLD

signals using the Balloon–Windkessel hemodynamic model [10]. Functional connectivity matrices $\widehat{FC}(\Theta)$ were then computed from the simulated BOLD time series by taking pairwise Pearson correlations between all cortical regions, in direct analogy to the empirical FC derived from resting-state fMRI. See Appendix A.3 for implementation details. The DMF equations thus induced, via simulation and hemodynamic transformation, a mapping

$$\Theta \;\longmapsto\; \widehat{FC}(\Theta),$$

from model parameters to predicted functional connectivity. Parameter fitting was formulated as an inverse problem: finding Θ such that the FC generated by the DMF dynamics matches the empirical FC as closely as possible.

The optimization objective minimized the discrepancy between empirical and simulated FC:

$$\mathcal{L}(\Theta) = 1 - \mathrm{corr}\!\left(\mathrm{vec}(FC),\ \mathrm{vec}\!\left(\widehat{FC}(\Theta)\right)\right), \tag{3}$$

where $\mathrm{vec}(\cdot)$ extracts the upper triangular elements of the FC matrices. Equation (3) directly links the DMF dynamics to the optimization criterion: the DMF equations generate neural activity and BOLD time series, these time series define $\widehat{FC}(\Theta)$, and the loss function measures how well the resulting large-scale correlation structure matches empirical resting-state connectivity. See Appendix A.4 for details of the fitness computation.

This FC-based objective follows standard practice in whole-brain modeling, where functional connectivity serves as a compact summary of long-range statistical dependencies induced by the underlying neural dynamics [7, 19, 24]. Minimizing $\mathcal{L}(\Theta)$ therefore corresponds to calibrating the biophysical parameters of the DMF model so that its emergent RSN-level behavior reproduces the observed functional organization of the human brain.

3.4 Optimization Curricula

This study compares multiple evolutionary optimization curricula that differ only in the temporal ordering by which parameter subsets are optimized during search.

Homogeneous Baseline. A single set of 20 parameters was used for all RSNs. Thus, this baseline assumes that biophysical dynamics is homogeneous throughout the brain.

Heterogeneous Optimization. All 140 RSN-specific parameters (20 parameters for each of the seven RSNs) were optimized simultaneously from the outset, without any curriculum structure. This approach maximizes search dimensionality at all times and serves as a flat baseline for assessing the impact of curriculum structure.

Hierarchy-Informed Curriculum Optimization (HICO). Guided by the principal cortical gradient described by Margulies et al. [20], parameters were optimized sequentially according to the following phases:

- **Phase I — Global Phase.** This phase establishes the homogeneous baseline as the starting point. Further optimizaiton is thus restricted to a low-dimensional, dynamically coherent subspace, which reduces the unstable fixed points and improves Lyapunov stability [7].
- **Phase II — Transmodal core (Default Mode and Limbic RSNs).**

- **Phase III — Frontoparietal control RSN.**
- **Phase IV — Attention RSNs (DorsAttn and SalVentAttn).**
- **Phase V — Visual RSN (VIS).**
- **Phase VI — Somatomotor RSN (SomMot).**

Within each phase, only the parameters associated with the active hierarchical segment were allowed to vary, while parameters optimized in earlier phases were held fixed. The total generation budget was divided evenly across phases, such that all curricula were matched for overall computational cost.

Reverse-Phased Curriculum. The reverse-phased curriculum retained an identical Phase I (global parameters) but executed the remaining five hierarchy-specific phases in reverse order, beginning with the transmodal core and progressing toward unimodal sensory systems. Thus, this curriculum tested whether the direction of hierarchical parameter release influences optimization outcomes.

Shuffled Curriculum. In the shuffled curriculum, Phase I (global parameters) was again identical to HICO, while the remaining five hierarchy-specific phases were randomly permuted for each subject. This control preserved the phased structure of optimization while removing any alignment with the cortical hierarchy.

3.5 Evolutionary Optimization Method

Model parameters were optimized using an elitist genetic algorithm (GA). Each candidate solution is a fixed-length parameter vector: $d = 20$ for the homogeneous model and $d = 140$ for heterogeneous models.

To ensure comparable search resolution across parameters with different physical units, each parameter was represented in a normalized coordinate in $[0, 1]$ with 1000-step precision and mapped via an affine transform to its literature-supported valid range, as established in prior DMF modeling studies [3, 7, 34]. This representation yields an effective discrete search space of 10^{60} possible configurations for the homogeneous model and 10^{420} for the heterogeneous formulation. This scale makes the evolutionary approach a good choice for optimizing these models.

Although Eq. (3) is written as a loss function, evolutionary optimization operated by *maximizing* the corresponding fitness, defined as the Pearson correlation between empirical and simulated functional connectivity matrices (i.e., $1 - \mathcal{L}(\Theta)$). Throughout this work, the terms *fitness* and *score* refer to this correlation value, with higher scores indicating closer agreement between simulated and empirical FC.

Across all optimization curricula, the GA configuration was held constant as

- **Population size:** $N = 100$ individuals per generation.
- **Elitism:** the top $E = 20$ individuals (highest fitness) were carried forward unchanged to the next generation.
- **Parent selection:** tournament selection with tournament size $k = 3$.
- **Variation operators:** uniform crossover and per-gene mutation.
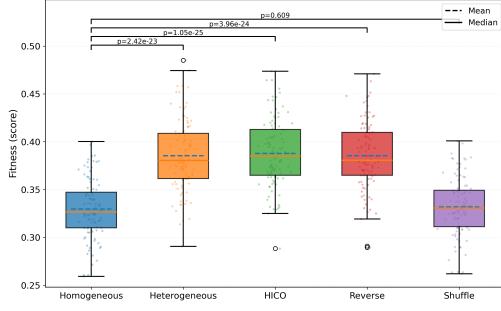- **Mutation rate:** $p_{\mathrm{mut}} = 0.1$ (probability of mutating each gene).

Figure 2: Distribution of fitness scores across optimization strategies when models were optimized separately for each individual subject. The points represent subjects and the boxes show the interquartile range (IQR) with median (solid line) and mean (dashed line). Unpaired statistical tests against the homogeneous baseline indicate that the heterogeneous, HICO, and reverse curricula achieve significantly higher fitness, demonstrating the advantage of increased model dimensionality. In contrast, the shuffled curriculum fails to realize this benefit, suggesting that the curriculum needs to be systematic to take advantage of such dimensionality.

- **Generation budget:** a fixed total number of generations $G_{\text{total}} = 120$ was used for all methods to ensure compute-matched comparisons.

For curriculum-based strategies, G_{total} was divided evenly across six hierarchy-defined phases (as was described in Section 3.4). Within each phase, only the parameter subset corresponding to the active RSN was allowed to vary, while parameters optimized in prior phases were held fixed. This schedule yielded an equal overall computational budget while controlling the effective search dimensionality over time.

All experiments were parallelized using a hub-and-spoke distributed evolutionary architecture with four concurrent evolutionary engines [30]. In this setup, multiple worker spokes evaluate candidate solutions in parallel and synchronize with a central hub that maintains population state and applies selection and variation. This design increases throughput without changing the underlying optimization dynamics, and it ensures identical GA settings across homogeneous and heterogeneous curricula.

4 Results

This section evaluates how different evolutionary optimization strategies influence model fit to individual subjects, cross-subject generalization, parameter geometry, and behavioral relevance in whole-brain DMF models. First, the approaches are evaluated in fitting to individual subjects in comparison against a homogeneous baseline. Second, generalization to new subjects is tested in leave-one-out (LOO) experiments. Third, to elucidate the mechanisms underlying observed failures and successes, the geometry of learned parameter distributions is examined using low-dimensional embeddings. Fourth, behavioral relevance of optimized parameters is assessed through RSN-level prediction analysis, and fifth, through permutation-calibrated statistical testing. Together, the results demonstrate that while increased model dimensionality improves within-subject fit, the *temporal structure* imposed by HICO is critical for achieving stable, generalizable, and behaviorally informative solutions.
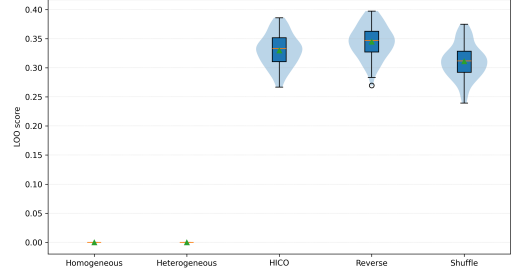

Figure 3: Leave-one-out (LOO) fitness distributions across optimization strategies. Violin plots show the distribution of LOO fitness scores across subjects, with embedded boxplots indicating median and interquartile range; triangles denote means. Homogeneous and flat heterogeneous strategies collapse to zero LOO fitness across the board, reflecting dynamical instability LOO fitness calculation. In contrast, curriculum-based strategies—particularly HICO and reverse-phased curricula—maintain robust, non-degenerate LOO performance, indicating improved cross-subject generalization.

4.1 Fitting to Individual Subjects

Figure 2 summarizes the distribution of fitness scores obtained under each optimization approach when parameters were optimized separately for each individual subject.

Relative to the homogeneous baseline, flat heterogeneous optimization, HICO, and the reverse-phased curriculum all exhibit upward shifts in both median and mean fitness. Paired t-tests against the homogeneous baseline indicate that these improvements are highly significant (Heterogeneous: $p = 2.42 \times 10^{-23}$; Phased: $p = 1.05 \times 10^{-25}$; Reverse: $p = 3.96 \times 10^{-24}$), confirming that introducing RSN-specific parameters substantially improves fit to individual subjects. In contrast, the shuffled curriculum shows substantial overlap with the homogeneous distribution and does not achieve a statistically significant improvement over baseline ($p = 0.609$).

These results establish two key points. First, increasing model complexity by moving from homogeneous to heterogeneous parameterizations yields a robust and statistically significant gain in fitting to individual subjects. Second, imposing a biologically informed curriculum does not compromise this fit: both HICO and reverse-phased optimization achieve fitness levels comparable to or exceeding those of flat heterogeneous optimization. This finding provides an essential baseline for subsequent analyses, demonstrating that curriculum-based strategies retain strong within-subject performance while enabling further investigation of their effects on generalization, stability, and behavioral relevance.

4.2 Generalization across Subjects

Cross-subject generalization was evaluated using a leave-one-out (LOO) parameter averaging based on a two-sided trimmed mean estimator ($p = 0.1$). The trimmed mean provides a robust estimate of central tendency by discarding extreme values from both tails of the parameter distribution, reducing sensitivity to outlier solutions while preserving the dominant structure shared across subjects [16]. For each left-out subject, parameters fitted on the remaining subjects were averaged and evaluated against the empirical data of the left-out subject.
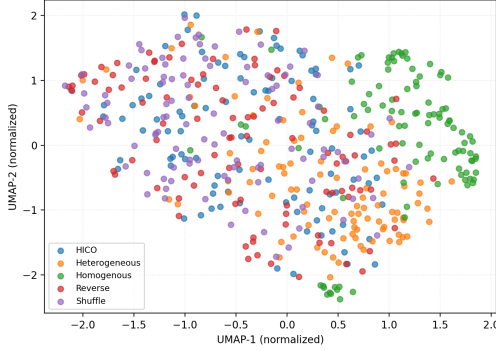
Figure 4: UMAP embedding of subject-level parameter vectors (averaged by parameter type). Each point is a subject, colored according to the approach used. Curriculum-based methods (HICO, Reverse, Shuffled) occupy a large overlapping region of parameter space, whereas homogeneous and heterogeneous approaches form distinct, focused clusters shifted away from this region. As seen in Fig. 3, the generalization properties are very different in these two areas.

Figure 3 reveals a divergence between flat and curriculum-based strategies under this evaluation. Despite achieving competitive fitness in fitting to individual subjects, both the homogeneous baseline and flat heterogeneous optimization frequently produced averaged parameter vectors with LOO fitness scores at zero. This collapse indicates that the averaged parameters lie in dynamically unstable regimes of the DMF model, consistent with failure to solve the associated Lyapunov equations and resulting in degenerate dynamics. In contrast, all curriculum-based strategies consistently avoided such instability, yielding no averaged parameter vectors with degenerate or unstable dynamics.

These findings identify an interesting generalization failure mode for unconstrained high-dimensional optimization: parameter configurations that fit individual subjects well may not survive averaging across subjects. Curriculum-guided optimization meets this challenge by constraining search trajectories toward dynamically coherent regions of parameter space that remain stable when averaged across subjects.

4.3 Geometry of Solutions

To probe how different optimization strategies shape the geometry of the learned parameter space, subject-level parameter vectors were embedded into two dimensions using Uniform Manifold Approximation and Projection (UMAP) [21]. UMAP preserves local neighborhood structure while providing a qualitative view of large-scale organization, making it possible to see how the different methods populate parameter space.

Figure 4 reveals marked differences between flat and curriculum-based strategies. Homogeneous and flat heterogeneous optimizations form relatively compact and well-separated clusters, indicating that these methods converged toward narrow regions of parameter space across subjects. Such concentration suggests that the parameter manifolds are fragile or degenerate: small perturbations may push the system outside dynamically stable regimes.

In contrast, HICO and other phased curricula (Reverse-Phased and Shuffled) produce embeddings that are more broadly distributed

and substantially overlapping. Rather than collapsing onto a single narrow cluster, these methods populate an extended region of parameter space, suggesting that curriculum-based optimization facilitates access to a wider set of dynamically viable solutions. This organization is robust under cross-subject parameter averaging, as illustrated in Section 4.2.

4.4 Predicting Behavior

Behavioral relevance of optimized DMF parameters was assessed by predicting three behavioral targets from RSN-level parameter summaries: fluid reasoning ability (PMAT24_A_CR), behavioral problems related to internalization (inwardly directed affective and somatic symptoms (ASR_Intn_Raw)), and externalization (outwardly directed behavioral tendencies (ASR_Extn_Raw)). For each approach, parameters for the seven RSNs were averaged into a single 20 dimensional vector and ridge regression models were fit to each behavioral target separately. Predictive performance was then quantified using the coefficient of determination (R^2), which measures the proportion of behavioral variance explained by the model, relative to a baseline that predicts the target's mean across subjects.

Figure 5 shows a consistent advantage for HICO across all behavioral domains. For fluid reasoning (Fig. 5A), HICO achieved the highest R^2 values across all the strategies, substantially exceeding both the homogeneous baseline and the flat heterogeneous optimization, with statistically significant improvements after multiple-comparison correction. A similar pattern was observed for behavioral problem scales, internalization scale - ASR_Intn_Raw (Fig. 5B) and externalization scale - ASR_Extn_Raw (Fig. 5C).

Across all three behavioral targets, reverse-phased curricula achieved intermediate performance, whereas shuffled curricula had lower and more variable R^2 values. These results indicate that merely increasing parameter dimensionality is insufficient to recover behaviorally meaningful representations. Instead, the ordering imposed by hierarchy-informed curricula plays a critical role in shaping parameter configurations that capture individual differences relevant to cognition and behavior. Together, these findings demonstrate that HICO not only stabilizes cross-subject generalization but also results in solutions that can be interpreted behaviorally.

4.5 Reliability of Predictions

To assess whether the observed associations between optimized DMF parameters and behavioral measures exceeded chance levels, prediction performance was evaluated using a permutation-calibrated significance framework. More specifically, the R^2 values obtained in Section 4.4 were evaluated against an empirical null distribution generated by repeatedly permuting the behavioral labels across subjects.

Figure 6 summarizes the resulting RSN-wise null distributions for the three behavioral targets. Within each panel, histograms represent the distribution of R^2_{null} values obtained from 10,000 permutations, while the vertical line indicates the observed R^2_{true} for the corresponding RSN and optimization approach. The accompanying annotations report the observed R^2 as well as permutation-derived p-values and false discovery rate−corrected q-values.
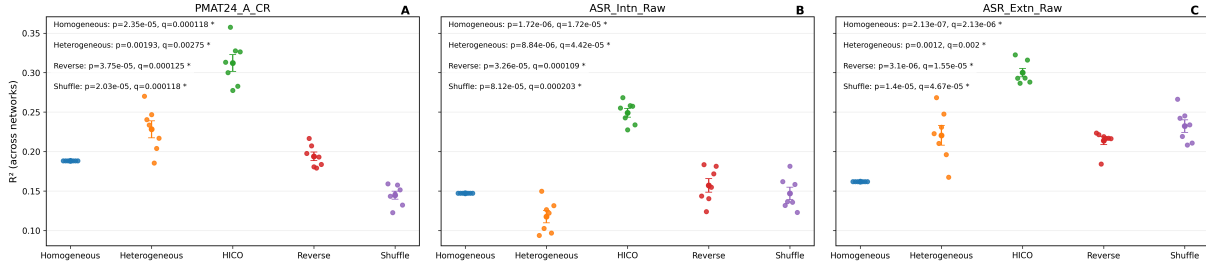
Figure 5: Predicting behavior based on solutions obtained by the different optimization strategies. Each subplot reports ridge regression R^2 values (across RSNs) for a different behavioral target: (A) fluid reasoning ability (PMAT24_A_CR), (B) inwardly directed affective and somatic symptoms (ASR_Intn_Raw), and (C) outwardly directed behavioral tendencies (ASR_Extn_Raw). For each optimization strategy, colored points show the R^2 values obtained for the seven individual RSNs, and error bars indicate the mean ± standard deviation across RSNs. Reported p-values and q-values correspond to permutation-based significance testing with False Detection Rate (FDR) correction. Across all targets, only HICO produced solutions that encode behaviorally relevant information.

Across all three behavioral targets, HICO produced the most consistent and pronounced deviations from the null distribution. Multiple RSNs under HICO had R^2 values well in the upper tail of the permutation distribution. These effects are most robust for fluid reasoning ability. In contrast, other optimization strategies yielded R^2 values that fell near the center of the null distribution, indicating weak or unreliable parameter–behavior relationships.

Importantly, this permutation-based analysis complements the behavioral prediction results by localizing where among the RSNs a behaviorally meaningful signal emerges. The findings demonstrate that the superior predictive performance of HICO is not driven by diffuse, nonspecific effects, but rather by structured and statistically considerable associations within specific cortical systems. Together, these results indicate that solutions found by the biologically motivated curriculum approach predict behavior in a statistically reliable manner.

5 Discussion and Future Work

The experimental results demonstrate that the curricular structure of parameter optimization is crucial in constructing whole-brain biophysical models that generalize across subjects and can be used to predict behavior. Although evolutionary algorithms are designed to explore high-dimensional search spaces without explicit inductive biases, the order in which parameter subspaces are explored can strongly shape the resulting search trajectory. In Dynamic Mean Field (DMF) models, where nonlinear interactions give rise to emergent large-scale dynamics, curriculum structure acts as an effective constraint that promotes stability, cross-subject generalization, and behavioral relevance.

A natural next step is to scale this framework to larger and more diverse datasets. In particular, extending the analysis to the full Human Connectome Project (with $n \approx 1200$ subjects) will allow characterizing interindividual variability more precisely, thus making behavior predictions more accurate. In parallel, applying the proposed optimization framework to translational and clinical datasets will make it possible to study how interventions such as neuromodulation or pharmacological treatments alter the underlying dynamical regimes, providing a pathway for using whole-brain models to assess treatment-induced changes in large-scale brain organization.

Complementary to these data-driven extensions, two methodological directions emerge naturally from the evolutionary optimization perspective. First, extending the HICO framework to substantially higher-dimensional parameterizations may further improve biological significance of the solutions. As whole-brain models incorporate finer-grained region and circuits, understanding how curricular structure scales with dimensionality will be essential for promoting stability, generalization, and relevance.

Second, the HICO framework may be combined with distributed attention-based approaches such as BLADE [30]. Applying BLADE within each optimization phase could selectively reduce the effective search dimensionality while preserving hierarchical structure, thereby compensating for the increased complexity of larger models. Such a hybrid strategy offers a principled way to balance expressivity and efficiency, and makes it possible to compare curriculum-guided attention-based evolution and unconstrained optimization of higher-dimensional models. Exploring this interaction may yield more scalable evolutionary optimization schemes for large, structured dynamical systems.

6 Conclusion

This study demonstrated that the curricular structure of parameter optimization plays a critical role in generality and relevance of whole-brain biophysical models. Biologically motivated structure acts as a powerful inductive bias in the evolutionary optimization: it uncovers parameter regimes that generalize effectively across subjects, and can be used to predict behavior. Curriculum-informed evolutionary optimization thus provides a principled approach for scaling complex biophysical models. This framework offers a promising direction for future work in large-scale computational neuroscience and other high-dimensional scientific modeling domains.
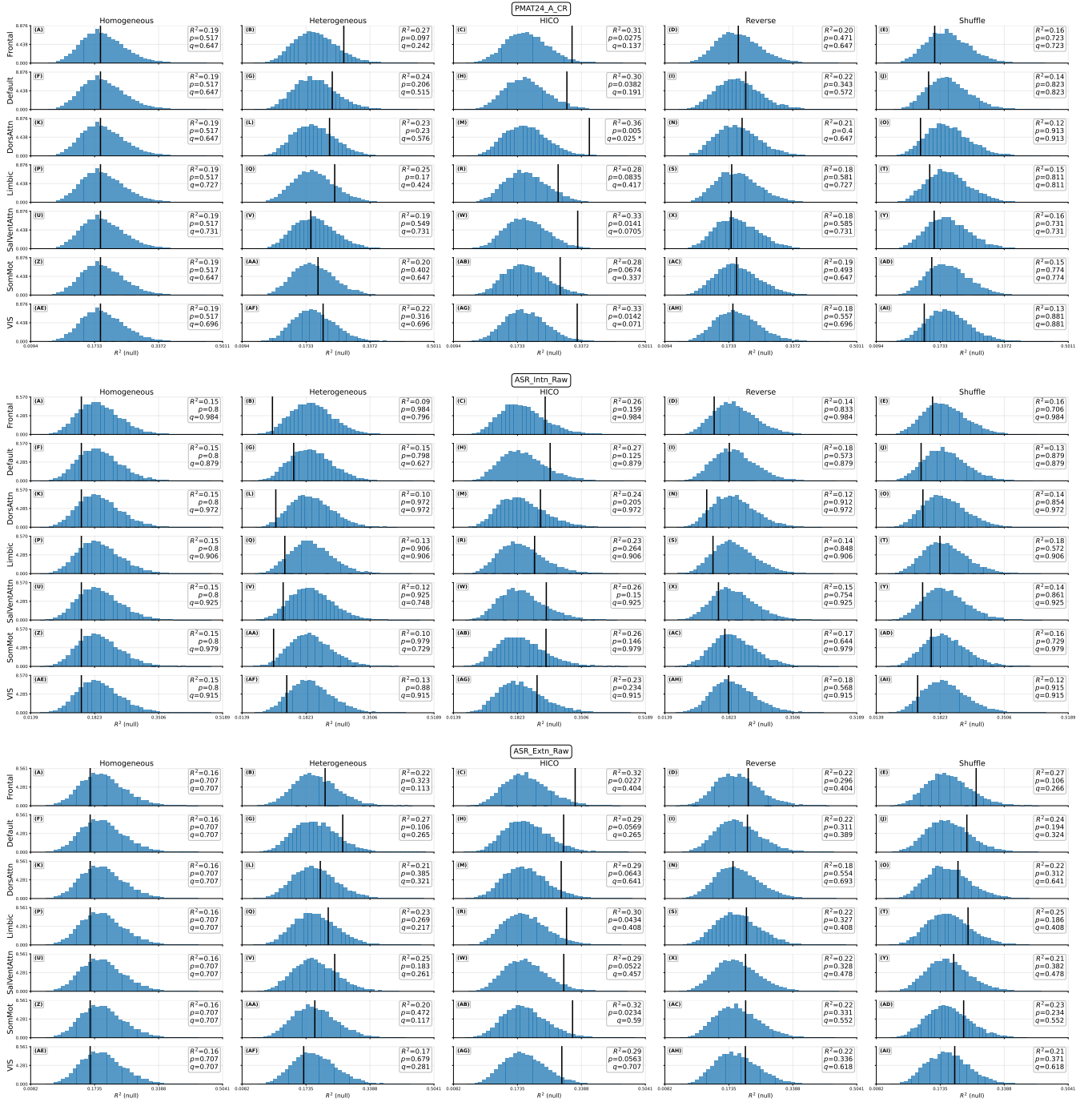
Acknowledgments

Figure 6: Localization of permutation-calibrated parameter–behavior associations across the different RSNs. Top: fluid reasoning ability (PMAT24_A_CR). Middle: inwardly directed behavioral problems (ASR_Intn_Raw). Bottom: outwardly directed behavioral problems (ASR_Extn_Raw). For each RSN and optimization approach, histograms show the null distribution of ridge regression R^2 obtained from 10,000 permutations of behavioral labels, while vertical lines indicate the observed R^2 from the true data. Reported p- and q-values quantify statistical significance relative to the permutation null. As indicated by the vertical lines, biology-informed optimization results in parameter solutions that consistently exceed chance levels and provide a strong foundation for predicting behavioral.

References

[1] Pablo Achard and Erik De Schutter. 2006. Complex parameter landscape for a complex neuron model. *PLoS Computational Biology* 2, 7 (2006). doi:10.1371/journal.pcbi.0020094

[2] Thomas M. Achenbach, Masha Y. Ivanova, and Leslie A. Rescorla. 2017. Empirically based assessment and taxonomy of psychopathology for ages 1½–90+ years: Developmental, multi-informant, and multicultural findings. *Comprehensive Psychiatry* 79 (2017), 4–18. doi:10.1016/j.comppsych.2017.03.006 Advances in Transdiagnostic Psychopathology Research.

[3] Michael Breakspear. 2017. Dynamic models of large-scale brain activity. *Nature Neuroscience* 20, 3 (2017), 340–352. doi:10.1038/nn.4497

[4] Joana Cabral, Morten L. Kringelbach, and Gustavo Deco. 2014. Exploring the network dynamics underlying brain activity at rest. *Philosophical Transactions of the Royal Society B* 369, 1653 (2014). doi:10.1098/rstb.2013.0518

[5] Kalyanmoy Deb and Cornelius Myburgh. 2017. Discrete Optimization. *European Journal of Operational Research* 261, 2 (2017), 460–474. doi:10.1016/j.ejor.2017.02.015

[6] Gustavo Deco, Viktor K. Jirsa, and Anthony R. McIntosh. 2011. Emerging concepts for the dynamical organization of resting-state activity in the brain. *Nature Reviews Neuroscience* 12, 1 (2011), 43–56. doi:10.1038/nrn2961

[7] Gustavo Deco, Viktor K. Jirsa, Anthony R. McIntosh, Olaf Sporns, and Rolf Kötter. 2013. Key role of coupling, delay, and noise in resting brain fluctuations. *Proceedings of the National Academy of Sciences* 110, 12 (2013). doi:10.1073/pnas.1221022110

[8] Gustavo Deco, Adrian Ponce-Alvarez, Dante Mantini, Guilherme L. Romani, Patric Hagmann, and Maurizio Corbetta. 2014. Resting-state functional connectivity emerges from structurally and dynamically shaped slow linear fluctuations. *Journal of Neuroscience* 34, 25 (2014), 7886–7897. doi:10.1523/JNEUROSCI.3288-13.2014

[9] Shaul Druckmann, Yochai Banitt, Albert Gidon, Felix Schürmann, Henry Markram, and Idan Segev. 2007. A novel multiple objective optimization framework for constraining conductance-based neuron models by experimental data. *Frontiers in Neuroscience* 1, 1 (2007), 7–18. doi:10.3389/neuro.01.1.001.2007

[10] Karl J. Friston, Lee Harrison, and William Penny. 2003. Dynamic causal modelling. *NeuroImage* 19, 4 (2003), 1273–1302. doi:10.1016/S1053-8119(03)00202-7

[11] Faustino Gomez and Risto Miikkulainen. 1997. Incremental Evolution of Complex General Behavior. *Adaptive Behavior* 5, 3-4 (1997), 317–342. doi:10.1177/105971239700500305

[12] Nikolaus Hansen. 2016. The CMA Evolution Strategy: A Tutorial. *arXiv preprint arXiv:1604.00772* (2016). https://arxiv.org/abs/1604.00772

[13] Desmond J. Higham. 2001. An Algorithmic Introduction to Numerical Simulation of Stochastic Differential Equations. *SIAM Rev.* 43, 3 (2001), 525–546. doi:10.1137/S0036144500378302

[14] John H. Holland. 1975. *Adaptation in Natural and Artificial Systems*. University of Michigan Press. https://mitpress.mit.edu/9780262581110/adaptation-in-natural-and-artificial-systems/

[15] Christopher J. Honey, Olaf Sporns, Leila Cammoun, Xavier Gigandet, Jean-Philippe Thiran, Reto Meuli, and Patric Hagmann. 2009. Predicting human resting-state functional connectivity from structural connectivity. *Proceedings of the National Academy of Sciences* 106, 6 (2009), 2035–2040. doi:10.1073/pnas.0811168106

[16] Peter J. Huber and Elvezio M. Ronchetti. 2009. *Robust Statistics* (2nd ed.). John Wiley & Sons, Hoboken, NJ, USA. doi:10.1002/9780470434697

[17] Naomi Keren, Noam Peled, and Alon Korngreen. 2005. Automated parameter estimation of multi-compartment models of single neurons. *Journal of Neuroscience Methods* 146 (2005), 133–144. doi:10.1016/j.jneumeth.2005.02.006

[18] Xiaodong Ma, Yuhui Li, Dunwei Gong, Jun Zhang, and Jun Liu. 2019. A Survey on Cooperative Co-Evolutionary Algorithms. *IEEE Transactions on Evolutionary Computation* 23, 3 (2019), 421–441. doi:10.1109/TEVC.2018.2878671

[19] Kaitlin Maile, Manish Saggar, and Risto P. Miikkulainen. 2019. Implementing Evolutionary Optimization to Model Neural Functional Connectivity. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion (GECCO '19 Companion)*. ACM, 1731–1733. doi:10.1145/3319619.3326826

[20] Daniel S. Margulies, Satrajit S. Ghosh, Alexandros Goulas, Magdalena Falkiewicz, Julia M. Huntenburg, Georg Langs, Gleb Bezgin, Simon B. Eickhoff, F. Xavier Castellanos, Michael Petrides, Elizabeth Jefferies, and Jonathan Smallwood. 2016. Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proceedings of the National Academy of Sciences* 113, 44 (2016), 12574–12579. doi:10.1073/pnas.1608282113

[21] Leland McInnes, John Healy, Nathaniel Saul, and Lukas Grossberger. 2018. UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software* 3, 29 (2018), 861. doi:10.21105/joss.00861

[22] Risto Miikkulainen. 2019. Creative AI Through Evolutionary Computation. *arXiv preprint* (2019). arXiv:1901.03775 [cs.AI] https://arxiv.org/abs/1901.03775

[23] T. M. Moore, S. P. Reise, R. E. Gur, H. Hakonarson, and R.C. Gur. 2014. Psychometric Properties of the Computerized Neurocognitive Battery. Proc. Natl. Acad. Sci. U.S.A. doi:10.1037/neu0000093 Includes the Penn Matrix Reasoning Test (PMAT) as measure of nonverbal reasoning ability..

[24] John D. Murray, Alberto Bernacchia, David J. Freedman, Ranulfo Romo, Jonathan D. Wallis, Xue-Lian Cai, Camillo Padoa-Schioppa, Tatiana Pasternak, Hyojung Seo, Daeyeol Lee, and Xiao-Jing Wang. 2014. A hierarchy of intrinsic timescales across primate cortex. *Nature Neuroscience* 17, 12 (2014), 1661–1663. doi:10.1038/nn.3862

[25] Martin Pelikan. 2005. *Bayesian Optimization Algorithm*. Springer Berlin Heidelberg, Berlin, Heidelberg, 31–48. doi:10.1007/978-3-540-32373-0_3

[26] Mitchell A. Potter and Kenneth A. De Jong. 2000. Cooperative Coevolution: An Architecture for Evolving Coadapted Subcomponents. *Evolutionary Computation* 8, 1 (2000), 1–29. doi:10.1162/106365600568086

[27] Alexander Schaefer, Ru Kong, Evan M Gordon, Timothy O Laumann, Xi-Nian Zuo, Avram J Holmes, Simon B Eickhoff, and B T Thomas Yeo. 2017. Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cerebral Cortex* 28, 9 (07 2017), 3095–3114. doi:10.1093/cercor/bhx179

[28] Hormoz Shahrzad, Daniel Fink, and Risto Miikkulainen. 2018. Enhanced optimization with composite objectives and novelty selection. In *Artificial Life Conference Proceedings*. MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA, 616–622. doi:10.1162/isal_a_00113

[29] Hormoz Shahrzad, Babak Hodjat, Camille Dollé, Andrei Denissov, Simon Lau, Donn Goodhew, Justin Dyer, and Risto Miikkulainen. 2019. Enhanced Optimization with Composite Objectives and Novelty Pulsation. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*. 507–515. doi:10.1145/3321707.3321836

[30] Hormoz Shahrzad and Risto P Miikkulainen. 2023. Accelerating Evolution Through Gene Masking and Distributed Search. In *Proceedings of the Genetic and Evolutionary Computation Conference* (Lisbon, Portugal) *(GECCO '23)*. Association for Computing Machinery, New York, NY, USA, 972–980. doi:10.1145/3583131.3590508

[31] Kenneth O. Stanley and Risto Miikkulainen. 2002. Evolving Neural Networks Through Augmenting Topologies. *Evolutionary Computation* 10, 2 (2002), 99–127. doi:10.1162/106365602320169811

[32] Werner Van Geit, Michael Gevaert, Giuseppe Chindemi, Christian Rössert, Jean-Denis Courcol, Eilif B. Muller, Felix Schürmann, Idan Segev, and Henry Markram. 2016. BluePyOpt: Leveraging open source software and cloud infrastructure to optimize model parameters in neuroscience. *Frontiers in Neuroinformatics* 10 (2016), 17. doi:10.3389/fninf.2016.00017

[33] Marie C. Vanier and James M. Bower. 1999. A comparative survey of automated parameter-search methods for compartmental neural models. *Journal of Computational Neuroscience* 7, 2 (1999), 149–171. doi:10.1023/A:1008972005316

[34] Kong-Fatt Wong and Xiao-Jing Wang. 2006. A recurrent network mechanism of time integration in perceptual decisions. *Journal of Neuroscience* 26, 4 (2006), 1314–1328. doi:10.1523/JNEUROSCI.3733-05.2006

[35] B. T. Thomas Yeo, Fenna M. Krienen, Jorge Sepulcre, Mert R. Sabuncu, Danial Lashkari, Marisa Hollinshead, Joshua L. Roffman, Jordan W. Smoller, Lilla Zöllei, Jonathan R. Polimeni, Bruce Fischl, Hesheng Liu, and Randy L. Buckner. 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology* 106, 3 (2011), 1125–1165. doi:10.1152/jn.00338.2011

A Dynamic Mean Field Parameters and Fitness Computation

This appendix specifies the Dynamic Mean Field (DMF) model parameters, their biological interpretation, default values and ranges, and the procedure used to compute functional connectivity (FC)–based fitness.

A.1 DMF State Variables and Fixed Constants

Each cortical region i is modeled as a coupled excitatory–inhibitory neural population. The dynamical state is described by synaptic gating variables $S_i(t)$ (excitatory) and $S_i^I(t)$ (inhibitory), evolving according to the DMF equations defined in Section 2.2.

The following constants were fixed across all experiments and adopted from prior DMF studies [3, 7, 24].

Symbol	Description	Value
τ_E	Excitatory synaptic time constant	100 ms
τ_I	Inhibitory synaptic time constant	10 ms
γ	Synaptic gain	0.641
a_E	Excitatory gain slope	$310\ \mathrm{nC}^{-1}$
b_E	Excitatory firing threshold	125 Hz
d_E	Excitatory curvature parameter	0.16 s
I_b	Background input current	fixed

The sigmoidal transfer function $H(\cdot)$ follows the standard DMF formulation [7], mapping total synaptic input to population firing rate.

A.2 Optimized Parameters per RSN

The Dynamic Mean Field (DMF) model parameters optimized in this study follow the formulation of Deco et al. [6, 7], Murray et al. [24]. Each RSN is associated with an identical ordered block of 20 parameters.

For clarity, the parameters are grouped into (i) *local circuit parameters* (Table 1), governing excitatory–inhibitory dynamics within each region, and (ii) *global and long-range coupling parameters* (Table 2), governing noise, synaptic scaling, and interactions between regions. Reported ranges correspond to bounds used in prior DMF studies; default values refer to canonical settings commonly adopted before optimization.

Notes on parameter usage. In the heterogeneous formulation, Tables 1 and 2 together define a 20-dimensional parameter block that is replicated independently across the seven RSNs, yielding a total of 140 free parameters. In the homogeneous baseline, a single shared instance of this block is optimized and applied uniformly across all RSNs. All parameters were represented internally in normalized coordinates in $[0, 1]$ and affinely mapped to the ranges listed above during fitness evaluation.

A.3 Simulation Pipeline

For a candidate parameter vector Θ, the DMF system was simulated for five minutes of biological time using Euler–Maruyama integration [13] with step size $\Delta t = 0.1$ ms. The resulting excitatory synaptic activity $S_i(t)$ was transformed into BOLD signals using the Balloon–Windkessel hemodynamic model [10].

Index	Parameter	Default / Range	Description
0	a_E	310 [200–400]	Gain of the excitatory firing-rate input–output function.
1	b_E	125 [100–150]	Offset (bias) of the excitatory firing-rate function.
2	d_E	0.16 [0.1–0.3]	Curvature parameter controlling slope of excitatory nonlinearity.
3	W_E	1.0 [0.5–2.0]	Scaling of recurrent excitatory input.
4	τ_E	100 ms [50–200]	Time constant of excitatory synaptic gating.
5	a_I	615 [400–800]	Gain of the inhibitory firing-rate function.
6	b_I	177 [150–220]	Offset (bias) of the inhibitory firing-rate function.
7	d_I	0.087 [0.05–0.15]	Curvature parameter of inhibitory nonlinearity.
8	W_I	0.7 [0.3–1.5]	Scaling of recurrent inhibitory input.
9	τ_I	10 ms [5–20]	Time constant of inhibitory synaptic gating.
10	w_{EE}	1.4 [0.5–2.5]	Excitatory-to-excitatory recurrent coupling strength.
11	w_{EI}	1.0 [0.5–2.0]	Inhibitory input strength onto excitatory population.
12	w_{IE}	1.0 [0.5–2.0]	Excitatory input strength onto inhibitory population.
13	w_{II}	0.5 [0.1–1.5]	Inhibitory-to-inhibitory recurrent coupling strength.

Table 1: Local circuit parameters optimized per RSN. These parameters control the excitatory–inhibitory population dynamics within each cortical region.

Index	Parameter	Default / Range	Description
14	I_b	0.382 [0.2–0.6]	Background (external) input current to excitatory population.
15	J	0.15 [0.05–0.3]	Synaptic scaling factor converting firing rate to synaptic current.
16	γ	0.641 [0.3–1.0]	Gain of excitatory synaptic gating dynamics.
17	γ_I	1.0 [0.5–1.5]	Gain of inhibitory synaptic gating dynamics.
18	σ	0.01 [0.001–0.05]	Noise amplitude driving stochastic synaptic fluctuations.
19	g	2.5 [0–5]	Global coupling strength applied to structural connectivity.

Table 2: Global and long-range coupling parameters optimized per RSN. These parameters regulate noise, synaptic scaling, and interregional communication through the structural connectome.

Functional connectivity matrices $\widehat{\mathrm{FC}}(\Theta)$ were computed as Pearson correlations between regional BOLD time series, matching the construction of empirical FC from resting-state fMRI.

A.4 Fitness Definition and Evaluation

Although the loss function in Eq. (2) is written as a minimization objective, evolutionary optimization was performed by maximizing the Pearson correlation between empirical and simulated FC matrices:

$$\text{fitness}(\Theta) = \text{corr}\left(\text{vec}(\text{FC}),\ \text{vec}\left(\widehat{\mathrm{FC}}(\Theta)\right)\right).$$

Throughout the paper, the terms *fitness* and *score* refer to this correlation value.

For leave-one-out (LOO) evaluation, subject-specific parameter vectors were aggregated using a two-sided trimmed mean ($p = 0.1$), producing a cohort-level parameter estimate $\tilde{\theta}^{(-j)}$ that was then evaluated on the held-out subject using the same FC-based fitness computation. The moments-based fitness implementation follows that of Murray et al. [24].

A.5   Methodological Clarifications

All simulations used subject-specific structural connectivity matrices derived from diffusion-weighted imaging. Behavioral prediction analyses and permutation-based significance testing were performed after all subject-level evolutionary runs were completed and did not influence optimization dynamics. This separation ensures that the learned DMF parameters define task-agnostic whole-brain dynamical models suitable for downstream analysis across multiple behavioral domains.