bioRxiv preprint doi: https://doi.org/10.1101/780163; this version posted September 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Aging, computation, and the evolution of neural regeneration processes

Aina Ollé-Vila[1,2], Luís F Seoane[3], Ricard Solé[1,2,4*]
[1] *ICREA-Complex Systems Lab, Universitat Pompeu Fabra, 08003 Barcelona*
[2] *Institut de Biologia Evolutiva (CSIC-UPF), Psg Maritim Barceloneta, 37, 08003 Barcelona*
[3] *Instituto de Física Interdisciplinar y Sistemas Complejos IFISC (CSIC-UIB), Campus UIB, 07122, Palma de Mallorca, Spain and*
[4] *Santa Fe Institute, 399 Hyde Park Road, Santa Fe NM 87501, USA*

Metazoans are capable of gathering information from their environments and respond in predictable ways. These computational tasks are achieved by means of more or less complex networks of neurons. Task performance must be reliable over an individual's lifetime and must deal robustly with the finite lifespan of cells or with connection failure – rendering aging a relevant feature in this context. How do computations degrade over an organism's lifespan? How reliable can computations remain throughout? In order to answer these questions, here we approach the problem under a multiobjective (Pareto) optimization approach. We consider a population of digital organisms equipped with a neural network that must solve a given computational task reliably. We demand that they remain functional (as reliable as possible) for an extended lifespan. Neural connections are costly (as an associated metabolism in living beings) and degrade over time. They can also be regenerated at some expense. We investigate the simultaneous minimization of the metabolic burden (due to connections and regeneration costs) and the computational error and the tradeoffs emerging thereof. We show that Pareto optimal designs display a broad range of potential solutions: from small networks with high regeneration rate, to larger and redundant circuits that regenerate slowly. The organism's lifespan and the external damage rates are found to act as an evolutionary pressure that improve the exploration of the space of solutions and poses tighter optimality conditions. We also find that large damage rates to the circuits can constrain the space of the possible and pose organisms to commit to unique strategies for neural systems maintenance.

I. INTRODUCTION

A major revolution in the history of multicellular life was the emergence of neurons, a new class of cells that enhanced the processing and storage of information beyond the genetic level [1]. Such revolution enabled fast adaptation to environmental fluctuations. Combining these apt building blocks, a more short-sighted sensor-actuator logic was soon backed by more complex networks, resulting in yet further increased capabilities for information processing and representation of the external environment [2]. Alongside, organism size and longevity also increased. Both required regeneration mechanisms that would sustain organismal coherence over extended periods of time, far beyond the cellular lifespan. An alternative (or complementary) way to guarantee reliable computation is through an appropriate connectivity pattern – e.g. faulty functioning due to loss of connections could be counterbalanced by redundant wires, among other mechanisms. How this can be achieved was early explored by the first generation of mathematicians dealing with unreliable computers [3, 4]. More recent works have addressed some of these questions from the perspective of reliability theory [5], particularly in relation to the role played by parallel versus sequential topologies [6].

The ability to regenerate the nervous system presents an extraordinary variation in Metazoa. Regarding vertebrate species, all of them can produce new neurons postnatally in specific regions of their nervous system, but only some lower vertebrates (fish and amphibians) can significantly repair several neural structures. Some regenerative ability, however, is found also in reptiles and birds, and even in mammals [7]. Remarkably, replacement of all or part of the nervous system has been documented in a few invertebrate phyla, including coelenterates, flatworms, annelids, gastropods and tunicates [8]. Such re-grown neural networks are parsimoniously integrated within the rest of the circuitry, stressing how phenotypic functionality is recovered. Although this ability is largely absent in higher vertebrates, evidence piles up that the potential might lay dormant [9–11]. Another open question concerns whether new neurons are being created throughout our lives in the absence of damage. While few or no new neurons seem to grow in most parts of human brains, there is evidence of limited neurogenesis in the hippocampus [12, 13] and the olfactory bulb [14]. Several species, notably fish, present large rates of neurogenesis, often requiring apoptosis of older cells to make place for the new ones [15].

The origin of this diversity of strategies around nervous system maintenance is a major challenge [16]. What are the evolutionary drivers behind each solution? The metabolic cost of wires readily comes to mind, and has already been explored as a major constraint of neural architecture [17–19], while the regeneration costs are also obvious. The relevance of these factors must be analyzed under the light of a phenotypic function. This, in neural circuits, traces back to computations that must be implemented within reasonable error bounds. This computa-

*ricard.sole@upf.edu

bioRxiv preprint doi: https://doi.org/10.1101/780163; this version posted September 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

tional performance constitutes a third dimension relevant to our research questions.

What is the optimal tradeoff between these factors for reliable neural circuits? Is the range of solutions parsimonious, or are there more locally stable designs that hinder the access to other possibilities? In order to answer these questions, we asses the maintenance of reliable computations over extended lifespans while enduring an aging process (inflicted through an external damage). Interestingly, the interplay between the lifespan of the whole organism versus the time scale of its constituent parts brings in an extra factor in our study. Its relevance becomes apparent in the empirical record, notably in the apoptosis of older neurons sought by some fish species [15] – implying that a valid regeneration strategy actively shortens the useful life of the organism's building blocks. How are design spaces affected by these different factors – computational performance, organismal lifespans (and its relation to component time scales), external damages, and the severity of metabolic costs? Early and current research has studied how given computational functions are implemented by evolved neural networks [20–25], but these studies seldom connect wiring costs with possible repair processes.

As noted above, our research questions can only be addressed given a computational task that the underlying circuitry must solve. In living organisms this further results in diverse anatomical patterns and neural plans, with the computational tasks ingrained in each organism's phenotype. This additional diversity falls beyond the scope of this paper. To gain some insight in the tradeoffs behind neural circuits maintenance, we resort to minimal toy models based on networks of Boolean units that solve archetypal tasks. Simple Boolean models have been used to explore the basic principles of neural functions and the role played by architecture, including information propagation thresholds [26], locomotion and gating [27–29], pattern formation [30], or the emergence of modularity [31]. In a more general context of evolved circuits for artificial agents, evolved neural networks play a central role in the development of biologically inspired robotic systems [32, 33].

In this spirit, we test feed-forward networks of Boolean units. We set a fixed number of layers, a varying number of units in each layer and connections across layers, and a range of regeneration rates. These toy neural circuits are tasked with solving a series of computations while responding to the three evolutionary forces outlined above: i) a cost stemming from computational errors, ii) another cost associated to wiring, and iii) the cost associated to the regeneration of damaged structures. To integrate these evolutionary forces without introducing unjustified biases that would assign more importance to a factor over the others, we adopt a (Pareto) Multi-Objective Optimization (MOO) approach [34–36]. This framework explores designs that simultaneously minimize all costs involved. The solution of a MOO problem is a restricted region in the space of possible designs. This solution embeds the diversity of somehow optimal strategies in a mathematical object whose geometry has been linked to phase transitions [37–41] and criticality [39, 41]. These phenomena give us some insights about how accessible the range of optimal solutions are: whether evolutionary biases can navigate them smoothly as they vary, or whether locally optimal designs dominate under discrete value regions of the different costs so that changes only happen abruptly.

In section II we introduce the elements of our toy model: i) the computational tasks explored, ii) the implementation of the feed-forward networks and their aging process, and iii) how all of this comes together under MOO. In section III we go over the results, including how each evolutionary pressure affects the shape of MOO solutions in design space, and how this relates to the biology of the problem. Section IV concludes discussing our results within existing literature.

II. METHODS

A. Computational tasks

We ask our toy neural networks to compute some arbitrary Boolean function:

$$\varphi : \Sigma^N \longrightarrow \Sigma^M, \tag{1}$$

with $\Sigma = \{0, 1\}$. In this paper we used $N = 3$, $M = 1$, and two archetypal computations (figure 1): i) The *multiplexer* (*Mux*) where the "selector" bit i_3 chooses one of the other inputs (i_1 or i_2) as output. ii) The *majority rule* (*Maj*), which returns 1 if most input bits are 1 and 0 otherwise. *Maj* is well known within the study of circuit redundancy and error correction.

B. Feed-forward neural networks

We subjected a population of Boolean, feed-forward neural networks to an evolutionary process that optimizes diverse design aspects according to the MOO logic detailed below. Here we clarify the general network architecture, how they compute, and some of the diversity allowed.

Each network consisted of three input units ($I \equiv \{i_1, i_2, i_3\}$), two hidden layers with a varying number of units in each layer (from 1 to $h_{max} = 15$) and connections across layers, and one output unit ($O \equiv \{o_1\}$). Each connection carried a weight $\omega_{ij} = \pm 1$, and each unit had an activation threshold $\theta_i \in [0.01, 1]$. Networks computed following a McCulloch-Pitts function [42]:

$$S_i = H\left(-\theta_i + \textstyle\sum_j \omega_{ij} S_j\right), \tag{2}$$

where j runs over input signals to unit i, and $H(\cdot)$ represents the Heaviside step function.

bioRxiv preprint doi: https://doi.org/10.1101/780163; this version posted September 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

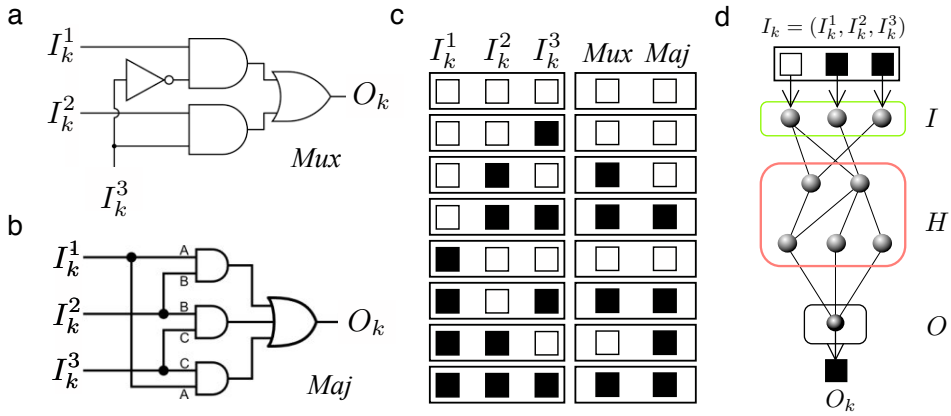

FIG. 1: **Computational tasks and neural implementation**. The agents used in our evolutionary algorithm are tested with two different, three-input, one output logic functions: (a) the multiplexer (*Mux*) and (b) the majority rule (*Maj*). These are traditionally implemented using logic gates as shown. Each circuit has an associated logic table that fully defines each performed computation. (c) Left, the eight possible binary sets of inputs ■ = 0 and □ = 1. Each possible input string, such as ■■□, results in a unique output. The outputs for *Mux* and *Maj* are shown (right). (d) In our evolving system, Boolean gates are replaced by feed-forward neural networks with several layers including input (I), output (O), and hidden (H) ones.

Thus each network computes a Boolean function. With the variety allowed across networks (different number of units, connections, and θ_i), there are several ways to implement a desired function. Different network designs will incur in different costs due to the expense of wiring or regeneration (see below), and their computational reliability will degrade differently for distinct topologies as they *age* (also, see below). This constitutes the basis of our MOO exploration.

C. Aging process

The networks are evaluated over an extended period $t = 1, \ldots, \tau$ during which their connections are eroded. τ defines the required durability for the whole network, which would correspond to the lifespan of a modeled organism. At the beginning of each evaluation step, each connection is removed with a probability δ, defining a *damage rate*. Also, the network restores each missing link with probability ρ, thus defining a *regeneration rate*. Recovered connections display their original weight (i.e. some detailed memory is never lost).

After knocking off some connections and regenerating others, we assess the reliability of each network in computing φ (eq. 1, i.e. *Mux* or *Maj*). An approximate mean-field model (see Supplementary Material, section II) shows how the damage/regeneration process eventually results in an average steady number of missing connections.

We performed experiments with different δ and τ. Notice that δ defines a damage rate, but also contributes to setting an average lifespan ($\tau_{link} \sim 1/\delta$) for the network connections. This, together with the timescale of the network lifespan can be combined into a dimensionless ratio $r_\tau \equiv \delta \cdot \tau \sim \tau/\tau_{link}$. This ratio is an informative index in analyzing some results, but note that the actual τ_{link} is also affected by regeneration values.

This aging process might result in plainly unfeasible networks – i.e. graphs that become disconnected such that information cannot flow from input to output. We track the proportion of time that a network γ is thus broken through a *feasibility function* $F_f(\gamma) \in [0, 1]$:

$$F_f(\gamma) \;=\; \frac{1}{\tau}\sum_t \xi(\gamma, t); \qquad (3)$$

where $\xi(\gamma, t) = 1$ if γ remains feasible (connections exist from input to output, and from all input units to the first hidden layer after damage and regeneration at t). Otherwise, $\xi(\gamma, t) = 0$ (check section III-E in Supplementary Material for more details).

D. Multiobjective Optimization

Given the set Γ of all allowed networks ($\gamma \in \Gamma$), we seek the subset $\Pi \subset \Gamma$ of designs $\gamma^\pi \in \Pi$ that simultaneously minimize all relevant costs without introducing artificial biases. This subset (Π) is the solution of a MOO or Pareto optimization problem [34–36]. Pareto optimal networks $\gamma^\pi \in \Pi$ are such that you cannot find two of them $\gamma_i^\pi, \gamma_j^\pi \in \Pi$ with one being better than the other with respect to all optimization targets. Pareto optimal designs constitute the best tradeoff possible between the studied traits: we can only improve one of them by worsening some other.

We are invested in the minimization of three optimization targets ($\{T_1, T_2, T_3\}$):

- **i Error** $\epsilon(\gamma) \in [0, 1]$ in implementing the computational task φ (eq. 1). At each evaluation step ($t = 1, \ldots, \tau$), after damage and regeneration, the

bioRxiv preprint doi: https://doi.org/10.1101/780163; this version posted September 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

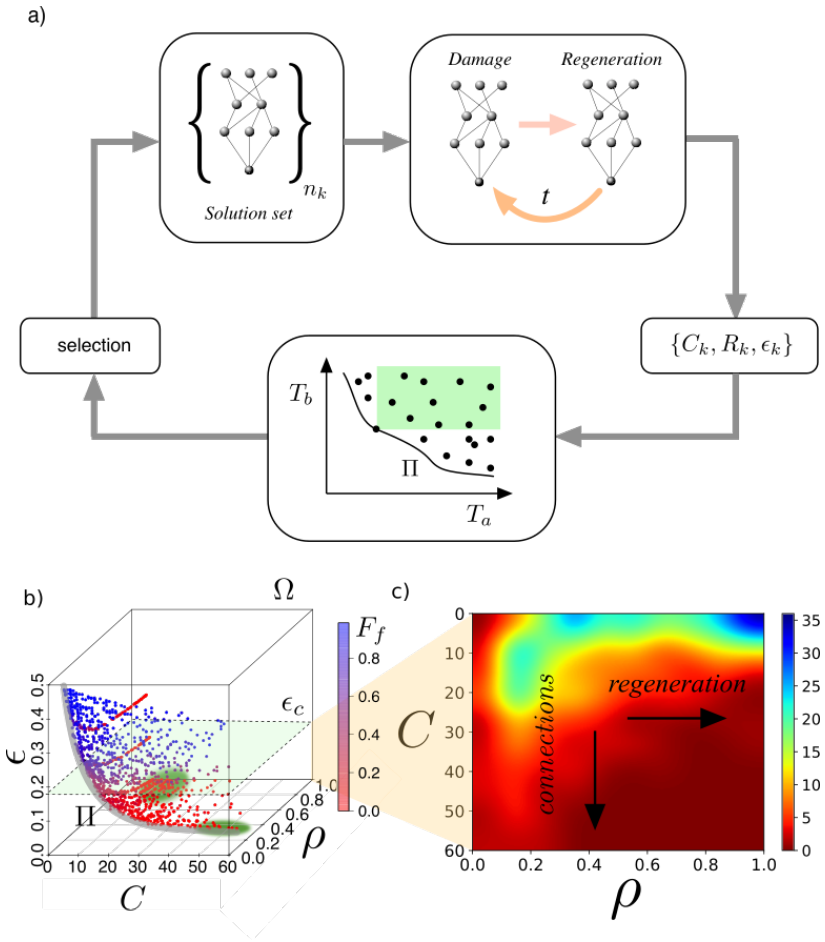

FIG. 2: **Evolutionary algorithm for neural regeneration and output processing**. (a) Iterative process of the MOO algorithm. A population $\hat{\Gamma}$ of $N = 480$ networks $\gamma \in \hat{\Gamma}$ is subjected to a damage: at each of τ time steps, their connections are lost with probability δ; meanwhile, the network attempts to compute. Computational performance of this damaged network ($T_1 \equiv \epsilon(\gamma)$) is measured, along with costs associated to wiring ($T_2 \equiv C(\gamma)$) and regeneration of lost connections ($T_3 \equiv \rho(\gamma)$). These *optimization targets* should be minimized. Bad solutions have a worse performance in all those targets simultaneously (green shaded area): they are *dominated*. Better network designs cannot perform optimally in all senses. They are rather better in trading good performance in a target for worse performance in others, thus avoiding dominance. Less dominated networks at a given iteration of the MOO algorithm are carried to the next generation and used to produce offspring. This algorithm proceeds for g^{max} iterations. (b) We execute this algorithm ten times for each fixed conditions of damage (δ) and length of the aging process (τ). This results in a combined population $\hat{\Gamma}$ whose non-dominated subset $\hat{\Pi}^{end} \subset \hat{\Gamma}$ approximates the optimal tradeoff between the targets involved. Blurred green circles indicate extreme phenotypes (investment is maximal in regeneration and minimal in connectivity, or vice versa). A steep relation between the computation error (T_1) and the structural feasibility of the network (as captured by $F_f(\gamma)$, see SM section VII, figures S18-S25) suggests a robust threshold delimiting acceptable computation $\epsilon_c = 0.2$. (c) Density of network designs in $\hat{\Gamma}^{end}$ over the $\rho - C$ plane that compute acceptably ($\epsilon(\gamma) < \epsilon_c$). These plots will be used to explore the tradeoff between regeneration and high connectivity.

network γ is fed all 2^M possible input bits I_i ($i = 1, \ldots, 2^3$ for *Mux* and *Maj*, figure 1). The network output $O^\gamma(I_i, t)$ is then compared to $\varphi(I_i)$ to compute the average number, over inputs and network lifespan, of mistaken outputs:

$$T_1 \equiv \epsilon(\gamma) \;=\; \frac{\sum_t \sum_{i=1}^{2^M} ||\varphi(I_i) - O^\gamma(I_i, t)||}{\tau \cdot 2^M}, \qquad (4)$$

where $||\cdot||$ represents absolute value. We average T_1 over 10 independent realizations of the aging process.

ii Connectivity given by the number of links at $t = 0$ before any damage:

$$T_2 \equiv C(\gamma) \;=\; \sum_\omega (1 - \delta_{\omega,0}), \qquad (5)$$

where $\delta_{\omega,0}$ is Kronecker's delta and the sum runs over all possible weights such that each non-zero connection incurs in one unit cost. This target embodies a metabolic burden entailed by wiring costs.

iii Each network is created with a unique **regenera-**

bioRxiv preprint doi: https://doi.org/10.1101/780163; this version posted September 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

tion rate:

$$T_3 \equiv \rho(\gamma) \ \in \ [\rho^{min}, 1]. \tag{6}$$

$\rho(\gamma)$ specifies the probability with which the network recovers each damaged link at each $t = 1, \ldots, \tau$ of the aging process. A lower bound $\rho^{min} = 0.01 > 0$ was chosen as such minimum regeneration power seems to be always present in living systems [16].

The set Γ of possible networks is combinatorially vast. To explore it, we resort to an evolutionary MOO algorithm (figure 2). For a seed (random) population $\hat{\Gamma}(0) \subset \Gamma$ of networks $\gamma \in \hat{\Gamma}(0)$ we evaluate their performance in each target $\{T_1(\gamma), T_2(\gamma), T_3(\gamma)\}$, select the fittest ones according to a Pareto dominance criterion, and produce diverse offspring through crossover and mutation. This evolves our network ensemble towards the MOO solution Π over generations $g = 0, \ldots, g^{max}$.

Our *target space* (with $\{T_1, T_2, T_3\}$ as axes, figure 2**b**) allows us to compare networks without favoring performance in any T_k over the others: Given $\gamma_i, \gamma_j \in \Gamma$, we say that γ_i *dominates* γ_j (and note it $\gamma_i \prec \gamma_j$) if γ_i is no worse than γ_j in all targets ($T_k(\gamma_i) \leq T_k(\gamma_j) \forall k = 1, 2, 3$) and it is strictly better in at least one target ($\exists k', T_{k'}(\gamma_i) < T_{k'}(\gamma_j)$).

Using these guidelines (following [43]), we conducted an MOO with a population of $N = 480$ networks over $g^{max} = 4000$ generations. At each generation, $\{T_1(\gamma), T_2(\gamma), T_3(\gamma)\}$ were used to calculate dominance scores: $D(\gamma_i, g) \equiv ||\{\gamma_j \in \hat{\Gamma}(g), \gamma_j \prec \gamma_i\}||$ – i.e. the number of designs $\gamma_j \in \hat{\Gamma}(g)$ that dominate $\gamma_i \in \hat{\Gamma}(g)$. Pareto optimal networks have $D(\gamma^{\pi} \in \Pi) = 0$ (but not the other way around). Poor-performing networks soon become dominated by others. We *copied* all γ_i with $D(\gamma_i, g) = 0$ into $\hat{\Gamma}(g+1)$. This *elitist* strategy ensures that we never lose the fittest designs. The half of the population with largest $D(\gamma_i, g)$ is discarded. Networks in the remaining half are combined to bring $\hat{\Gamma}(g+1)$ to its full size ($N = 480$). The resulting children are mutated (see Supplementary Material, section III, for further implementation details). We ran this algorithm 10 times for each (δ, τ) condition. The final population of all 10 runs are combined in a unique set noted $\hat{\Gamma}^{end} \equiv \hat{\Gamma}(g = g^{max})$. The results reported correspond to the Pareto-optimal networks of this merged data set, $\hat{\Pi}^{end} \equiv \{\gamma_k \in \hat{\Gamma}^{end}, D(\gamma_k) = 0\}$ (see Supplementary Material, section IV, to observe all the final results for each (δ, τ) condition). We assume $\hat{\Pi}^{end} \simeq \Pi$, but full convergence cannot be guaranteed.

Figure 2**b** illustrates this for particular (δ, τ) conditions. $\hat{\Pi}^{end}$ (and actually Π itself) includes designs that compute very badly (large $T_1 \equiv \epsilon(\gamma)$) but have been selected because of their negligible regeneration cost and number of links. Pareto dominance offers no principled way to dispose of these networks, even though they would fade away in a biological setting because they plainly fail to function. Interestingly, we observed that computational errors are often associated to deeper structural breakdowns, measured by F_f. Figure 2**b** shows, color-coded, the feasibility $F_f(\gamma)$ of each network. Those with large computational errors often cannot even convey information from input to output. Plotting $\epsilon(\gamma)$ vs $F_f(\gamma)$ (see Supplementary Material, section VII) we noted that this degradation of computation capabilities and network structure follows a logistic curve with a marked threshold $\epsilon_c \sim 0.24$. This value turned out to be similar across realizations of the algorithm and for different (δ, τ) conditions (both for *Mux* and *Maj* functions; see Supplementary Material, figures S18 to S25). We took it as a heuristic limit (horizontal plane in figure 2**b**) to select *acceptably working* designs (we took $\epsilon_c = 0.2$ to be on the safe side). Figure 2**d** shows all networks with a performance better than this threshold ($T_1 \equiv \epsilon(\gamma) < \epsilon_c = 0.2$) in a $C - \rho$ map. (See Supplementary Material, section V, to observe the density plots for each (δ, τ) condition, section I, to check on all the parameters of the model, and section III to check on further implementation details.)

III. RESULTS

A. Tradeoff between connectivity and regeneration

Figure 3 shows abundance of networks $\gamma \in \hat{\Pi}^{end}$ with $\epsilon(\gamma) < \epsilon_c \sim 0.2$ (i.e. Pareto optimal designs that further satisfy the reasonable phenotypic performance ϵ_c, which also entails a persisting structural integrity – large feasibility $F_f(\gamma)$) for a low damage δ condition, both for *Mux* and *Maj* as computational tasks. The plotted abundance of designs across the $\rho - C$ plane captures the tradeoff between connectivity ($T_2 \equiv C(\gamma)$) and regeneration ($T_3 \equiv \rho(\gamma)$). A broad range of optimal solutions compatible with proper functionality is showcased.

Both panels (and further plots in SM) show a higher density of solutions around areas with less connections and higher regeneration (e.g. peak in the upper right corner, figure 3**a**). Graphs labeled 1 and 2 for *Mux* (figure 3**a**) and 6, 7, and 8 for *Maj* (figure 3**b**) illustrate minimal circuits implementing those tasks. The density of designs found with lower regeneration rates (which demands higher connectivity) is notably smaller in comparison. Some graphs (3, 4, and 5 for *Mux*; 9 and 10 for *Maj*) sample this more sparsely occupied region with lower regeneration levels and more densely connected circuits. Trends are shared among the two Boolean functions tested, such as the lower density of solutions in this region of phenotype space, and the resilience that is obtained trough abundant, duplicated links. This hints us at general patterns emerging despite the potential variety in phenotypes imposed by different computational tasks.

bioRxiv preprint doi: https://doi.org/10.1101/780163; this version posted September 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

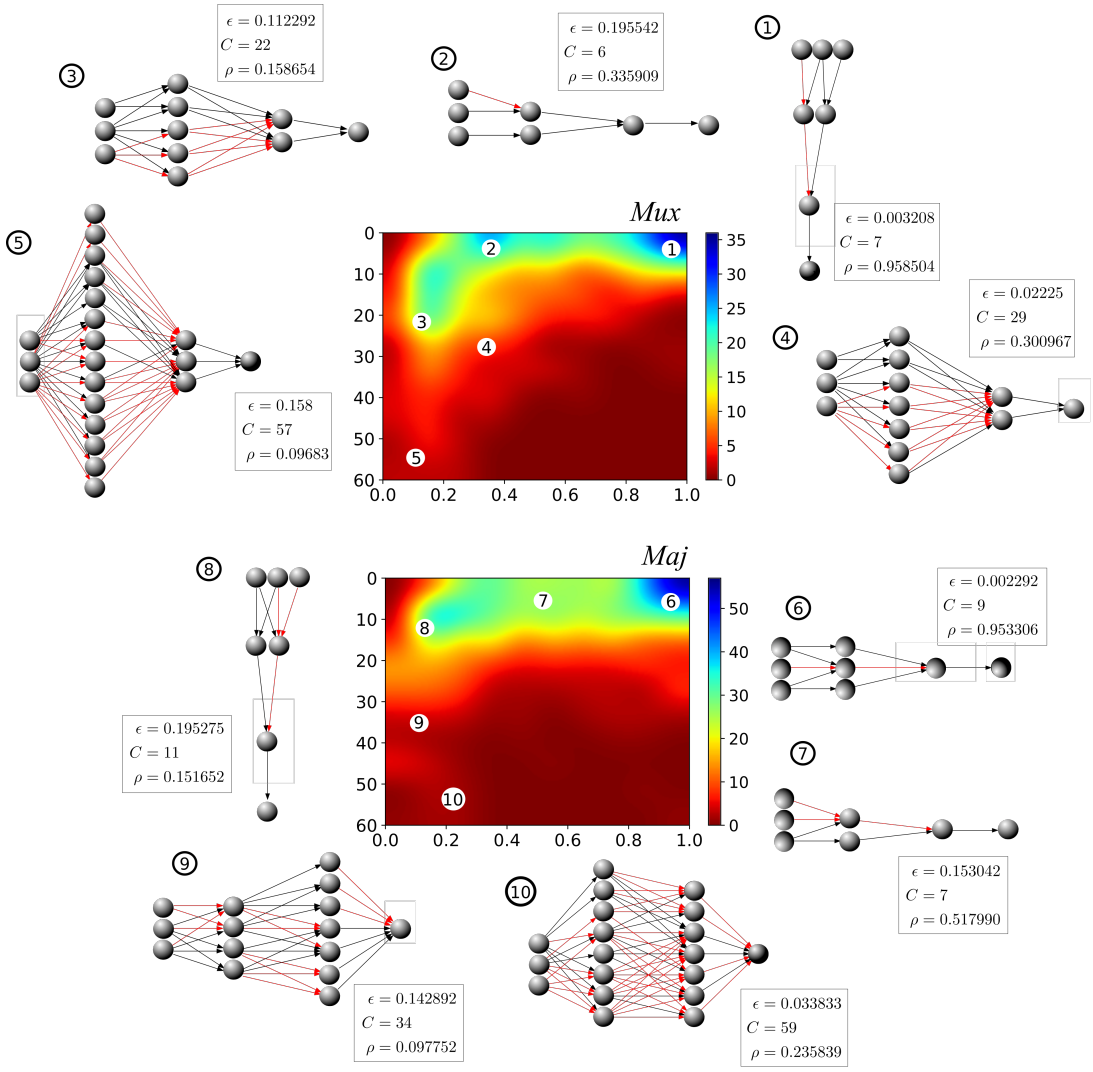

FIG. 3: **Connectivity-regeneration ($C-\rho$) tradeoff as a means to achieve reliable computation.** Two selected density plots for *Mux* (top) and *Maj* (bottom) of solutions $\gamma \in \hat{\Gamma}^{end}$ with reliable computation ($\epsilon(\gamma) < \epsilon_c$) throughout the $\rho - C$ plane. Alongside, representative networks from characteristic regions of phenotype space. Performance of these networks in all three targets is shown (enclosed squares). Results obtained for $\delta = 0.04$, $\tau = 300$ for *Mux* and $\delta = 0.02$ and $\tau = 1500$ for *Maj*. (See SM for other values.)

B. Network lifespan and external damage act as evolutionary pressures

As discussed above, our aim is to asses the influence of additional distinct features on our evolutionary framework, namely the lifespan of networks and the damage inflicted to their links during the aging process. The former offers a window to the effects of life length (τ) on the selection process whereas the latter, implemented by δ, allows including the stochastic perturbations that hamper proper computation. What is the impact of lifetime (τ) or damage rates (δ) in the optimal space of solutions resulting from our model?

We have found that, for a fixed damage rate, longer agent life times act as an evolutionary pressure, such that the algorithm better explores the optimal tradeoff by pushing harder our evolving population against it. The same happens in the opposite case, e.g. increasing damage rates for fixed life times. To capture this, we compute the amount of non-dominated versus dominated solutions in $\hat{\Gamma}^{end}$:

$$\begin{aligned}\Psi(\delta,\tau) &= \frac{||\{\gamma_k \in \hat{\Gamma}^{end}, D(\gamma_k) > 0\}||}{||\{\gamma_k \in \hat{\Gamma}^{end}, D(\gamma_k) = 0\}||} \\ &= \frac{||\hat{\Gamma}^{end} - \hat{\Pi}^{end}||}{||\hat{\Pi}^{end}||},\end{aligned} \tag{7}$$

where $D(\cdot)$ is again the dominance score. Each of our evolution takes place under fixed (δ, τ) conditions. Those instances that result in larger Ψ retain more network

bioRxiv preprint doi: https://doi.org/10.1101/780163; this version posted September 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

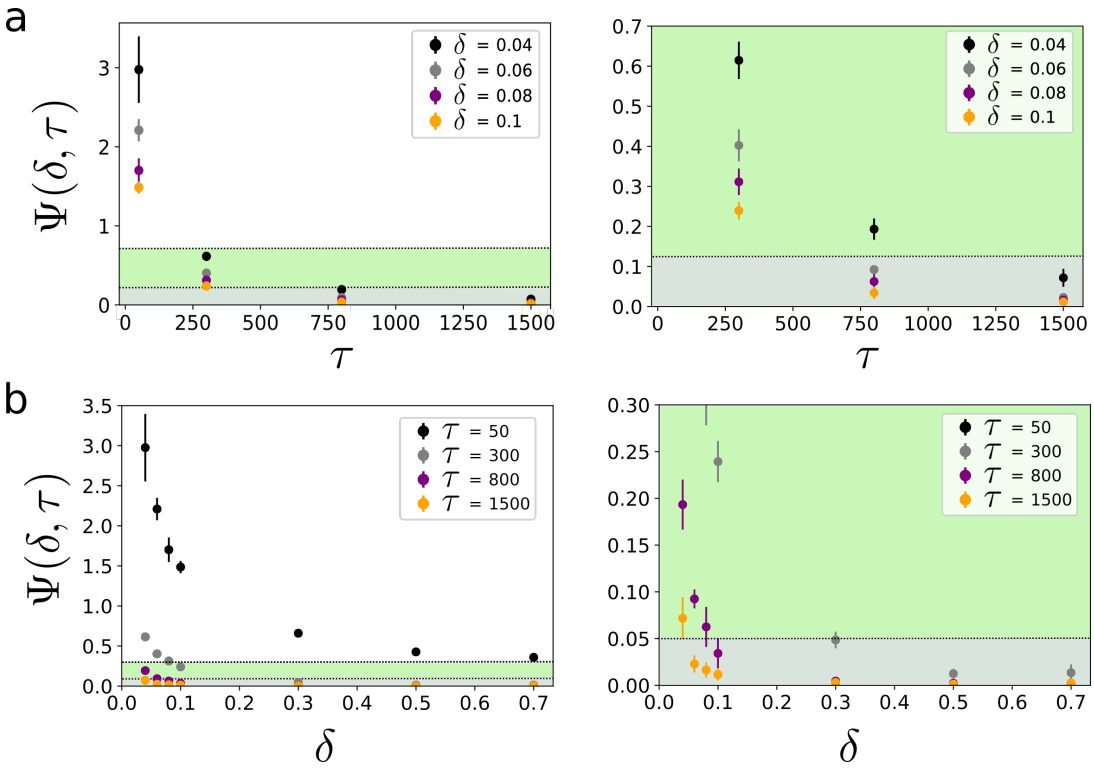

FIG. 4: **Life time τ and damage rate δ act as evolutionary pressures.** (a) Ratio $\Psi(\delta,\tau)$ between the number of non-Pareto and Pareto-optimal solutions in $\hat{\Gamma}^{end}$ as a function of the τ for fixed values of damage rate δ (for *Mux*). Ψ decreases for increasing τ. Only a sample of δ values (0.04, 0.06, 0.08 and 0.1) is shown to help visualization (other conditions, in SM section VI, also for *Maj*; all results in Supplementary Material support the conclusions in the main text). (b) Same ratio for fixed values of τ and varying δ. Again, Ψ drops as δ increases. Decreases in Ψ capture a higher evolutionary pressure towards Pareto optimality.

designs that are suboptimal (in the Pareto sense) with respect to other surviving designs. This is, such (δ,τ) settings are less strict during selection, such that networks that perform relatively worse in all targets simultaneously are still retained. Opposed to this, (δ,τ) values resulting in a smaller Ψ are more severe regarding Pareto optimality selection – i.e. evolutionary pressure to select Pareto optimal solutions is higher. See Fig. 4a to check the effect of fixed damage rates and increasing life times, and Fig. 4b to illustrate the effect of fixed lifetime and increasing damage. However, both effects are present in both plots. (Check Supplementary Material, section VI, to check the rest of parameter values either for *Mux* and *Maj*.)

Which are the reasons for such observation? Despite both network lifespan and damage rate exert a pressure of the same nature (as measured by $\Psi(\delta,\tau)$, the ultimate reasons for such observation might be different. An explanation for large τ as an evolutionary pressure for Pareto optimality could lay in the fact that longer life times are synonym of dealing with more extended and numerous threats entailing a larger information retrieval from the environment. Therefore, improvements during evolutionary time can have a larger impact on this longer living populations compared to shorter living ones. Regarding the influence of the degree of damage inflicted to networks (δ), the reason for such evolutionary pressure is probably rooted to the inherent harsher survival conditions imposed in larger damage regimes.

Importantly, the interplay of both features might also be acting as an evolutionary pressure (as measured by $\Psi(\delta,\tau)$). As presented in section II C, damage rates and network lifespans can be collapsed into a dimensionless ratio, $r_\tau \equiv \delta\cdot\tau \sim \tau/\tau_{link}$, as the lifespan of the connections τ_{link} grows monotonously with $1/\delta$, thus capturing a relationship between time-scales proper of whole organisms versus those of its parts. An increase of this ratio also entails lower values of $\Psi(\delta,\tau)$, meaning that an increase in the difference between the lifespan of the components versus that of the organism might be playing a role in the described observations.

C. Damage rates influence the overall shape of the optimal tradeoff

The resulting $\hat{\Pi}^{end}$ embody the optimal tradeoff between all targets involved. This optimal tradeoff can be visualized if plotted in target space, where its shape can be different for each (δ,τ) fixed conditions. Our exhaus-

bioRxiv preprint doi: https://doi.org/10.1101/780163; this version posted September 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

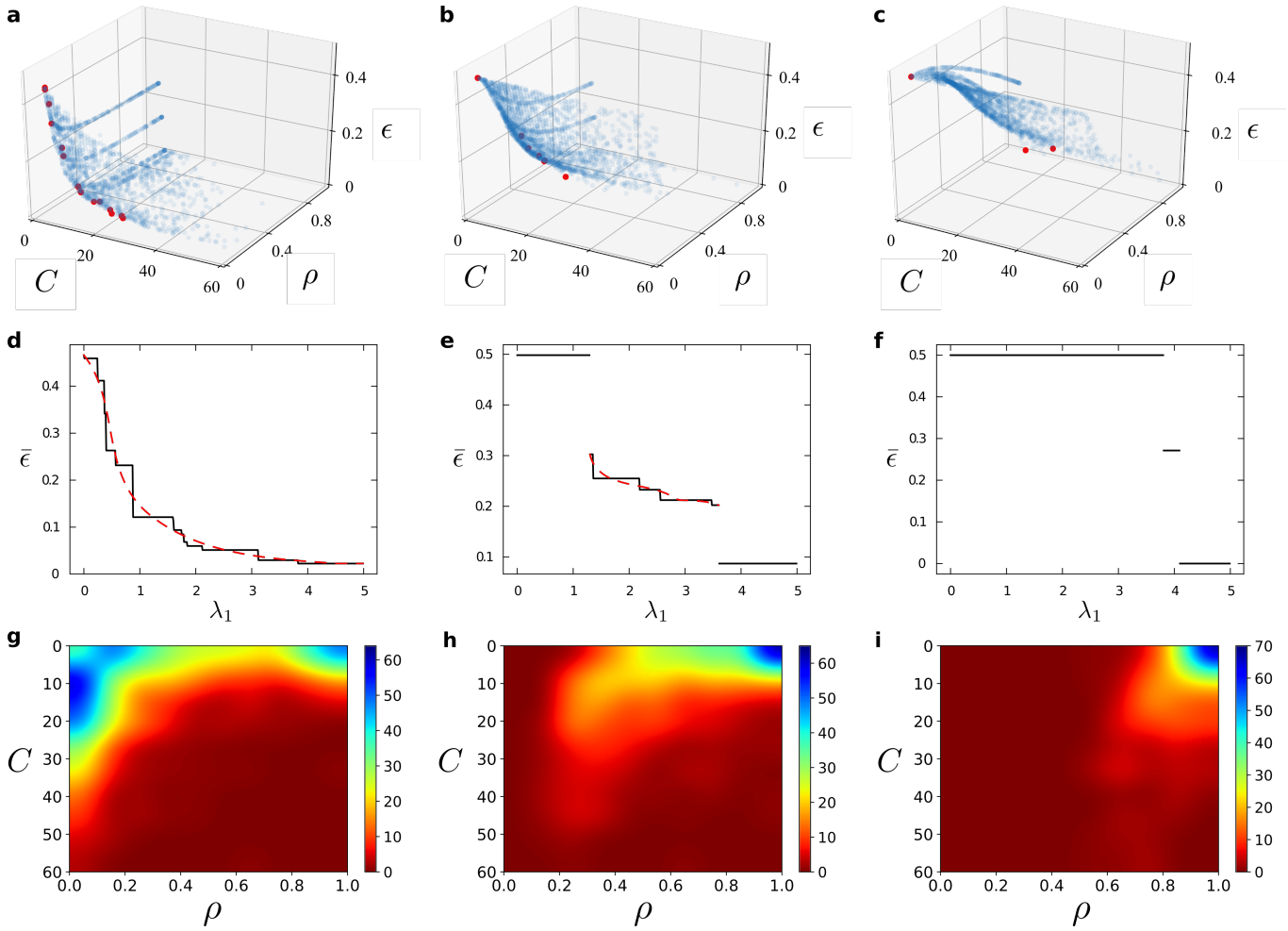

FIG. 5: **Overall shape of the optimal tradeoff and accessibility of phenotype space.** Pareto optimal strategies $\{\gamma \in \hat{\Pi}^{end}\}$ under fixed conditions ((a) $\delta = 0.01$, $\tau = 1500$; (b) $\delta = 0.1$, $\tau = 300$; (c) $\delta = 0.7$, $\tau = 50$) plotted in *target space*. Each blue dot represents $\{T_1(\gamma), T_2(\gamma), T_3(\gamma)\}$ for a given $\gamma \in \hat{\Pi}^{end}$. The shape of each embedding surface determines how accessible phenotype space is if minimizers of a global utility function $\Omega = \sum \lambda_i T_i$ were sought. Red dots represent such global optima as $\lambda_2 = 0.009$ and $\lambda_3 = 2$ are kept fixed and $\lambda_1 \in [0, 5]$. Increasing damage ((a) $\delta = 0.01$, (b) $\delta = 0.1$, (c) $\delta = 0.7$) results in increasingly more rugged tradeoffs. Smooth tradeoffs (a) are sampled more evenly by global optimizers, so that successive global optima are similar to each other. Whatever property we plot of them, e.g. their computational error $\epsilon(\gamma)$ (d, solid black lines), varies relatively smoothly with λ_1. Some discreteness remains due to the numerical nature of our experiments (dashed red lines illustrate the underlying continuous dependency). The corresponding density plot (g) with solutions fulfilling $\epsilon(\gamma) < \epsilon_c$ shows the tradeoff $\rho - C$ available for the low damage regime (already shown in figure 3). Tradeoffs with cavities (b, c) leave some phenotypes unsampled by global optima. These often take leaps in phenotype space, resulting in sudden jumps in any measured property of such solutions as a function of λ_1 (e, f). These plots show how a heightened damage rate has an ability to hamper access to existing phenotypes and fracture the continuity of global optimal designs. The corresponding density plots (h, i) with solutions fulfilling $\epsilon(\gamma) < \epsilon_c$ show the constraints imposed by increasing damage regimes in the $C - \rho$ projection. Crucially, these available designs can be further constrained by the fracture of the continuity of global optimal designs. See S2-S7 in SM for further evidence linking ruggedness of the optimal tradeoff to increased damage rate, both for *Mux* and *Maj* functions.

tive exploration of (δ, τ) combinations (see SM section IV) shows a clear overall change in the shape of this optimal tradeoff as damage increases – a change, furthermore, that has consequences in terms of phenotype space accessibility and exploration as we discuss below. Figure 5**a-c** shows networks in $\hat{\Pi}^{end}$ for increasing damage regimes. Designs with $\epsilon(\gamma) > \epsilon_c$ are retained here. $\hat{\Pi}^{end}$ appears smooth for the low damage regime ($\delta = 0.01$), meaning that the optimal tradeoff does not present large cavities or singular points when plotted in target space (figure 5**a**). This surface becomes more rugged (i.e. its curvature changes, thus generating cavities) for increasing damage rates ($\delta = 0.1, 0.7$, figure 5**b, c**).

This varying ruggedness of optimal tradeoffs can tell us something about how accessible our space of optimal solutions is. We could weight all targets linearly

bioRxiv preprint doi: https://doi.org/10.1101/780163; this version posted September 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

into a global optimization function: $\Omega(\gamma, \Lambda) = \lambda_1 T_1(\gamma) + \lambda_2 T_2(\gamma) + \lambda_3 T_3(\gamma)$, where $\Lambda \equiv \{\lambda_1, \lambda_2, \lambda_3\}$ represent explicit evolutionary biases towards a specific target. For example, a large λ_2 versus low $\lambda_{1,3}$ indicates that, in a given environment, minimizing the initial metabolic cost of links is outstandingly important. Giving values to the λ_k could thus define a specific set of external environmental pressures. Then the minimization of $\Omega(\gamma, \Lambda)$ selects one single solution $\bar{\gamma}(\Lambda) \in \hat{\Pi}^{end}$ out of the Pareto optimal tradeoff. This is, the imposition of such specific constraints would force evolution towards a determined region of phenotype space.

Cavities and singular points in Π ($\sim \hat{\Pi}^{end}$) have been linked to phase transitions [37–41] and critical phenomena [39, 41] that arise as the λ_k are varied – i.e. as potential biases in a niche change (perhaps over time, or perhaps because the species has left that niche). A phase transition in our problem would indicate that certain network designs are persistently stable under a range of external evolutionary conditions, and that a switch from a network design to another would sometimes be drastic even if those external conditions would vary just slightly.

Such cavities and singular points are absent for the low δ example in figure 5**a** (the same regime was shown in figure 3), meaning that the whole phenotypic space of nervous system maintenance strategies can be smoothly visited as evolutionary pressures vary. To illustrate this we have set λ_2 and λ_3 to a fixed value while varying $\lambda_1 \in [0, 5]$. This means that we sample evolutionary pressures that initially disregard good computation but progress towards situations in which computing correctly becomes more pressing. Network designs $\bar{\gamma} \in \hat{\Pi}^{end}$ that minimize $\Omega(\gamma, \Lambda)$ for the range of λ_1 are displayed in red in figure 5**a**. Figure 5**d** shows the computational error for these absolute optima $\bar{\epsilon} \equiv \epsilon(\bar{\gamma}(\Lambda))$ as a function of λ_1. The numerical nature of our experiments introduces some unavoidable discreteness in $\bar{\epsilon}(\lambda_1)$; but overall we can see how $\hat{\Pi}^{end}$ is rather smoothly sampled (as compared to the next cases) and optimal solutions progress parsimoniously into each other as external evolutionary pressures change.

Additionally, figure 5**a** shows how $\hat{\Pi}^{end}$ becomes virtually flat for $T_1 \equiv \epsilon \to 0$. Decrementally small improvements in computation can only be achieved by increasingly larger investment on $T_2 \equiv C$ and $T_3 \equiv \rho$. Perfect computation ($\epsilon = 0$) could be reached ultimately. But this plot reveals a decreasing return for low errors. Organisms eventually need exaggerated investments to achieve negligible computational improvements. This is reflected by a sparse sampling of those costly areas of phenotype space, also reflected in the plots of abundance of designs across the $\rho - C$ plane (figure 3).

Optimal tradeoffs present cavities for the examples in larger damage regimes (figures 5**b, c**). Red dots show absolute optima as the same values of Λ are sampled. The cavities in $\hat{\Pi}^{end}$ entail that a lower number of different absolute optima are recovered for $\delta = 0.1$, and even less for $\delta = 0.7$ under the same procedure. This is so because certain phenotypes remain optimal over a wide range of λ_1, thus preventing us from visiting part of phenotypic space. Furthermore, when such paramount designs stop being global optima, the new preferred network is found far apart in $\hat{\Pi}^{end}$. A small change in λ_1 would then demand a prompt yet drastic adaptation to accommodate the new best. This is reflected in $\bar{\epsilon}(\lambda_1)$ plots (figures 5**e-f**) through huge improvements in performance as the bias towards minimizing the computational error (λ_1) increases.

Notice that in these cases (figures 5**b, c**) the overall shape of the Pareto optimal front is also constraining to a much reduced area the set of designs which retain reliable computations ($\epsilon < \epsilon_c$, see the net effect on the $\rho - C$ density plots in figures 5**h, i**), if compared with the low damage rate regime shown in figures 3 and 5**a,g**. Thus, increased damage (besides constraining the available phenotypic space due to emerging phase transitions) also results in a much more narrower space when looking only at *acceptably performing* designs, disregarding of any global optimization of Ω.

Importantly, it cannot be discarded that either the network lifespan (τ) or the interplay between it and the lifespan of the connections (τ_{link}), measured through r_τ, may also have an influence on the change in the overall shape of optimal tradeoffs ($r_\tau = 15, 30, 35$ in figures 5**a,b,c**, respectively). However, the observations retrieved from the set of tested parameters (see Supplementary Material, section IV) show that the clear driver of the observed effect is the increase in damage rates regimes.

IV. DISCUSSION

In this paper we attempt to provide insights on which are the evolutionary pressures or drivers that may underpin the evolution of nervous systems and its regeneration capabilities. Due to the generality of the model, we do not aim to answer specific questions but rather extract general principles which might pave the way to future research directions. The simplifications we make are considerable, starting from the use of artificial feedforward neural networks to represent a nervous system. Regarding the dynamics of the system, the so-called aging process to which our networks are exposed and their subsequent response is a simplification of the processes of axonogenesis and neurogenesis [16] observed in real nervous systems. While only axonogenesis is explicitly incorporated in the model, neurogenesis can be considered to be present as a secondary response linked to the recovery of connections (whose loss can cause the breakdown of neuron functionality). On the other hand, it is also known that regeneration response is dependent on anatomical location and type of damage [16], but such a feature has not been incorporated in any way, being another simplification of the model. And yet our schematic framework retains what we think are some key factors to understand the evolutionary drivers of neural systems

bioRxiv preprint doi: https://doi.org/10.1101/780163; this version posted September 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

maintenance. We can thus disentangle some pressures at play as well as isolate relevant factors for future studies. This would probably be more difficult with more complicated models.

First, our results show that, in the domain of reliable computation, a tradeoff might be at play between high density of connections and high regeneration capabilities. This tradeoff shapes the phenotypic space of possible designs for network maintenance, and it reminds us of the variety of such strategies in the natural world. The region of phenotype space with higher connectivity presumably achieves robust computation through specific redundant connections or degenerate mechanisms [44, 45] The alternative, which relies on large regeneration rates, is found in the region of phenotype space where networks are smaller. We observe that these are also the networks preferentially explored by our MOO algorithm. This is so even if, intuitively, a higher number of connections could result in a potentially larger amount of different architectures that solve a same task.

This suggests that regeneration can be a more reliable strategy – at least at the scales explored. Strategies betting on duplicate pathways might need to deal, e.g., with emerging interactions between surviving connections as faulty ones are removed. Such problems could be bypassed by alternate computational paradigms – e.g. distributed computation [46, 47] or reservoir computing [48–52]. The latter can take explicit advantage of such emergent behaviors. Such alternatives could unlock further regions of phenotypic space with a high number of similar solutions. But such computational strategies depend crucially on huge numbers of units, and might be unreachable for our experiments; thus returning us to the observed bias towards regeneration.

Secondly, we have found that both the extent of external damage and the network lifespan are acting as active evolutionary pressures over the whole population so that it eventually contains more Pareto optimal individuals. The increase of both factors gives rise to more strict evolutionary scenarios, in which the pressure to select Pareto optimal solutions is higher, resulting in more diverse optimal strategies (better exploration of the optimal tradeoffs). Interestingly, the relationship r_τ between organismal lifetimes (τ) and the timescale of its component parts ($\tau_{link} \sim 1/\delta$) is also suggested to be acting as an evolutionary pressure towards this direction. Focus on this ratio becomes a very enticing research avenue if we wonder at what level (organismal versus component part) can a Darwinian process store the information gathered as evolution proceeds.

Third, from the observations retrieved we have found that regimes of increasing damage (which can be considered an ecological factor) result in more rugged tradeoffs. Such tradeoffs present cavities, which is the characteristic signature of first order phase transitions [37–41]. Under varying evolutionary biases, the presence of such transitions can result in history dependency effects similar to hysteresis [53]. In such phenomena, evolutionary pressures confronted by a species would vary just slightly and yet a preferred global optimum would change drastically. The species might need to adapt swiftly and retain suboptimal aspects due to frozen accidents, thus resulting in increased evolutionary path dependency.

All of this suggests that i) discrete phenotypic space, ii) drastic changes expected under varying external evolutionary pressures, iii) phenotypic space becoming less accessible, and iv) heightened path dependency in evolution should all become more prominent as the external erosion of our system is higher. Overall, this would mean that higher damage rates could induce organisms to commit to specific nervous systems maintenance strategies, potentially renouncing an ability to switch options with relative ease. Damage could change swiftly as living beings migrate to more benign environments or are suddenly locked on harsher conditions. Similarly to an inherent shorter life-time of component parts, a heightened external damage rate has the ability to harshen the strictness of the selection process (again, in a Pareto optimality sense; thus implying population-wide phenomena) and also of promptly shifting the shape of the optimal tradeoff. Future work will be required to explore these results in a more general context of multicellularity (natural and synthetic) where cognitive complexity is a well-defined dimension [54].

Acknowledgments

The authors thank the members of the Complex Systems Lab for useful discussions, specially Jordi Piñero. This work has been supported by an ERC Advanced Grant Number 294294 from the EU seventh framework program (SYNCOM), by the Botín Foundation by Banco Santander through its Santander Universities Global Division, a MINECO grant FIS2015-67616 fellowship cofunded by FEDER/UE, by the Universities and Research Secretariat of the Ministry of Business and Knowledge of the Generalitat de Catalunya and the European Social Fund, and by the Santa Fe Institute.

V. REFERENCES

[1] E. Jablonka and M. J. Lamb, *Journal of Theoretical Biology*, 2006, **239**, 236–246.

[2] D. Bickerton, *Language and species*, University of Chicago Press, 2018.

bioRxiv preprint doi: https://doi.org/10.1101/780163; this version posted September 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[3] J. Von Neumann, *Automata studies*, 1956, **34**, 43–98.
[4] S. Winograd and J. D. Cowan, *Reliable computation in the presence of noise*, Mit Press Cambridge, Mass., 1963.
[5] H. Barlow and W. R. Levick, *The Journal of physiology*, 1965, **178**, 477–504.
[6] L. A. Gavrilov and N. S. Gavrilova, *Journal of theoretical Biology*, 2001, **213**, 527–545.
[7] P. Ferretti, *European Journal of Neuroscience*, 2011, **34**, 951–962.
[8] S. B. Moffet, *Nervous system regeneration in the invertebrates*, Springer Science & Business Media, 2012, vol. 34.
[9] M. Levin, *Trends in cell biology*, 2007, **17**, 261–270.
[10] M. Levin, Seminars in cell & developmental biology, 2009, pp. 543–556.
[11] A.-S. Tseng, W. S. Beane, J. M. Lemire, A. Masi and M. Levin, *Journal of Neuroscience*, 2010, **30**, 13192–13200.
[12] P. S. Eriksson, E. Perfilieva, T. Björk-Eriksson, A.-M. Alborn, C. Nordborg, D. A. Peterson and F. H. Gage, *Nature medicine*, 1998, **4**, 1313.
[13] H. van Praag, A. F. Schinder, B. R. Christie, N. Toni, T. D. Palmer and F. H. Gage, *Nature*, 2002, **415**, 1030.
[14] M. A. Hack, A. Saghatelyan, A. de Chevigny, A. Pfeifer, R. Ashery-Padan, P.-M. Lledo and M. Götz, *Nature neuroscience*, 2005, **8**, 865.
[15] G. Zupanc, *Journal of Comparative Physiology A*, 2006, **192**, 649.
[16] E. M. Tanaka and P. Ferretti, *Nature Reviews Neuroscience*, 2009, **10**, 713.
[17] B. L. Chen, D. H. Hall and D. B. Chklovskii, *Proceedings of the National Academy of Sciences*, 2006, **103**, 4723–4728.
[18] D. B. Chklovskii, T. Schikorski and C. F. Stevens, *Neuron*, 2002, **34**, 341–347.
[19] A. Raj and Y.-h. Chen, *PloS one*, 2011, **6**, e14832.
[20] G. F. Miller, P. M. Todd and S. U. Hegde, ICGA, 1989, pp. 379–384.
[21] P. J. Angeline, G. M. Saunders and J. B. Pollack, *IEEE transactions on Neural Networks*, 1994, **5**, 54–65.
[22] X. Yao, *Proceedings of the IEEE*, 1999, **87**, 1423–1447.
[23] D. Floreano, P. Dürr and C. Mattiussi, *Evolutionary Intelligence*, 2008, **1**, 47–62.
[24] K. O. Stanley and R. Miikkulainen, *Evolutionary computation*, 2002, **10**, 99–127.
[25] K. O. Stanley, D. B. D'Ambrosio and J. Gauci, *Artificial life*, 2009, **15**, 185–212.
[26] J.-P. Eckmann, O. Feinerman, L. Gruendlinger, E. Moses, J. Soriano and T. Tlusty, *Physics Reports*, 2007, **449**, 54–76.
[27] G. Székely, *Acta physiologica Academiae Scientiarum Hungaricae*, 1965, **27**, 285.
[28] G. S. Stent, W. B. Kristan, W. O. Friesen, C. A. Ort, M. Poon and R. L. Calabrese, *Science*, 1978, **200**, 1348–1357.
[29] W. O. Friesen and G. S. Stent, *Annual review of biophysics and bioengineering*, 1978, **7**, 37–61.
[30] S.-i. Amari, *Biological cybernetics*, 1977, **27**, 77–87.
[31] J. Clune, J.-B. Mouret and H. Lipson, *Proceedings of the Royal Society b: Biological sciences*, 2013, **280**, 20122863.
[32] D. Floreano, S. Nolfi and F. Mondada, *Advances in the evolutionary synthesis of intelligent agents*, 2001, 273–306.
[33] R. Pfeifer, M. Lungarella and F. Iida, *science*, 2007, **318**, 1088–1093.
[34] C. Coello, S. De Computación and C. Zacatenco, *IEEE computational intelligence magazine*, 2006, **1**, 28–36.
[35] P. Schuster, *Complexity*, 2012, **18**, 5–7.
[36] L. F. Seoane *et al.*, *Ph.D. thesis*, Universitat Pompeu Fabra, 2016.
[37] L. F. Seoane and R. V. Solé, *arXiv preprint arXiv:1310.6372*, 2013.
[38] L. F. Seoane and R. Solé, *Physical Review E*, 2015, **92**, 032807.
[39] L. F. Seoane and R. Solé, *arXiv preprint arXiv:1510.08697*, 2015.
[40] L. F. Seoane and R. Solé, in *Proceedings of ECCS 2014*, Springer, 2016, pp. 259–270.
[41] L. F. Seoane and R. Solé, *Scientific reports*, 2018, **8**, year.
[42] R. Rojas, *Neural networks: a systematic introduction*, Springer Science & Business Media, 2013.
[43] A. Konak, D. W. Coit and A. E. Smith, *Reliability Engineering & System Safety*, 2006, **91**, 992–1007.
[44] G. Tononi, O. Sporns and G. M. Edelman, *Proceedings of the National Academy of Sciences*, 1999, **96**, 3257–3262.
[45] G. M. Edelman and J. A. Gally, *Proceedings of the National Academy of Sciences*, 2001, **98**, 13763–13768.
[46] J. Macía, F. Posas and R. V. Solé, *Trends in biotechnology*, 2012, **30**, 342–349.
[47] S. Regot, J. Macia, N. Conde, K. Furukawa, J. Kjellén, T. Peeters, S. Hohmann, E. De Nadal, F. Posas and R. Solé, *Nature*, 2011, **469**, 207.
[48] H. Jaeger, *Bonn, Germany: German National Research Center for Information Technology GMD Technical Report*, 2001, **148**, 13.
[49] W. Maass, T. Natschläger and H. Markram, *Neural computation*, 2002, **14**, 2531–2560.
[50] H. Jaeger, W. Maass and J. Principe, 2007.
[51] D. Verstraeten, B. Schrauwen, M. dHaene and D. Stroobandt, *Neural networks*, 2007, **20**, 391–403.
[52] M. Lukoševičius, H. Jaeger and B. Schrauwen, *KI-Künstliche Intelligenz*, 2012, **26**, 365–371.
[53] R. Solé, *Press. Princeton*, 2011.
[54] A. Ollé-Vila, S. Duran-Nebreda, N. Conde-Pueyo, R. Montáñez and R. Solé, *Integrative Biology*, 2016, **8**, 485–503.