How evolution learns to generalise: Using the principles of learning theory to understand the evolution of developmental organisation.

Kostas Kouvaris[1,*], Jeff Clune[2], Louis Kounios[1], Markus Brede[1], Richard A. Watson[1]
1 ECS, University of Southampton, Southampton, UK
2 University of Wyoming, Laramie, Wyoming, USA
∗ E-mail: kk6g11@soton.ac.uk

Abstract

One of the most intriguing questions in evolution is how organisms exhibit suitable phenotypic variation to rapidly adapt in novel selective environments. Such variability is crucial for evolvability, but poorly understood. In particular, how can natural selection favour developmental organisations that facilitate adaptive evolution in previously unseen environments? Such a capacity suggests foresight that is incompatible with the short-sighted concept of natural selection. A potential resolution is provided by the idea that evolution may discover and exploit information not only about the particular phenotypes selected in the past, but their underlying structural regularities: new phenotypes, with the same underlying regularities, but novel particulars, may then be useful in new environments. If true, we still need to understand the conditions in which natural selection will discover such deep regularities rather than exploiting 'quick fixes' (i.e. fixes that provide adaptive phenotypes in the short term, but limit future evolvability). Here we argue that the ability of evolution to discover such regularities is formally analogous to learning principles, familiar in humans and machines, that enable generalisation from past experience. Conversely, natural selection that fails to enhance evolvability is directly analogous to the learning problem of over-fitting and the subsequent failure to generalise. We support the conclusion that evolving systems and learning systems are different instantiations of the same algorithmic principles by showing that existing results from the learning domain can be transferred to the evolution domain. Specifically, we show that conditions that alleviate over-fitting in learning systems successfully predict which biological conditions (e.g., environmental variation, regularity, noise or a pressure for developmental simplicity) enhance evolvability. This equivalence provides access to a well-developed theoretical framework from learning theory that enables a characterisation of the general conditions for the evolution of evolvability.

Author Summary

A striking feature of evolving organisms is their ability to acquire novel characteristics that help them adapt in new environments. The origin and the conditions of such ability remain elusive and is a long-standing question in evolutionary biology. Recent theory suggests that organisms can evolve designs that help them generate novel features that are more likely to be beneficial. Specifically, this is possible when the environments that organisms are exposed to share common regularities. However, the organisms develop robust designs that tend to produce what had been selected in the past and might be inflexible for future environments. The resolution comes from a recent theory introduced by Watson and Szathmry that suggests a deep analogy between learning and evolution. Accordingly, here we utilise learning theory to explain the conditions that lead to more evolvable designs. We successfully demonstrate this by equating evolvability to the way humans and machines generalise to previously-unseen situations. Specifically, we show that the same conditions that enhance generalisation in learning systems have biological analogues and help us understand why environmental noise and the reproductive and maintenance costs of gene-regulatory connections can lead to more evolvable designs.

Introduction

Linking the Evolution of Evolvability with Generalisation in Learning Systems

Explaining how organisms adapt in novel selective environments is central to evolutionary biology [1–5]. Living organisms are both robust and capable of change. The former property allows for stability and reliable functionality against genetic and environmental perturbations, while the latter provides flexibility allowing for the evolutionary acquisition of new potentially adaptive traits [5–9]. This capacity of an organism to produce suitable phenotypic variation to adapt to new environments is often identified as a prerequisite for *evolvability*, i.e. the capacity for adaptive evolution [7, 10, 11]. It is thus important to understand the underlying variational mechanisms that enable the production of adaptive phenotypic variation [6, 7, 12–18].

Phenotypic variations are heavily determined by intrinsic tendencies imposed by the genetic and the developmental architecture [18–21]. For instance, developmental biases may permit high variability for a particular phenotypic trait and limited variability for another, or cause certain phenotypic traits to covary [6, 15, 22–26]. Developmental processes are themselves also shaped by previous selection. As a result, we may expect that past evolution could adapt the distribution of phenotypes explored by future natural selection to amplify promising variations and avoid less useful ones by evolving developmental architectures that are predisposed to exhibit effective adaptation [10, 13]. Selection though cannot favour traits for benefits that have not yet been realised. Moreover, in situations when selection can control phenotypic variation, it nearly always reduces such variation because it favours canalisation over flexibility [23, 27–29].

Developmental canalisation may seem to be intrinsically opposed to an increase in phenotypic variability. Some, however, view these notions as two sides of the same coin, i.e., a predisposition to evolve some phenotypes more readily goes hand in hand with a decrease in the propensity to produce other phenotypes [8,30,31]. Kirschner and Gerhart integrated findings that support these ideas under the unified framework of *facilitated variation* [8,32]. Similar ideas and concepts include the *variational properties* of the organisms [13], the *self-facilitation* of evolution [20] and evolution as *tinkering* [33] and related notions [6, 7, 10, 12]. In facilitated variation, the key observation is that the intrinsic developmental structure of the organisms biases both the amount and the direction of the phenotypic variation. Recent work in the area of facilitated variation has shown that multiple selective environments were necessary to evolve evolvable structures [25, 27, 34–36]. When selective environments contain underlying structural regularities, it is possible that evolution learns to limit the phenotypic space to regions that are evolutionarily more advantageous, promoting the discovery of useful phenotypes in a single or a few mutations [35, 36]. But, as we will show, these conditions do not necessarily enhance evolvability in novel environments. Thus the general conditions which favour the emergence of adaptive developmental constraints that enhance evolvability are not well-understood.

To address this we study the conditions where evolution by natural selection can find developmental organisations that produce what we refer to here as *generalised phenotypic distributions* — i.e., not only are these distributions capable of producing multiple distinct phenotypes that have been selected in the past, but they can also produce novel phenotypes from the same family. Parter et al. have already shown that this is possible in specific cases studying models of RNA structures and logic gates [34]. Here we wish to understand more general conditions under which, and to what extent, natural selection can enhance the capacity of developmental structures to produce suitable variation for selection in the future. We follow previous work on the evolution of development [25] through computer simulations based in gene-regulatory network (GRN) models. Many authors have noted that GRNs share common functionality to artificial neural networks [25, 37–40]. Watson et al. demonstrated a further result, more important to our purposes here; that the way regulatory interactions *evolve* under natural selection is mathematically equivalent to the way neural networks *learn* [25]. During evolution a GRN is capable of learning a memory of multiple phenotypes that were fit in multiple past selective environments by internalising their statistical correlation structure into its ontogenetic interactions, in the same way that learning neural networks store and recall training patterns. Phenotypes that were fit in the past can then be recreated by the network spontaneously (under genetic drift without selection) in the future or as a response to new selective environments that are partially similar to past environments [25]. An important aspect of the evolved systems mentioned above is modularity. Modularity has been a key feature of work on evolvability [6, 29, 41, 42] aiming to facilitate variability that respects the natural decomposable structure of the selective environment, i.e., keep the things together that need to be kept together and separate the things that are independent [6, 12, 20, 41]. Accordingly, the system can perform a simple form of generalisation by separating knowledge from the context in which it was originally observed and re-deploying it in new situations.

Here we show that this functional equivalence between learning and evolution predicts the evolutionary conditions that enable the evolution of generalised developmental organisations. We test this analogy between learning and evolution by testing its predictions. Specifically, we resolve the tension between canalisation of phenotypes that have been successful in past environments and anticipation of phenotypes that are fit in future environments by recognising that this is equivalent to prediction in learning systems. Such predictive ability follows simply from the ability to represent structural regularities in previously seen observations (i.e., the training set) that are also true in the yet-unseen ones (i.e., the test set). In learning systems, such generalization is commonplace and not considered mysterious. But it is also understood that successful generalisation in learning systems is not for granted and requires certain well-understood conditions. We argue here that understanding the evolution of development is formally analogous to model learning and can provide useful insights and testable hypotheses about the conditions that enhance the evolution of evolvability under natural selection [42, 43]. Thus, in recognising that learning systems do not really ‘see into the future’ but can nonetheless make useful predictions by generalising past experience, we demystify the notion that short-sighted natural selection can produce novel phenotypes that are fit for previously-unseen selective environments and, more importantly, we can predict the general conditions where this is possible. This functional equivalence between learning and evolution produces many interesting, testable predictions (Table 1).

In particular, the following experiments show that techniques that enhance generalisation in machine learning correspond to evolutionary conditions that facilitate generalised phenotypic distributions and hence increased evolvability. Specifically, we describe how well-known machine learning techniques, such as learning with noise and penalising model complexity, that improve the generalisation ability of learning models have biological analogues and can help us understand how noisy selective environments and the direct selection pressure on the reproduction cost of the gene regulatory interactions can enhance evolvability in gene regulation networks. This is a much more sophisticated and powerful form of generalisation than previous notions that simply extrapolate previous experience. The system does not merely extend its learned behaviour outside its past 'known' domain. Instead, we are interested in situations where the system can create new knowledge by discovering and systematising emerging patterns from past experience, and more notably, how the system separates that knowledge from the context in which it was originally observed, so that it can be re-deployed in new situations.

Some evolutionary mechanisms and conditions have been proposed as important factors for improved evolvability. Some concern the modification of genetic variability (e.g., [36, 44, 45] and [46]), while others concern the nature of selective environments and the organisation of development including multiple selective environments [36], sparsity [47], the direct selective pressure on the cost of connections (which can induce modularity [27, 44] and hierarchy [48]), low developmental biases and constraints [49] and stochasticity in GRNs [50]. In this paper, we focus on mechanisms and conditions that can be unified and better understood in machine learning terms, and more notably, how we can utilise well-established theory in learning to characterise general conditions under which evolvability is enhanced. We thus provide the first theory to characterise the general conditions that enhance the evolution of developmental organisations that generalise information gained from past selection, as required to enhance evolvability in novel environments.

	Learning Theory	**Evolutionary Theory**
(a)	Generalisation; ability to produce an appropriate response to novel situations by exploiting regularities observed in past experience (i.e., not rote learning).	Facilitated variation; predisposition to produce fit phenotypes in novel environments (i.e., not just canalisation of past selected targets). Confirmed by experiment Conditions that Facilitate Generalised Phenotypic Distributions.
(b)	The performance of online learning algorithms (i.e., processing one training example at a time) are learning-rate dependent. Both high and low learning rates can lead to situations of under-fitting; failure of the learning system to capture the regularities of the training data [51].	The evolution of generalised phenotypic distributions is dependent on the time-scale of environmental switching. Both high and low time-scales can lead to inflexible developmental structures that fail to capture the functional dependencies of the past phenotypic targets. Confirmed by experiment Rate of Environmental Switching (Learning Rates).
(c)	The problem of over-fitting: improved performance on the training set comes at the expense of generalisation performance on the test set. Over-fitting occurs when the model learns to focus on idiosyncrasies or noise in the training set [52]. Accordingly, the model starts learning the particular irrelevant relationships existing in the training examples rather than the 'true' underlying relationships that are relevant to the general class. This leads to memorisation of specific training examples, which decreases the ability to generalize, and thus perform well, on new data.	Failure of natural selection to evolve generalised developmental organisations: improved average fitness gained by decreasing the phenotypic variation of descendants comes at the expense of potentially useful variability for future selective environments. Favouring immediate fitness benefits would lead to robust developmental structures that canalise the production of the selected phenotypes in the current selective environment. Yet, this sets up a trade-off between robustness and evolvability, since natural selection would always favour inflexible developmental organisations that reduce phenotypic variability and thus hinder the discovery of useful phenotypes that can have fitness benefits in the future. Confirmed by experiment How Generalisation Changes over Evolutionary Time.
(d)	Conditions that alleviate the problem of over-fitting: (1) training with noisy data, i.e., adding noise during the learning phase (jittering), (2) regularisation (parsimony pressure), i.e., introducing a connection cost term into the objective function that favours connections of small values (L_2-regularisation) or fewer connections (L_1-regularisation).	Evolutionary conditions that facilitate the evolution of generalised phenotypic distributions, and thus evolvability: (1) extrinsic noise in selective environments, (2) direct selection pressure on the cost of ontogenetic interactions, which favour simpler developmental processes and sparse network structures. Confirmed by experiments Conditions that Facilitate Generalised Phenotypic Distributions and How Generalisation Changes over Evolutionary Time.
(e)	L_2-regularisation results in similar behaviour as early stopping; an ad-hoc technique that prevents over-fitting by stopping learning when over-fitting begins [51].	Favouring weak connectivity via connection costs results in similar behaviour as stopping adaptation at an early stage. Confirmed by experiments Conditions that Facilitate Generalised Phenotypic Distributions and How Generalisation Changes over Evolutionary Time.
(f)	Training with noise results in similar behaviour to L_2-regularisation [51].	Noisy environments can enhance the evolution of generalised developmental organisation in a similar manner as favouring weak connectivity. Confirmed by experiments Conditions that Facilitate Generalised Phenotypic Distributions and How Generalisation Changes over Evolutionary Time.
(g)	Generalisation performance is dependent on the appropriate level of regularisation and the level of noise, i.e., it depends on the inductive biases, or prior assumptions about which models are more likely to be correct, such as a priori perference for simple models via parsimony pressures.	The evolution of generalised phenotypic distributions is dependent on the strength of selection pressure on the cost of connections and the level of environmental noise. Confirmed by experiment Sensitivity Analysis to Parameters Affecting Phenotypic Generalisation.

Table 1. Predictions Made By Porting Key Lessons of Learning Theory to Evolutionary Theory; Each is Confirmed by Our Experiments.

Experimental Setup

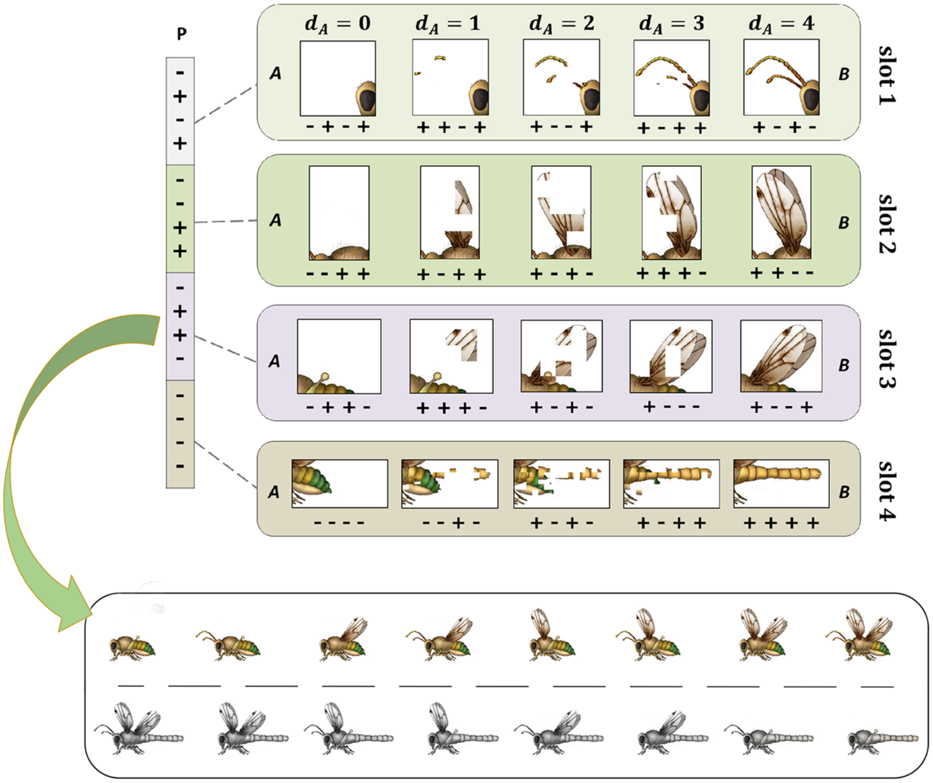


Fig 1. Pictorial representation of phenotypes. (Top) Schematic representation of mapping from phenotypic pattern sequences onto pictorial features. Each phenotypic 'slot' represents a set of features (here 4) controlling a certain aspect of the phenotype (e.g., front wings, halteres and antennae). Within the possible configurations in each slot (here 16), there are two particular configurations (state A and B) that are fit in some environment or another (see S2 Appendix). For example, '$+ + --$' in the second slot (from the top, green) of the phenotypic pattern encodes for a pair of front wings (state B), while '$- - ++$' encodes for their absence (state A). States A and B are the complement of one another, i.e., not neighbours in phenotype space. All of the other intermediate states (here 14) are represented by a random mosaic image of state A and B, based on their respective distance. d_A indicates the Hamming distance between a given state and state A. Accordingly, there exist $\binom{4}{d_A}$ potential intermediate states (i.e., 4 for $d_A = 1$, 6 for $d_A = 2$ and 4 for $d_A = 3$). (Bottom) Pictorial representation of all phenotypes that are perfectly adapted to each of eight different environments. Each target phenotype is analogous to an insect-like organism comprised of 4 functional features. The grey phenotypic targets correspond to bit-wise complementary patterns of the phenotypes on the top half of the space.For example, in the rightmost, top insect, the antennae, forewings, and hindwings are present, and the tail is not. In the rightmost, bottom insect (the bitwise complement of the insect above it), the antennae, forewings, and hindwings are absent, but the tail is present. We define the top row as 'the class' and we disregard the bottom complements as degenerate forms of generalisation.

The main experimental setup involves a non-linear recurrent GRN which develops an embryonic phenotypic pattern, G, into an adult phenotype, P_a, upon which selection can act [25]. An adult phenotype represents the gene expression profile that results from the dynamics of the GRN. Those dynamics are determined by the gene regulatory interactions of the network, B [38, 39, 47, 53, 54] (see SI: Developmental Model). We evaluate the fitness of a given genetic structure based on how close the developed phenotype is to the target phenotypic pattern, S. S characterises the direction of selection for each phenotypic trait, i.e., element of gene expression profile, in the current environment. The dynamics of selective environments are modelled by switching from one target phenotype to another every K generations. K is chosen to be considerably smaller than the overall number of generations simulated. Below, we measure evolutionary time in *epochs*, where each epoch denotes $N_T \times K$ generations and N_T corresponds to the number of target phenotypes. (Note that *epoch* here is a term we are borrowing from machine learning and does not represent geological timescale.)

In the following experiments all phenotypic targets are chosen from the same class (as in [25, 34]). This class consists of 8 different modular patterns that correspond to different combinations of sub-patterns. Each sub-pattern serves as a different function as pictorialised in Fig 1. This modular structure ensures that the environments (and thus the phenotypes that are fittest in those environments) share common regularities, i.e., they are all built from different combinations from the same set of modules. We can then examine whether the system can actually 'learn' these systematicities from a limited set of examples and thereby generalise from these to produce novel phenotypes within the same class. Our experiments are carried out as follows. The population is evolved by exposure to a limited number of selective environments (training). We then analyse conditions under which new phenotypes from the same family are produced (test). As an exemplary problem we choose a training set comprised of three phenotypic patterns from the class (see Fig 2 a).

One way to evaluate the generalisation ability of developmental organisations is to evolve a population to new selective environments and evaluate the evolved predisposition of the development system to produce suitable phenotypes for those environments (as per [34]). We do this at the end of experimental section. We also use a more stringent test and examine the spontaneous production of such phenotypes induced by development from random genetic variation. Specifically, we examine what phenotypes the evolved developmental constraints and biases B are predisposed to create starting from random initial gene expression levels, G. For this purpose, we perform a post-hoc analysis. First, we estimate the phenotypic distributions induced by the evolved developmental architecture under drift. Since mutation on the direct effects on the embryonic phenotypes (G) in this model is much greater than mutation on regulatory interactions (B) (see Methods), we estimate drift with a uniformly random distribution over G (keeping B constant). Then we assess how successful the evolved system is at producing high-fitness phenotypes, by seeing if the phenotypes produced by the evolved correlations, B, tend to be members of the general class (see Methods).

Results and Discussion

Conditions that Facilitate Generalised Phenotypic Distributions

In this section, we focus on the conditions that promote the evolution of adaptive developmental biases that facilitate generalised variational structures. To address this, we examine the distributions of potential phenotypic variants induced by the evolved developmental structure in a series of different evolutionary scenarios: 1) different time-scales of environmental switching, 2) environmental noise and 3) direct selection pressure for simple developmental processes applied via a the cost of ontogenetic interactions favouring i) weak and ii) sparse connectivity.

Rate of Environmental Switching (Learning Rates)

In this scenario, we assess the impact of the rate at which selective environments switch on the evolution of generalised developmental organisations. This demonstrates prediction (b) from Table 1. The total number of generations was kept fixed at 24×10^6, while the switching intervals, K, varied. In all reproductive events, G is mutated by adding a uniformly distributed random value drawn in $[-0.1, 0.1]$. Additionally, in half the reproduction events, all interaction coefficients are mutated slightly by adding a uniformly distributed value drawn from $[-0.1/(15N^2), 0.1/(15N^2)]$, where N corresponds to the number of phenotypic traits.

Prior work on facilitated variation has shown that the evolution of evolvability in varying selective environments is dependent on the time-scale of environmental change [34–36]. This is analogous to the sensitivity of generalisation to learning rate in learning systems. The longer a population is exposed to a selective environment, the higher the expected adaptation accumulated to that environment would be. Accordingly, the rate of change in a given environment (learning rate) can be controlled by the rate of environmental change (sample rate). Slow and fast environmental changes thus correspond to fast and

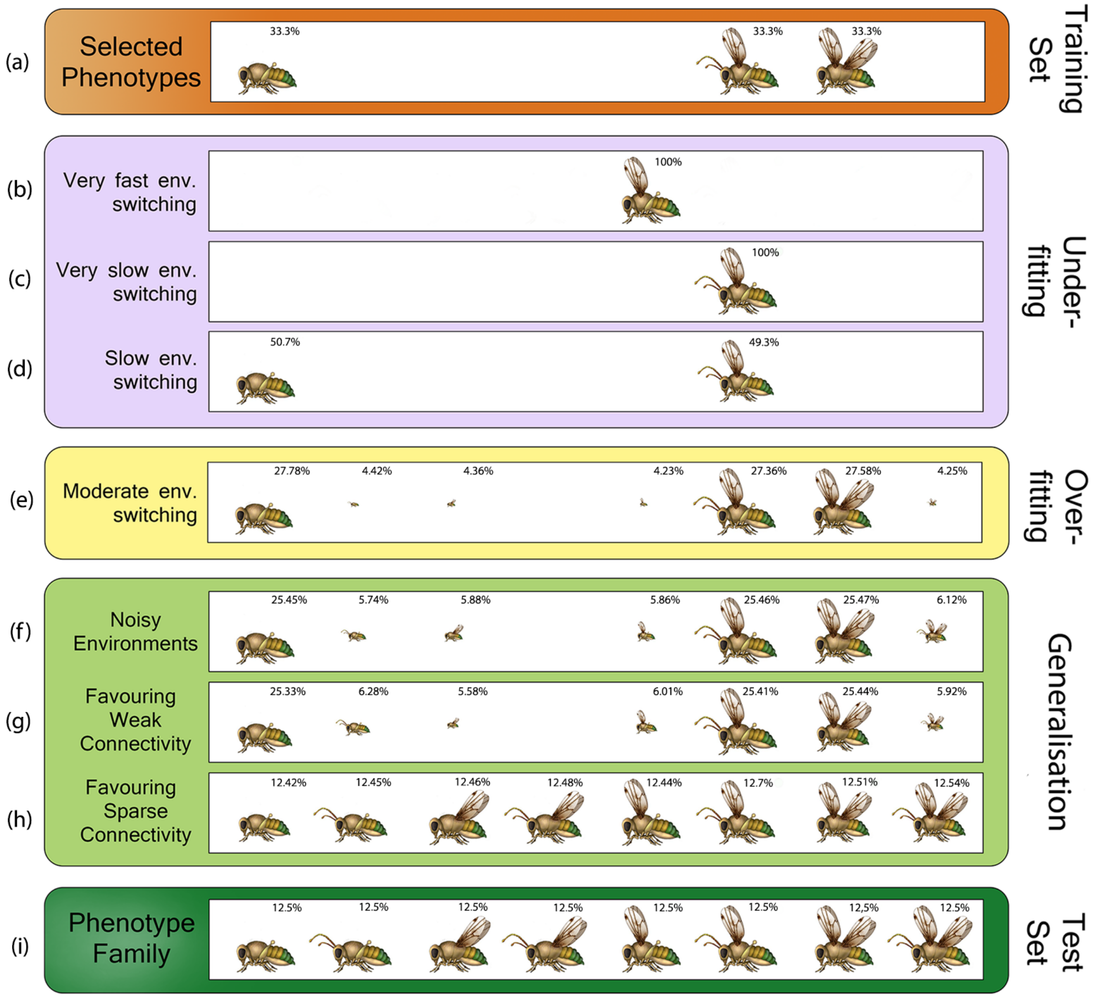


Fig 2. Conditions that facilitate generalised phenotypic distributions. Potential phenotypic distributions induced by the evolved developmental process under 1) different time-scales of environmental switching, 2) environmental noise ($\kappa = 35 \times 10^{-4}$) and 3) direct selection pressure for weak ($\lambda = 38$) and sparse connectivity ($\lambda = 0.22$). The organisms were exposed to three selective environments (a) from the general class (i). Developmental memorisation of past phenotypic targets clearly depends on the time-scale of environmental change. Noisy environments and parsimony pressures enhance the generalisation ability of development predisposing the production of previously unseen targets from the class. The size of the insect-like creatures describes relative frequencies and indicates the propensity of development to express the respective phenotype (phenotypes with frequency less than 0.01 were ignored). Note that the initial developmental structure represented all possible phenotypic patterns equally (here 2^{12}).

slow sample rates respectively.

We find that when the environments rapidly alternated from one to another (e.g., $K \sim 2$), natural selection canalised a single phenotypic pattern (Fig 2 b). This phenotype however did not correspond to any of the previously selected ones (Fig 2 a). Rather, this corresponds to the combination of phenotypic characters that occurs most in each of the seen target phenotypes. Hence, it does best on average over the past selective environments. For example, over the three patterns selected in the past it is more common that halteres are selected than a pair of back wings, or a pair of front wings is present more often than not and so on.

When environments changed very slowly (e.g., $K \sim 4 \times 10^6$), development canalised the first selective environment experienced, prohibiting the acquisition of any useful information regarding other selective environments (Fig 2 c). The situation was improved for a range of slightly faster environmental switching times (e.g., $K \sim 2 \times 10^6$), where natural selection also canalised the second target phenotype experienced, but not all three (Fig 2 d). Canalisation can therefore be opposed to evolvability, resulting in very inflexible models that failed to capture any or some of the relevant regularities in the past or current environments, i.e., *under-fitting*. Such developmental organisations could provide some limited immediate fitness benefits in the short-term, but are not good representatives of either the past, or the general class.

When the rate of environmental switching was intermediate (e.g., $K \sim 4 \times 10^4$), the organisms exhibited developmental memory [25] . Although initially all possible phenotypic patterns (here 2^{12}) were equally represented by development, the variational structure of development was adapted over evolutionary time to fit the problem structure of that past, by canalising the production of previously seen targets (Figure 2 e, see also S2 Fig). This holds for a wide range of intermediate switching intervals (see S3 Fig). This observations illustrates the ability of evolution to genetically acquire and utilise information regarding the statistical structure of previously experienced environments.

The evolved developmental constraints also exhibited generalised behaviour by allowing the production of three additional phenotypes that were not directly selected in the past, but share the same structural regularities with the target phenotypes. These new phenotypic patterns correspond to novel combinations of previously-seen phenotypic features. Yet, the propensity to express these extra phenotypes was still limited. The evolved variational mechanism over-represented past targets, failing to properly generalise to all potential, but yet-unseen selective environments from the same class as the past ones, i.e., over-fitted (see below). We find no rate of environmental variation capable of causing evolution by natural selection to evolve a developmental organisation that produces the entire class. Consequently, the rate of environmental change can facilitate the evolution of developmental memory, but does not always produce good developmental generalisation.

Here we argue that the problem of natural selection failing to evolve generalised phenotypic distributions in certain cases is formally analogous to the problem of learning systems failing to generalise due to either under- or over-fitting. In learning, under-fitting is observed when a learning system is incapable of capturing a set of exemplary observations. On the other hand, over-fitting is observed when a model is over-trained to memorise a particular set of exemplary observations, at the expense of predictive performance on previously unseen data from the class [51]. Over-fitting occurs when the model learns to focus on idiosyncrasies or noise in the training set [52]. Similarly, canalisation to past selective environments can be opposed to evolvability if canalised phenotypes from past environments are not fit in future environments. Specifically, canalisation can be opposed to evolvability by either 1) (first type of underfitting, from high learning rates) reducing the production of all phenotypic characters except those that are fit in the selective environments that happen to come early (Fig 2 c), 2) (second type of under-fitting, from low learning rates) reducing the production of all characters except those that are fit on average over the past selective environments (Fig 2 b), or 3) (over-fitting) successfully producing a sub-set of or all phenotypes that were fit in the past selective environments, but inhibiting the production of new and potentially useful phenotypic variants for future selective environments (Fig 2 d, e).

Below, we investigate the conditions under which an evolutionary process can avoid canalising the past and remain appropriately flexible to respond to novel selective environments in the future. To do so, we test whether techniques used to avoid under-fitting and over-fitting that improve generalisation to unseen test sets in learning models will likewise alleviate canalisation to past phenotypic targets and improve fit to novel selective environments in evolutionary systems. For this purpose, we choose the time scale of environmental change to be moderate ($K = 20000$). This constitutes our control experiment in the absence of environmental noise and/or any selective pressure on the cost of connections. In the following evolutionary scenarios, simulations were run for 150 epochs. This demonstrates prediction d,e, and f from Table 1.

Noisy Environments (Training with Noisy Data)

In this scenario, we investigate the evolution of generalised developmental organisations in noisy environments by adding Gaussian noise, $n_\mu \sim N(0,1)$ to the respective target phenotype, S, at each generation. The level of noise was scaled by parameter κ. In order to assess the potential of noisy selection to facilitate phenotypic generalisation, we show results for the optimal amount of noise (here $\kappa = 35 \times 10^{-4}$). Later, we will show how performance varies with the amount of noise.

We find that the distribution of potential phenotypic variants induced by the evolved development in noisy environments was still biased in generating past phenotypic patterns (Fig 2 f). However, it improved fit to other selective environments in the class slightly compared with Fig 2 e. The evolved developmental structure was characterised by more suitable variability, displaying higher propensity, compared to the control, in producing those variants from the class that were not directly selected in the past.

Masking spurious details in the training set by adding noise to the training samples during the training phase is a general method to combat the problem of over-fitting in learning systems. This technique is known as ‘training with noise’ or ‘jittering’ [51] and is closely related to the use of intrinsic noise in deep neural networks; a technique known as ‘dropout’ [55]. The intuition is that when noise is applied during the training phase, it makes it difficult for the optimisation process to fit the data precisely, and thus it inhibits capturing the idiosyncrasies of the training set. Training with noise is mathematically equivalent to a particular way of controlling model complexity known as Tikhonov regularisation [51].

Favouring Weak Connectivity (L_2-regularisation)

In this scenario, the developmental structure was evolved under the direct selective pressure for weak connectivity — favouring regulatory interactions of small magnitude, i.e., L_2-regularisation (see Methods). Weak connectivity is achieved by applying a direct pressure on the cost of connections that is proportion to their magnitude. This imposes constraints on the evolution of the model parameters by penalising extreme values.

Under these conditions natural selection discovered more general developmental structures. Specifically, developmental generalisation was enhanced in a similar manner as in the presence of environmental noise, favouring similar weakly generalised phenotypic distributions. The distribution of potential phenotypic variants induced by development displayed higher propensity in producing useful phenotypic variants for potential future selective environments (Fig 2 g).

Favouring Sparse Connectivity (L_1-regularisation)

In this scenario, the developmental structure was evolved under the direct selective pressure for sparse connectivity — favouring fewer regulatory interactions, i.e., L_1-regularisation. Sparse connectivity is achieved by applying an equal direct pressure on the cost of connections. This imposes constraints on the evolution of the parameters by decreasing all non-zero values equally, and thus favouring models using fewer connections.

We find that under these conditions the evolution of generalised developmental organisations was dramatically enhanced. The evolved phenotypic distribution (Fig 2 h) was a perfect representation of the class (Fig 2 i). We see that the evolved developmental process under the pressure for sparsity favoured the production of novel phenotypes that were not directly selected in the past. Those novel phenotypes were not arbitrary, but characterised by the time-invariant intra-modular regularities common to past selective environments. Although the developmental system was only exposed to three selective environments, it was able to generalise and produce all of the phenotypes from the class by creating novel combinations of previously-seen modules. More notably, we see that the evolved developmental process also pre-disposed the production of that phenotypic pattern missing under the conditions for weak connectivity and environmental noise due to strong developmental constraints.

Moreover, the parsimonious network topologies we find here arise as a consequence of a direct pressure on the cost of connections. The hypothesis that sparse network can arise through a cost minimisation process is also supported by previous theoretical findings advocating the advantages of sparse gene regulation networks [56]. Accordingly, natural selection favours the emergence of gene-regulatory networks of minimal complexity. In [56], Leclerc argues that sparser GRNs exhibit higher dynamical robustness. Thus, when the cost of complexity is considered, robustness also implies sparsity. In this study, however, we demonstrated that sparsity gives rise to enhanced evolvability. This indicates that parsimony on the connectivity of the GRNs is a desired property that may facilitate both robustness and evolvability.

Favouring weak and sparse connectivity belong in a general category of *regularisation* methods that alleviate over-fitting by penalising unnecessary model complexity via the application of a parsimony pressure that favours simple models with fewer assumptions on the data, i.e., imposing a form of Occam's razor on solutions (e.g., the Akaike [57] and [58] Bayesian information criteria, limiting the number of features in decision trees [59], or limiting the tree depth in genetic programming [60]). The key observation is that networks with too few connections will tend to under-fit the data (because they are unable to represent the relevant interactions or correlations in the data); whereas networks with more connections than necessary will tend to over-fit the idiosyncrasies of the training data, because they can memorize those idiosyncrasies instead of being forced to learn the underlying general pattern.

How Generalisation Changes over Evolutionary Time

We next asked why costly interactions and noisy environments facilitate generalised developmental organisations. To understand this, we monitor the match between the phenotypic distribution induced by the evolved developmental process and the ones that describe the past selective environments (training set) and all potential selective environments (test set) respectively over evolutionary time in each evolutionary setting (see Methods). Following conventions in learning theory, we term the first measure 'training error' and the second 'test error'. This demonstrates predictions c, e and f from Table 1.

The dependence of the respective errors on evolutionary time are shown in Fig 3. For the control scenario (panel A) we observe the following trend. Natural selection initially improved the fit of the phenotypic distributions to both distributions of past and future selective environments. Then, while the fit to past selective environments continued improving over evolutionary time, the fit to potential, but yet-unseen, environments started to deteriorate (see also S2 Fig). The evolving organisms tended to accurately *memorise* the idiosyncrasies of their past environments, at the cost of losing their ability to retain appropriate flexibility for the future, i.e., over-fitting. The dashed-line in Fig 3 A indicates when the problem of over-fitting begins,i.e., when the test error first increases. We see that canalisation can be opposed to the evolution of generalised phenotypic distributions in the same way over-fitting is opposed to generalisation. Then, we expect that preventing the canalisation of past targets can enhance the generalisation performance of the evolved developmental structure. Indeed, Fig 3 B,C,D confirm this hypothesis (predictions a-c from Table 1).

In the presence of environmental noise, the generalisation performance of the developmental structure was improved by discovering a set of regulatory interactions that corresponds to the minimum of the generalisation error curve of 0.34 (Fig 3 B). However, natural selection in noisy environments was only able to postpone canalisation of past targets and was unable to avoid it in the long term (see SI). Consequently, stochasticity improved evolvability by decreasing the speed at which over-fitting occurs, allowing for the developmental system to spend more time at a state which was characterised by high generalisation ability (see also S6 Fig). On the other hand, under the parsimony pressure for weak connectivity, the

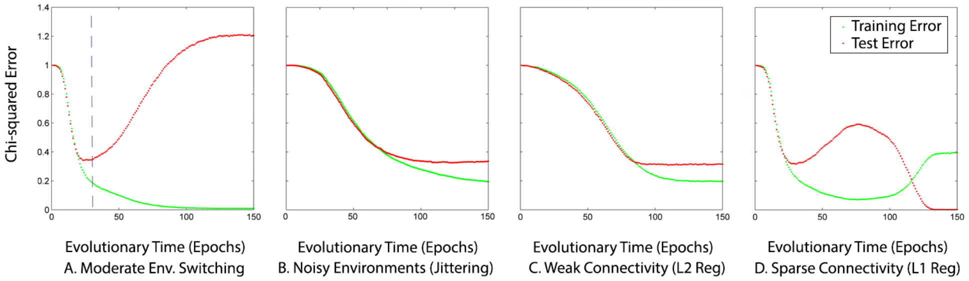
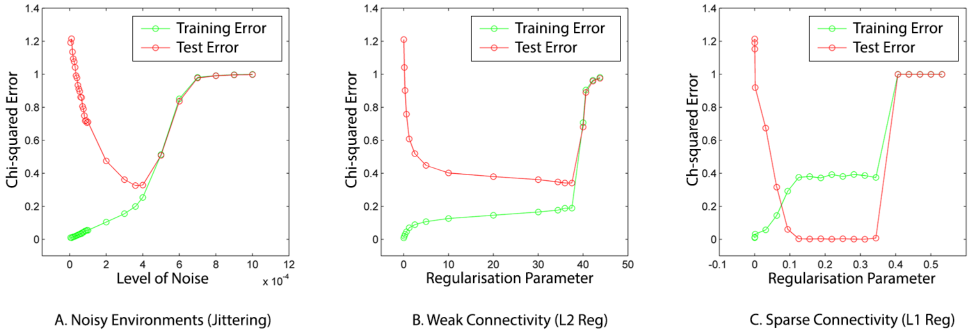


Fig 3. How generalisation changes over evolutionary time. The match between phenotypic distributions generated by evolved GRN and the target phenotypes of selective environments the developmental system has been exposed to (training error) and all selective environments (test error) against evolutionary time for (A) moderate environmental switching, (B) noisy environments, (C) favouring weak connectivity and (D) favouring sparse connectivity. The vertical dashed line denotes when the ad-hoc technique of early stopping would be ideal, i.e. at the moment the problem of over-fitting begins. Favouring weak connectivity and jittering exhibits similar effects as applying early stopping.

evolving developmental system maintained the same generalisation performance over evolutionary time. The canalisation of the selected phenotypes was thus prevented by preventing further limitation of the system's phenotypic variability. Note that the outcome of these two methods (Fig 3 B and C) resembles in many ways the outcome as if we stopped at the moment when the generalisation error was minimum, i.e., early stopping; an ad-hoc solution to preventing over-fitting [51]. Accordingly, learning is stopped before the problem of over-fitting begins (see also S6 Fig). Subject to parsimony pressure for sparse connectivity, we observe that the generalisation error of the evolving developmental system reached zero (Fig 3 D). Accordingly, natural selection successfully exploited the time-invariant regularities of the environment properly representing the entire class (Fig 2 h). Additionally, S4 Fig shows that the entropy of the phenotypic distribution reduces as expected over evolutionary time as the developmental process increasingly canalises the training set phenotypes. In the case of perfect generalisation to the class (sparse connectivity), this convergence reduces from 16 bits (the original phenotype space) to four bits, corresponding to four degrees of freedom where each of the four modules vary independently. In the other cases, overfitting is indicated by reducing to less than four bits.

Sensitivity Analysis to Parameters Affecting Phenotypic Generalisation

As seen so far, the generalisation ability of development can be enhanced under the direct selective pressure for both sparse and weak connectivity and the presence of noise in the selective environment, when the



A. Noisy Environments (Jittering)

B. Weak Connectivity (L2 Reg)

C. Sparse Connectivity (L1 Reg)

Fig 4. Role of the strength of parsimony pressure and the level of environmental noise. The match between phenotypic distributions and the selective environments the network has been exposed to (training error) and all possible selective environments of the same class (generalisation error) for (A) noisy environments against parameter κ and under the parsimony pressure weak (B) and sparse (C) connectivity against parameter λ.

strength of parsimony pressure and the level of noise were properly tuned. Different values of λ and κ denote different evolutionary contexts, where λ determines the relative burden placed on the fitness of the developmental system due to reproduction and maintenance of its elements, or other physical constraints and limitations, and κ determines the amount of extrinsic noise found in the selective environments (see Evaluation of Fitness).

In the following, we analyse the impact of the strength of parsimony pressure and the level of environmental noise on the evolution of generalised developmental organisations. Simulations were run for various values of parameters λ and κ. Then, the training and generalisation error were evaluated and recorded (Fig 4). This demonstrates prediction (g) from Table 1.

We find that in the extremes, low and high levels of parsimony pressures, or noise, gave rise to situations of over-fitting and under-fitting respectively (Fig 4). Very small values of λ, or κ, were insufficient at finding good regulatory interactions to facilitate high evolvability to yet-unseen environments, resulting in the canalisation of past targets, i.e., over-fitting. On the other hand, very large values of λ over-constrained the search process hindering the acquisition of any useful information regarding environment's causal structure, i.e., under-fitting. Specifically, with a small amount of L_1-regularisation, the generalisation error is dropped to zero. This outcome holds for a wide spectrum of the regularisation parameter $ln(\lambda) \in [0.15, 0.35]$. However, when λ is very high (here $\lambda = 0.4$), the selective pressure on the cost of connection was too large; this resulted in the training and the generalisation errors corresponds to the original 'no model' situation (Fig 4 C). Similarly, with a small amount of L_2-regularisation, the generalisation error quickly drops. In the range $[10, 38]$ the process became less sensitive to changes in λ, resulting in one optimum at $\lambda = 38$ (Fig 4 B). Similar results were also obtained for jittering (Fig 4 A). But the generalisation performance of the developmental process changes 'smoothly' with κ, resulting in one optimum at $\kappa = 35 \times 10^{-4}$ (Fig 4 A). Inductive biases need to be appropriate for a given problem, but in many cases a moderate bias favouring simple models is sufficient for non-trivial generalisation.

Generalised Developmental Biases Improve the Rate of Adaptation

Lastly we examine whether generalised phenotypic distributions can actually facilitate evolvability. For this purpose, we consider the rate of adaptation to each of all potential selective environments as the number of generations needed for the evolving entities to reach the respective target phenotype.

To evaluate the propensity of the organisms to reach a target phenotype as a systemic property of its developmental architecture, the regulatory interactions were kept fixed, while the direct effects on the embryonic phenotype were free to evolve for 2500 generations, which was empirically found to be sufficient for the organisms to find a phenotypic target in each selective environment (when that was allowed by the developmental structure). In each run, the initial gene expression levels were uniformly chosen at random. The results here were averaged over 1000 independent runs, for each selective environment and for each of the four different evolutionary scenarios (as described in the previous sections). Then, counts of the average number of generations to reach the target phenotype of the corresponding selective environment were taken. This was evaluated by measuring the first time the developmental system achieved maximum fitness possible. If the target was not reached, the maximum number of generations 2500 was assigned.

We find that organisms with developmental organisations evolved in noisy environments or the parsimony pressure on the cost of connections adapted faster than the ones in the control scenario (Fig 5). The outliers in the evolutionary settings of moderate environmental switching, noisy environments and favouring weak connectivity, indicate the inability of the developmental system to express the target phenotypic pattern for that selective environment due to the strong developmental constraints that evolved in those conditions. This corresponds to the missing phenotype from the class we saw above in the evolved phenotypic distributions induced by development (Fig 2 e, f, g). In all these three cases development allowed for the production of the same set of phenotypic patterns. Yet, developmental structures evolved in the presence of environmental noise or under the pressure for weak connectivity exhibited higher adaptability due to their higher propensity to produce other phenotypes of the structural family. In

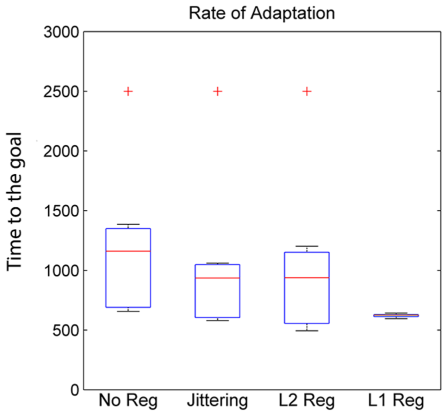


Fig 5. Generalised developmental organisations improve the rate of adaptation to novel selective environments. Boxplot of the generations taken for the evolved developmental systems to reach the target phenotype for all potential selective environments under different evolutionary conditions. The developmental architecture is kept fixed and only the direct effects on the embryonic phenotype are free to evolve. Organisms that facilitate generalised phenotypic distributions, such as the ones evolved in noisy environments or under the direct pressure on the cost connections, adapt faster to novel selective environments exhibiting enhanced evolvability. The outliers indicate the inability of the corresponding evolved developmental structures to reach that selective target due to strong developmental constraints.

particular, we see that for the developmental process evolved under the pressure for sparsity, the rate of adaptation of the organisms was significantly improved. The variability structure evolved under sparsity to perfectly represent the functional dependencies between phenotypic traits. Thus, it provided a selective advantage guiding phenotypic variation in more promising directions.

Conclusions

The above experiments demonstrated the transfer of predictions from learning models into evolution, by specifically showing that: a) the evolution of generalised phenotypic distributions is dependent on the time-scale of environmental switching, in the same way that generalisation in online learning algorithms is learning-rate dependent, b) the presence of environmental noise can be beneficial for the evolution of generalised phenotypic distributions in the same way training with corrupted data can improve the generalisation performance of learning systems with the same limitations, c) direct selection pressure for weak connectivity can enhance the evolution of generalised phenotypic distributions in the same way L_2-regularisation can improve the generalisation performance in learning systems, d) noisy environments result in similar behaviour as favouring weak connectivity, as Jittering can have similar effects to L_2-regularisation in learning systems, e) direct selection pressure for sparse connectivity can enhance the evolution of generalised phenotypic distributions in the same way that L_1-regularisation can improve the generalisation performance in learning systems, f) favouring weak connectivity (i.e., L_2-regularisation) results in similar behaviour to as early stopping and g) the evolution of generalised phenotypic distributions is dependent on the strength of selection pressure on the cost of connections and the level of environmental noise, in the same way generalisation is dependent on the level of inductive biases.

Learning is generally *contextual*; it gradually builds upon what *concepts* are already known. Here these concepts correspond to the repeated modular sub-patterns persisting over all observations in the training set which become encoded in the modular components of the evolved network. The inter-module connections determine which combinations of (sub-)attractors in each module are compatible and which are not. Therefore, the evolved network representation can be seen as dictating a higher-order conceptual (combinatorial) space based on previous experience. This enables the evolved developmental system to explore permitted combinations of features constrained by past selection. Novel phenotypes can thus arise through new combinations of previously selected phenotypic features explicitly embedded in the developmental architecture of the system [25]. Indeed, under the selective pressure for sparse connectivity, we observe that the phenotypic patterns generated by the evolved developmental process consisted of combinations of features from past selected phenotypic patterns. Thus, we see that the 'developmental memories' are stored and recalled in combinatorial fashion allowing generalisation.

We see that noisy environments and the parsimony pressure on the cost of connections led to more evolvable genotypes by internalising more general models of the environment into their developmental organisation. The evolved developmental systems did not solely capture and represent the specific idiosyncrasies of past selective environments, but internalised the regularities that remained time-invariant in all environments of the given class. This enabled natural selection to 'anticipate' novel situations by accumulating information about and exploiting the tendencies in that class of environments defined by the regularities. Peculiarities of past targets were generally represented by weak correlations between phenotypic characters as these structural regularities were not typically present in all of the previously-seen selective environments. Parsimony pressures and noise then provided the necessary selective pressure to neglect or de-emphasise such spurious correlations and maintain only the strong ones which tended to correspond to the underlying problem structure (in this case, the intra-module correlations only, allowing all combinations of fit modules). Enhancing evolvability by means of inductive biases is not for granted in evolutionary systems any more than such methods have guarantees in learning systems. The quality of the method depends on information about past targets and the strength of the parsimony pressure. Inductive biases can however constrain phenotypic evolution into more promising directions and exploit systematicities in the environment when opportunities arise.

In this study we demonstrated that canalisation can be opposed to evolvability in biological systems the same way under- or over-fitting can be opposed to generalisation in learning systems. We showed that conditions that are known to alleviate over-fitting in learning are directly analogous to the conditions that enhance the evolution of evolvability under natural selection. Specifically, we described how well-known techniques, such as learning with noise and penalising model complexity, that improve the generalisation ability of learning models can help us understand how noisy selective environments and the direct selection pressure on the reproduction cost of the gene regulatory interactions can enhance context-specific evolvability in gene regulation networks. This opens-up a well-established theoretical framework, enabling it to be exploited in evolutionary theory. This equivalence demystifies the basic idea of the evolution of evolvability by equating it with generalisation in learning systems. This framework predicts the conditions that will enhance generalised phenotypic distributions and evolvability in natural systems.

Methods

Evolution of GRNs

We model the evolution of a population of GRNs under strong selection and weak mutation where each new mutation is either fixed or lost before the next arises. This emphasises that the effects we demonstrate do not require lineage-level selection [61–63] — i.e., they do not require multiple genetic lineages to coexist long enough for their mutational distributions to be visible to selection. Accordingly a simple hill-climbing model of evolution is sufficient [25, 36].

The population is represented by a single genotype $[G, B]$ (the direct effects and the regulatory interactions respectively) corresponding to the average genotype of the population. Similarly, mutations in G and B indicate slight variations in population means. Consider that G' and B' denote the respective mutants. Then the adult mutant phenotype, P'_a, is the result of the developmental process, which is characterised by the interaction B', given the direct effects G'. Subsequently, the fitness of P_a and P'_a are calculated for the current selective environment, S. If $f_S(P'_a) > f_S(P_a)$, the mutation is beneficial and therefore adopted, i.e., $G_{t+1} = G'$ and $B_{t+1} = B'$. On the other hand, when a mutation is deleterious, G and B remain unchanged.

The variation on the direct effects, G, occurs by applying a simple point mutation operator. At each evolutionary time step, t, an amount of μ_1 mutation, drawn from $[-0.1, 0.1]$ is added to a single gene i. Note that we enforce all $g_i \in [-1, 1]$ and hence the direct effects are hard bounded, i.e., $g_i = min\{max\{g_i + \mu_1, -1\}, 1\}$. For a developmental architecture to have a meaningful effect on the phenotypic variation, the developmental constraints should evolve considerably slower than the phenotypic variation they control. We model this by setting the rate of change of B to lower values as that for G. More specifically, at each evolutionary time step, t, mutation occurs on the matrix with probability 1/15. The magnitude μ_2 is drawn from $[-0.1/(15N^2), 0.1/(15N^2)]$ for each element b_{ij} independently, where N corresponds to the number of phenotypic traits.

Evaluation of Fitness

Following the framework used in [64], we define the fitness of the developmental system as a benefit minus cost function.

The benefit of a given genetic structure, b, is evaluated based on how close the developed adult phenotype is to the target phenotype of a given selective environment. The target phenotype characterises a favourable direction for each phenotypic trait and is described by a binary vector, $S = \langle s_1, \ldots, s_N \rangle$, where $s_i \in \{-1, 1\}, \forall i$. For a certain selective environment, S, the selective benefit of an adult phenotype, P_a, is given by (modified from [25]):

$$b = w(P_a, S) = \frac{1}{2}\left(1 + \frac{P_a \cdot S}{N}\right), \tag{1}$$

where the term $P_a \cdot S$ indicates the inner product between the two respective vectors. The adult phenotype is normalised in $[-1, 1]$ by $P_a \leftarrow P_a/(\tau_1/\tau_2)$, i.e., $b \in [0, 1]$.

The cost term, c, is related to the values of the regulatory coefficients, $b_{ij} \in B$ [65]. The cost represents how fitness is reduced as a result of the system's effort to maintain and reproduce its elements, e.g., in *E. coli* it corresponds to the cost of regulatory protein production. The cost of connection has biological significance [27, 64–67], such as being related to the number of different transcription factors or the strength of the regulatory influence. We consider two cost functions proportional to i) the sum of the absolute magnitudes of the interactions, $c = \|B\|_1 = \sum_{i=1}^{N^2} |b_i j|/N^2$, and ii) the sum of the squares of the magnitudes of the interactions, $c = \|B\|_2^2 = \sum_{i=1}^{N^2} b_{ij}^2/N^2$, which put a direct selection pressure on the weights of connections, favouring sparse (L_1-regularisation) and weak connectivity (L_2-regularisation) respectively [68].

Then, the overall fitness of P_a for a certain selective environment S is given by:

$$f_S(P_a) = b - \lambda c, \tag{2}$$

where parameter λ indicates the relative importance between b and c. Note that the selective advantage of structure B is solely determined by its immediate fitness benefits on the current selective environment.

Chi-squared Error

The χ^2 measure is used to quantify the lack of fit of the evolved phenotypic distribution $\hat{P}_t(s_i)$ against the distribution of the previously experienced target phenotypes $P_t(s_i)$ and/or the one of all potential target phenotypes of the same family $P(s_i)$. Consider two discrete distribution profiles, the observed frequencies $O(s_i)$ and the expected frequencies $E(s_i)$, $s_i \in S, \forall i = 1, \ldots, k$. Then, the chi square error between distribution O and E is given by:

$$\chi^2(O, E) = \sum_i \frac{(O(s_i) - E(s_i))^2}{E(s_i)} \tag{3}$$

S corresponds to the training set and the test set when the training and the generalisation error are respectively estimated. Each $s_i \in S$ indicates a phenotypic pattern and $P(s_i)$ denotes the probability of this phenotype pattern to arise.

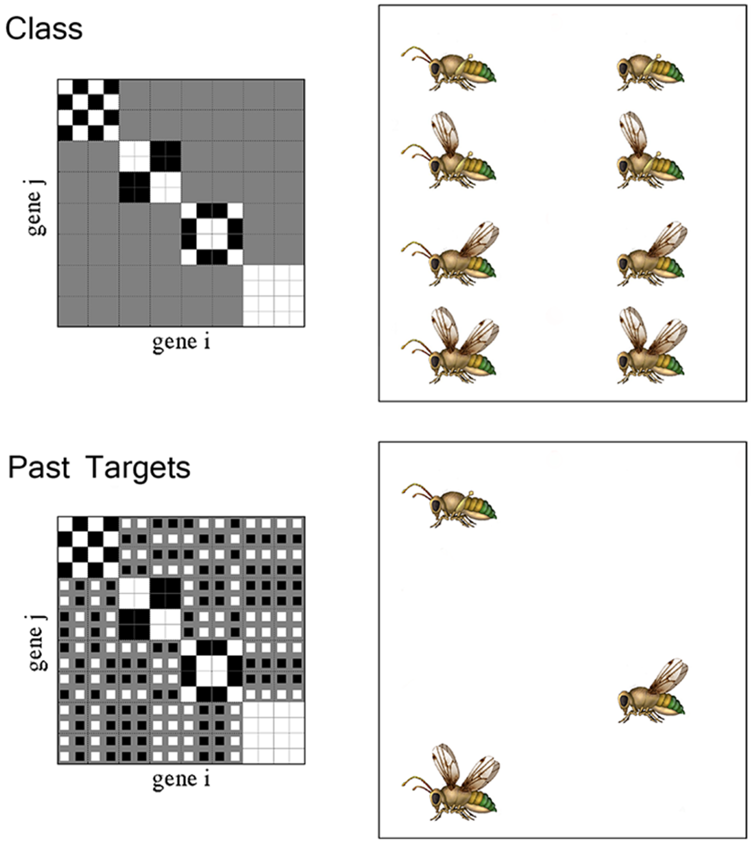
The samples, over which the distribution profiles are estimated, are uniformly drawn at random (see Estimating the Empirical Distributions). This guarantees that the sample is not biased and the observations under consideration are independent. Although the phenotypic profiles here are continuous variables, they are classified into binned categories (discrete phenotypic patterns). These categories are mutually exclusive and the sum of all individual counts in the empirical distribution is equal to the total number of observations. This indicates that no observation is considered twice, and also that the categories include all observations in the sample. Lastly, the sample size is large enough to ensure large expected frequencies, given the small number of expected categories.

Estimating the Empirical Distributions

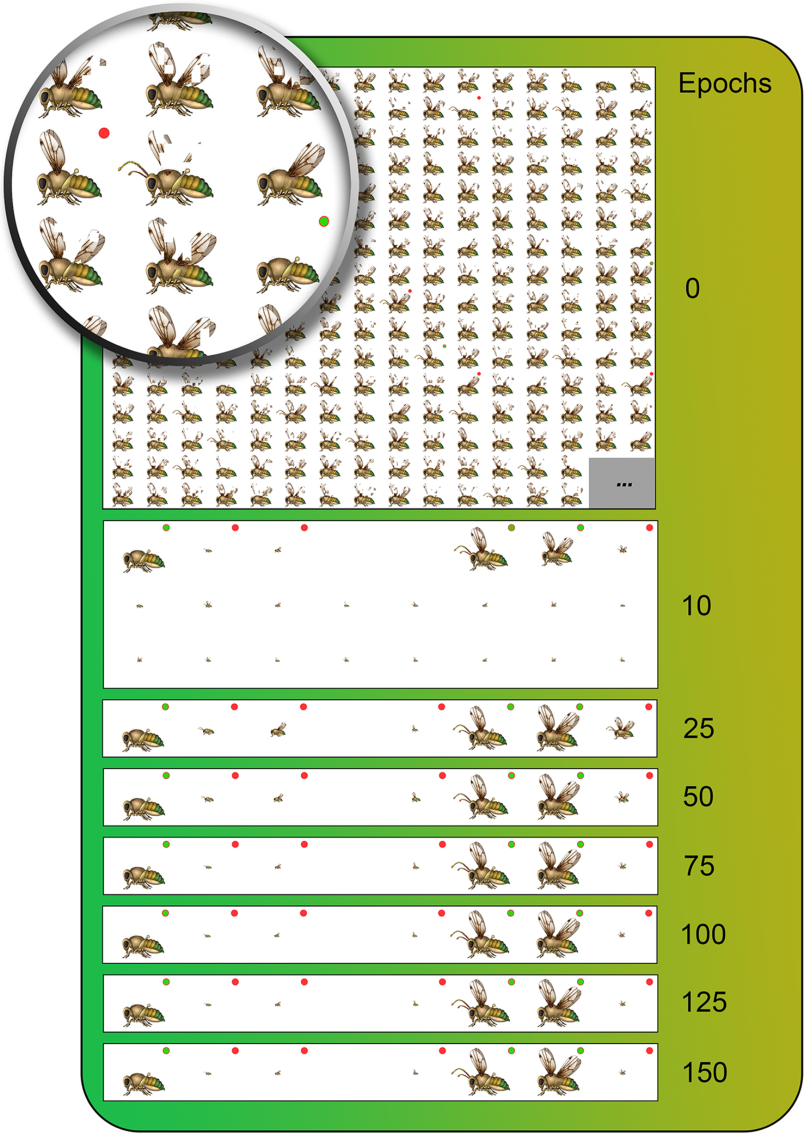
For the estimation of the empirical (sample) probability distribution of the phenotypic variants over the genotypic space, we follow the Classify and Count (CC) approach [69]. Accordingly, 5000 embryonic phenotypes, $P(0) = G$, are uniformly generated at random in the hypercube $[-1, 1]^N$. Next, each of these phenotypes is developed into an adult phenotype and the produced phenotypes are categorised by their closeness to target patterns to take counts. Note that the development of each embryonic pattern in the sample is unaffected by development of other embryonic patterns in the sample. Also, the empirical distributions are estimated over all possible combinations of phenotypic traits, and thus each developed phenotype in the sample falls into exactly one of those categories. Finally, low discrepancy quasi-random sequences (Sobol sequences; [70]) with Matousek's linear random scramble [71] were used to reduce the stochastic effects of the sampling process, by generating more homogeneous fillings over the genotypic space.

Supporting Information Legends

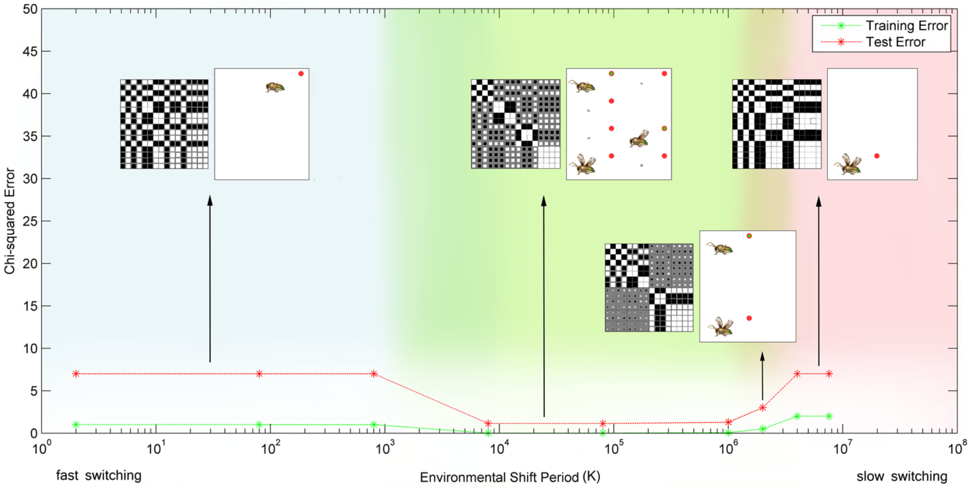
S1 Appendix. Supporting Figures



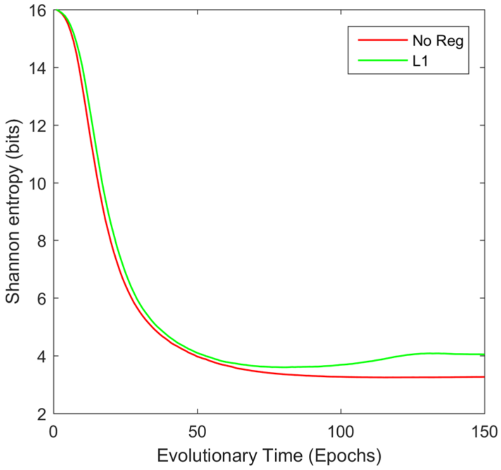
S1 Fig. The underlying correlational structure of the class and the training set. (Top) Hinton diagram of the variance-covariance matrix and phenotypic distribution of all potential future phenotypic targets. The true underlying structure of the given problem set which is comprised of all 8 possible phenotypic targets is described by the block diagonal interaction matrix. Accordingly, the traits within each module that encode for each functional part of the organism (e.g., front wings) are strongly correlated with each other (positively or negatively depending on the combination of signs in the particular phenotypic pattern used), and no correlations between one module and another (e.g., the production of halteres is functionally independent from the production of front wings). (Bottom) Hinton diagram of the variance-covariance matrix and phenotypic distribution of past phenotypic targets. The structure of the training set which is comprised of 3 phenotypic targets is described by an interaction matrix with non-zero off-diagonal elements. Those elements correspond to spurious correlations that describe functional phenotypic dependencies between modules that are present in the past selected phenotypic targets (e.g., the production of front wings is positively correlated with the production of antennas). Such developmental structures will appropriately represent the 3 past selected targets, but fail to generate all 8 phenotypes from the class. The colour and the size of the squares in Hinton's representation indicate the sign and the magnitude of the respective correlations.



S2 Fig. The evolution of phenotypic distribution for moderate environmental switching. Pictorial representation of the phenotypic distributions induced by the evolving developmental process over evolutionary time for moderate environmental switching. Green circles indicate past selected targets, while red circles indicate previously-unseen phenotypes from the same phenotype family as the past ones. Phenotypes outside of the class are represented by distorted mosaic images. The size of the insect-like creatures indicates the propensity of development to express the respective phenotype. At the beginning (epoch 0), development equally predisposes the production of all possible phenotypic patterns (here 2^{12}), i.e., no developmental biases. The evolving developmental structure initially starts canalising only phenotypes from the class. After epoch 25 however it further canalises the production of past selected phenotypes, by reducing the propensity of producing those phenotypes from the class that were not selected in the past, i.e., over-fitting.



S3 Fig. Fast and Slow Environmental Switching Fail to Evolve Developmental Memory. The match between phenotypic distributions and the selective environments the network has been exposed to (training error) and all selective environments (generalisation error) against different environmental switching intervals (K). The insets illustrate the Hinton diagram of the evolved interaction matrix for each regime (indicated by different background colour) and the respective phenotypic distribution induced by the evolved developmental process.



S4 Fig. Entropy of the phenotypic distribution reduces over evolutionary time. Shannon entropy [72] of the phenotypic distribution induced by the evolving developmental process for moderate environmental switching and sparse connectivity. Overfitting is indicated by reducing to less than four bits. For the case of sparse connectivity entropy converges to four bits indicating that each of the four modules vary independently. The sample size was 5×10^5.

S2 Appendix. Developmental Model

Following previous work [25], we describe the development of the embryonic phenotype to an adult phenotype by a continuous, non-linear and recurrent (i.e., it allows for feed-back connections) model of gene-regulatory networks [38, 39].

At each developmental time step, t, the phenotype of an individual organism is characterised by a collection of phenotypic traits, $P_t = \langle p_{t,1}, \dots, p_{t,N} \rangle$, where $p_{t,i} \in \mathbb{R}, \forall i$. The genotype is comprised of two parts: the direct effects on the embryonic phenotypic traits, $G_t = \langle g_{t,1}, \dots, g_{t,N} \rangle$, where $g_{t,i} \in \{-1, 1\}, \forall i$ and the regulatory interactions between the genes, b_{ij}, that determine the dynamical developmental process [41, 64, 73]. The regulatory interactions are represented by the matrix B.

The dynamics of the expression level for each gene depend on 1) the gene expression levels of the genes that is connected to and 2) the its pattern of connections, i.e., how strongly the respective gene is connected to its neighbouring genes. In the first time step, the embryonic phenotype is solely characterised by the direct effects of G ($P_0 = G$). Thereafter, at every developmental step the phenotypic traits are developed under the following set of difference equations [25, 74]:

$$p_{t+1,i} = p_{t,i} + \tau_1 \sigma(\sum_j b_{ij} p_{t,j}) - \tau_2 p_{t,i}, \tag{4}$$

where $\tau_1 = 1$ and $\tau_2 = 0.2$ indicate the maximal expression rate and the constant rate of degradation of the given gene product respectively. The second term in the right-hand side of equation (4) corresponds to the interaction term, the activity of which is limited by a non-linear, monotonic and bounded (sigmoid) activation function, $\sigma(x) = tanh(\alpha x)$, where $\alpha = 0.5$. Then, over a fixed number of developmental time steps, T (here $T = 10$), the embryonic phenotype is transformed into an adult phenotype, $P_a = P_T$, upon which selection can act. Both G and B are initialised at zero.

S3 Appendix. Varying Selective Environments

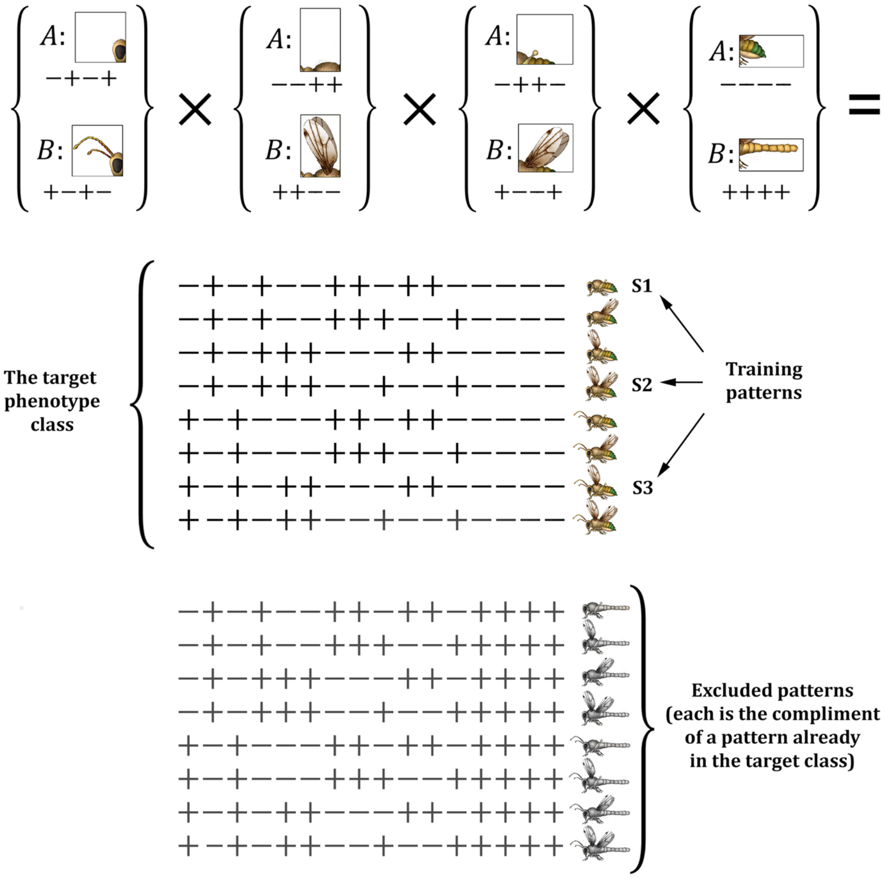
In this work, a set of related phenotypic targets is considered from the same family (as in [25, 34]). This guarantees that the environment changes in a systematic manner (i.e., shares common regularities invariant over time) — something which is ubiquitous in natural environments.

Since we are interested in modelling phenotypic variability, traits that are under constant selection are omitted from our model. We choose a simple family of modularly-varying targets. Modularity is widespread in the natural world and provides a simple way to test for generalised developmental oragnisations that are biological relevant [27, 29, 29, 41, 64, 75–77]. For simplicity, to model selection that varies in a modular manner, we assume an extreme form of modularity, namely separable modules [78]. Accordingly, selection on any trait is strongly interdependent with selection on other traits in the same module, but independent of selection on traits in other modules. Specifically, when a change in the environment occurs, if the direction of selection on a given trait changes, the direction of selection on all other traits in the same module also changes (this defines the modules). Selection thus favours two complementary states for each module that confer high fitness in different environments. Since the selection on each module is independent of selection on other modules, this means that there are 2^k possible high-fitness phenotypes, where k corresponds to the number of modules.

Here we assume a class of phenotypes consisted of equal sized modules (4 modules of 4 phenotypic traits each). The particular patterns chosen are irrelevant. So we pick one phenotype of 16 traits arbitrarily, here $(-+-+--++-++-----)$, and divide it into 4 equal modules (i.e., $(-+-+)$, $(--++)$, $(-++-)$ and $(----)$). Accordingly, for the phenotypic patterns that belong in the class, each module (block) can have 2 states: A or B; denoting a particular phenotypic sub-pattern or sub-goal (e.g., here the sub-goal for the first module can be either $(-+-+)$ (A) or $(+-+-)$ (B)). The class is thus comprised of 16 different modular patterns; all possible combinations of the sub-patterns (blocks) (see S5 Fig).

The time-invariant regularities here are the correlations between traits within any one module. The actual underlying structure of the given problem can thus be described by the block diagonal interaction matrix (see S1 Fig). The colour and the size of the squares in Hinton's representation indicate the sign and the magnitude of each correlation respectively. This clearly shows that selection on the traits within each module are strongly correlated with each other (positively or negatively depending on the combination of signs in the particular phenotypic pattern used), and no correlations between one module and another.

Complementary patterns here are also stable states of the evolved dynamical system as a result of Equation 4. The map described in Equation 4 is an odd function (i.e., symmetric with respect to the origin) since $f(-x) = -f(x)$. Accordingly, if R is a stable state of the system, i.e., $R = f(R)$, then $-R$ is also stable since $-R = -f(R) = f(-R)$. In order to focus on the more interesting (non-trivial) attractors that may arise, we limit the phenotypic space so as to ignore complementary targets (i.e., thus removing 8 of the patterns). Specifically, without loss of generality, we consider the phenotypic targets in which the sub-pattern in the last slot (trait positions:$13-16$) corresponds to state A: $\{-,-,-,-\}$, i.e., we focus on the top-half of the class as arranged in the lower part of Fig 1. Accordingly, each member of the other



S5 Fig. Modularly-varying environment. Target phenotypes varying from one another in a modular fashion. Each target phenotype consists of 4 modules of 4 phenotypic traits (i.e., 16 phenotypic traits in total). Each module can take two (complementary) states: A or B; denoting particular sub-patterns favoured by selection in different selective environments. The complete set of phenotypes is thus comprised of $2^4 = 16$ phenotypes, differing from one another in a modular fashion. The signs of phenotypic traits correspond to the direction favoured by selection in a given environment. Eight of the 16 possible phenotypes are designated as the target class (the other eight are merely the complement of a pattern already in the target class). For the main experiments, three patterns from the target class are used as 'training' patterns, i.e., selected for.

half of the class is the bit-wise complement of a member in the top half.

In this work, we want to examine the ability of the developmental system to 'learn' from past selective environments and generalise to new environments by producing novel phenotypes within the same class. Accordingly, to assay generalisation and the conditions that promote it, the population is evolved by exposure to a limited number of selective environments (< 8, i.e., a strict sub-set of the class). Otherwise, generalisation would not be relevant, since the population would have been exposed to all possible selective environments (i.e., all phenotypes in the class are presented). For this paper, we use the following example from this problem domain as a training set:

$$\begin{aligned} S_1 &= \{-,+,-,+\},\{-,-,+,+\},\{-,+,+,-\},\{-,-,-,-\}. \\ S_2 &= \{-,+,-,+\},\{+,+,-,-\},\{+,-,-,+\},\{-,-,-,-\}. \\ S_3 &= \{+,-,+,-\},\{+,+,-,-\},\{-,+,+,-\},\{-,-,-,-\}. \end{aligned} \tag{5}$$

In S5 Appendix, we explore sensitivity to this particular choice by examining generalisation from training on all possible proper subsets of the class.

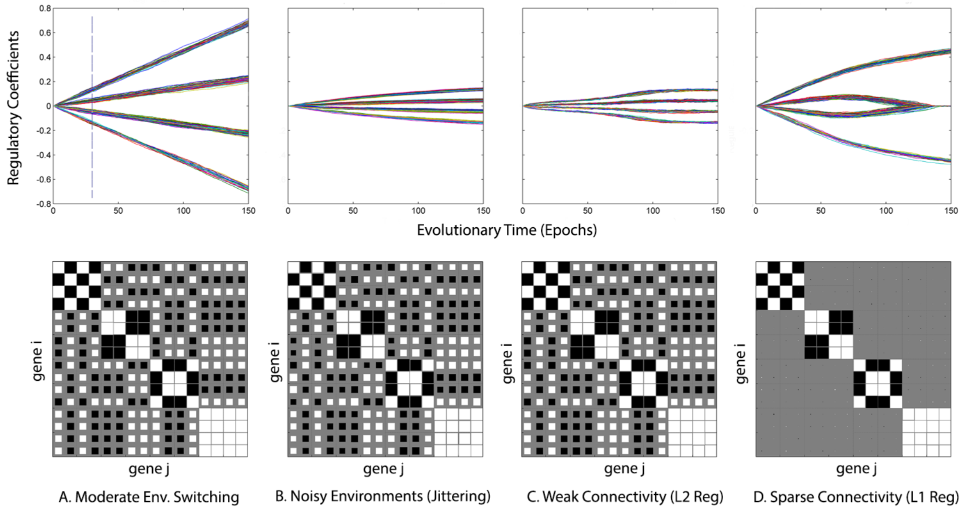
S4 Appendix. The Structure of Developmental Organisation

Here we show how costly interactions and noisy environments facilitate the emergence of more general and parsimonious developmental models. For this purpose, we monitor the evolution of regulatory interactions over evolutionary time in each evolutionary setting. The regulatory coefficients here correspond to the free parameters of the developmental model that determine the functional organisation of development.

We first analyse the evolution of regulatory coefficients in the control scenario, i.e., moderate rate of environmental change. Figure S6 Fig A shows that the ontogenetic interactions evolved under natural selection to reflect the correlations in the previously-experienced selective environments. As seen, the Hinton diagram of the evolved regulatory matrix appropriately matched the variance-covariance matrix of the past phenotypic targets (S6 Fig). The colour and the size of the squares in Hinton's representation indicate the sign and the magnitude of the respective correlations.

Yet natural selection did not directly select either *for* correlations, or *for* matching the exploration distribution to the fitness distribution of the phenotypic variants (i.e., training error minimisation). Natural selection selected *for* immediate fitness differences depending on how well adapted the organism was to its current selective environment; i.e., how close the produced adult phenotype was to the respective target phenotype. The evaluation of the developmental process performed here against the training and the test set was a post hoc analysis, and hence not part of the actual evolutionary dynamics.

In the same fashion as the nervous system [79], evolution does not try to analyse anything. It just tries to generate appropriate behaviour. The observed (correlation) learning behaviour of evolution can



A. Moderate Env. Switching

B. Noisy Environments (Jittering)

C. Weak Connectivity (L2 Reg)

D. Sparse Connectivity (L1 Reg)

S6 Fig. Evolution of regulatory coefficients in noisy environments and under parsimony pressure. The evolution of regulatory coefficients over evolutionary time and the Hinton diagram of the evolved regulatory coefficients (after epoch 150) for (A) moderate environmental switching, (B) noisy environments, (C) favouring weak connectivity and (D) favouring sparse connectivity. The vertical dashed line denotes when the ad-hoc technique of early stopping is used, i.e., the moment the problem of over-fitting begins. Favouring sparsity ignores the weak spurious correlations of the finite sampling noise and maintains the time-invariant ones.

be seen as a by-product of developmental systems' effort to produce high-fitness phenotypic variants in varied selective environments — optimise the actual functionality of the system. The system does not explicitly aim at inferring the target function, namely, the ideal G-P map that gives rise to proper system functionality in long-term (over certain genetic and environmental conditions). Nevertheless, we see that under certain conditions the system may discover a hypothesis (i.e., set of regulatory coefficients) closer to the target function, by producing phenotypic variants that are fitter in short term.

Figure S6 Fig B shows that under the presence of environmental noise, the regulatory interactions evolved towards smaller in magnitude weights. In particular, we observe that the rate of evolutionary change was decreased with evolutionary time giving rise to a plateau in the test error in Fig 3 B. The set of evolved regulatory coefficients here corresponds to the one we get if we stopped evolution the moment over-fitting begins, i.e., at the vertical dashed line in Fig 3 A. From Hinton diagram we can see that the relative importance between strong and weak correlations remained the same as in the case of the control run, i.e., only the magnitudes changed. Therefore, noise had a beneficial role on the evolution of genetic structures by making it difficult for natural selection to find configurations that over-fit past phenotypic

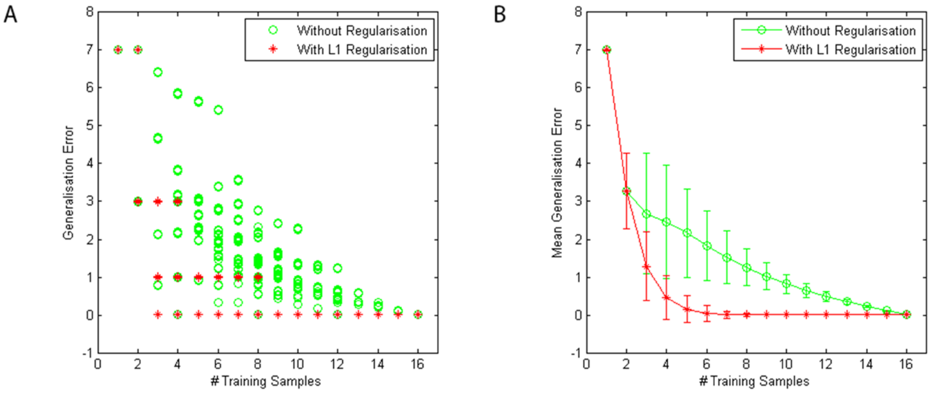
targets.

We observe similar results for the evolution of regulatory interactions under the pressure for weak connectivity (Figure S6 Fig C). In contrast to environmental stochasticity, however, favouring weak connectivity imposes strict constraints on the evolution of regulatory coefficients that prohibit them from growing bigger, i.e., providing a hard bound determined by the strength of parsimony pressure (see below). Accordingly, the regulatory coefficients initially increased until they reached a level that the further increase in the reproduction and maintenance cost of interactions was greater than the benefit of the developmental structure. Moreover, when properly tuned favouring weak connectivity exhibits the same behaviour as stopping early. Favouring weak connectivity (L_2-regularisation) can be understood as imposing inductive biases (i.e., additional constraints) in the evolution of regulatory interactions, punishing interactions (parameters) with extreme (high) magnitudes by applying a penalty proportional to their current magnitudes (as in weight-decay).

Lastly, Figure S6 Fig D illustrates how favouring sparse connectivity can exhibit a form of feature selection emphasising the relative importance of the strong correlations against the weak correlations. Specifically, we see that only the strongest (time-invariant) correlations persisted, while the weak (spurious) correlations, which arose as a result of the sampling process, were eliminated over evolutionary time. The strong correlations here (i.e., the block diagonal of the interaction matrix) correspond to the actual underlying modular structure of the environmental variation that remain invariant over time. Consequently, if the strength of parsimony pressure is large enough to ignore the spurious correlations, the evolved associations are (almost) identical to the variance-covariance matrix that describes the phenotypes family (see Figure S6 Fig). Favouring sparse connectivity (L_1-regularisation) can be understood as punishing interactions by equally applying a fixed penalty to all of the weights of the network. The amount of reduction is controlled by the hyper-parameter λ (see below); the higher its value, the higher the penalty applied, and hence the higher the level of sparsity. When properly tuned, favouring sparse connectivity leads to many zero weights, and thus the complexity of the model is reduced by removing degrees of freedom.

S5 Appendix. Favouring Sparse Connectivity in Different Training Sets

Experiments were also carried out for every possible training set as a strict sub-set of the test set. Firstly, all possible combinations, $\sum_{0\leq k\leq N}\binom{N}{k}=2^N$, were explicitly enumerated, where N indicates the



S7 Fig. Favouring sparse connectivity enhances phenotypic generalisation. (A) Phenotypic generalisation with and without the parsimony pressure for sparsity (L_1-regularisation) against all possible evolutionary scenarios (training sets), i.e., all possible combinations of distinct past selective environments drawn from the class. (B) Means and error bars of the generalisation performance of the evolved networks with and without the parsimony pressure for sparsity against different numbers of previously experienced selective environments. The cost of connection significantly enhanced evolvability in the majority of the cases. The interaction matrices here were determined using Hebb's rule.

number of patterns in the test set. Then, the respective developmental systems were determined following Hebb's rule with and without the selective pressure on the cost of connections (for optimal λ values). Hebbian learning was used here for computational tractability (65536 possible combinations), since it has been shown before that the interaction matrix evolves under natural selection in a Hebbian manner [25]. According to Hebb's rule, the pair-wise interactions are increased (or decreased) if the phenotypic traits are aligned (or not). The Hebbian matrix can be computed by computing the outer-product over the training inputs, i.e., the auto-correlation matrix. For the sake of comparison, the respective coefficient matrices were also tuned to be of the same average magnitude level as in the experiments above. These simulations allow us to draw some more general conclusions.

Overall, we find that the cost of connection significantly enhanced evolvability in the majority of the cases (Figure S7 Fig). As the number of observations is increased we observe an increase on average in evolvability, reaching zero generalisation error when $k = N$, even without incorporating the cost of connection. Interestingly, this was also true for some cases of 4, 8 and 12 patterns. We therefore see that different training sets entailed different information about the class, some of which were better representatives than others. For training sets consisted of more than half of the patterns in the class, we also observe that (optimally tuned) parsimony pressure for sparsity certainly resulted in perfect generalisation. On the other hand, in situations like the ones of 1 or 2 patterns the parsimony pressure had no effect on the generalisation performance of the network, and in some situations between 3 to 8 patterns it had little effect.

Acknowledgments

This work was partly funded by DSTLX-1000074615. Jeff Clune was supported by an NSF CAREER award (CAREER: 1453549). No data sets are associated with this publication.

References

1. Bedau MA, McCaskill JS, Packard NH, Rasmussen S, Adami C, et al. (2000) Open problems in artificial life. Artificial life 6: 363–376.

2. Adami C, Ofria C, Collier TC (2000) Evolution of biological complexity. Proceedings of the National Academy of Sciences 97: 4463–4468.

3. Lenski RE, Ofria C, Pennock RT, Adami C (2003) The evolutionary origin of complex features. Nature 423: 139–144.

4. Bedau MA (2009) The evolution of complexity. Springer.

5. Moczek AP, Sultan S, Foster S, Ledón-Rettig C, Dworkin I, et al. (2011) The role of developmental plasticity in evolutionary innovation. Proceedings of the Royal Society B: Biological Sciences : rspb20110971.

6. Wagner GP, Altenberg L (1996) Perspective: Complex adaptations and the evolution of evolvability. Evolution : 967–976.

7. Conrad M (1979) Bootstrapping on the adaptive landscape. BioSystems 11: 167–182.

8. Kirschner MW, Gerhart JC (1998) Evolvability. Proceedings of the National Academy of Sciences 95: 8420–8427.

9. Schlichting CD, Murren CJ (2004) Evolvability and the raw materials for adaptation. Plant Adaptation: Molecular genetics and ecology NRC research Press, Ottawa : 18–29.

10. Conrad M (1972) The importance of molecular hierarchy in information processing. Towards a theoretical biology 4: 222–228.

11. Pigliucci M (2008) Is evolvability evolvable? Nature Reviews Genetics 9: 75–82.

12. Riedl R, Jefferies RPS (1978) Order in living organisms: a systems analysis of evolution. Wiley New York.

13. Altenberg L (1995) Genome growth and the evolution of the genotype-phenotype map. In: Evolution and biocomputation, Springer. pp. 205–259.

14. Toussaint M (2002) On the evolution of phenotypic exploration distributions. In: FOGA. Citeseer, pp. 169–182.

15. Brakefield PM (2006) Evo-devo and constraints on selection. Trends in Ecology & Evolution 21: 362–368.

16. Gerhart J, Kirschner M (2007) The theory of facilitated variation. Proceedings of the National Academy of Sciences 104: 8582–8589.

17. Toussaint M, von Seelen W (2007) Complex adaptation and system structure. BioSystems 90: 769–782.

18. Braendle C, Baer CF, Félix MA (2010) Bias and evolution of the mutationally accessible phenotypic space in a developmental system. PLoS genetics 6: e1000877.

19. Smith JM, Burian R, Kauffman S, Alberch P, Campbell J, et al. (1985) Developmental constraints and evolution: a perspective from the mountain lake conference on development and evolution. Quarterly Review of Biology : 265–287.

20. Conrad M (1998) Towards high evolvability dynamics introduction. In: Evolutionary systems, Springer. pp. 33–43.

21. Yampolsky LY, Stoltzfus A (2001) Bias in the introduction of variation as an orienting factor in evolution. Evolution & development 3: 73–83.

22. Hansen TF (2003) Is modularity necessary for evolvability?: Remarks on the relationship between pleiotropy and evolvability. Biosystems 69: 83–94.

23. Pavlicev M, Cheverud JM, Wagner GP (2010) Evolution of adaptive phenotypic variation patterns by direct selection for evolvability. Proceedings of the Royal Society B: Biological Sciences : rspb20102113.

24. Pavlicev M, Hansen TF (2011) Genotype-phenotype maps maximizing evolvability: Modularity revisited. Evolutionary Biology 38: 371–389.

25. Watson RA, Wagner GP, Pavlicev M, Weinreich DM, Mills R (2014) The evolution of phenotypic correlations and developmental memory. Evolution 68: 1124–1138.

26. Pavličev M, Cheverud JM (2015) Constraints evolve: Context-dependency of gene effects allows evolution of pleiotropy. Annual Review of Ecology, Evolution, and Systematics 46.

27. Clune J, Mouret JB, Lipson H (2013) The evolutionary origins of modularity. Proceedings of the Royal Society b: Biological sciences 280: 20122863.

28. Clune J, Misevic D, Ofria C, Lenski RE, Elena SF, et al. (2013) Natural selection fails to optimize mutation rates for long-term adaptation on rugged fitness landscapes. In: GECCO (Companion). pp. 25–26.

29. Wagner GP, Pavlicev M, Cheverud JM (2007) The road to modularity. Nature Reviews Genetics 8: 921–931.

30. Brigandt I (2007) Typology now: homology and developmental constraints explain evolvability. Biology & Philosophy 22: 709–725.

31. Draghi JA, Parsons TL, Wagner GP, Plotkin JB (2010) Mutational robustness can facilitate adaptation. Nature 463: 353–355.

32. Kirschner MW, Gerhart JC (2006) The plausibility of life: Resolving Darwin's dilemma. Yale University Press.

33. Jacob F (1977) Evolution and tinkering. Science .

34. Parter M, Kashtan N, Alon U (2008) Facilitated variation: how evolution learns from past environments to generalize to new environments. PLoS Computational Biology 4: e1000206.

35. Kashtan N, Alon U (2005) Spontaneous evolution of modularity and network motifs. Proceedings of the National Academy of Sciences of the United States of America 102: 13773–13778.

36. Kashtan N, Noor E, Alon U (2007) Varying environments can speed up evolution. Proceedings of the National Academy of Sciences 104: 13711–13716.

37. Wagner A (1996) Does evolutionary plasticity evolve? Evolution : 1008–1023.

38. Vohradský J (2001) Neural model of the genetic network. Journal of Biological Chemistry 276: 36168–36173.

39. Vohradský J (2001) Neural network model of gene expression. The FASEB Journal 15: 846–854.

40. Fierst JL, Phillips PC (2015) Modeling the evolution of complex genetic systems: The gene network family tree. Journal of Experimental Zoology Part B: Molecular and Developmental Evolution 324: 1–12.

41. Lipson H, Pollack JB, Suh NP (2002) On the origin of modular variation. Evolution 56: 1549–1556.

42. Watson RA, Mills R, Buckley C, Kouvaris K, Jackson A, et al. (2015) Evolutionary connectionism: algorithmic principles underlying the evolution of biological organisation in evo-devo, evo-eco and evolutionary transitions. Evolutionary Biology : 1–29.

43. Watson RA, Szathmáry E (2015) How can evolution learn? Trends in Ecology and Evolution .

44. Friedlander T, Mayo AE, Tlusty T, Alon U (2013) Mutation rules and the evolution of sparseness and modularity in biological systems. PloS one 8: e70444.

45. Livnat A (2013) Interaction-based evolution: how natural selection and nonrandom mutation worktogether. Biology direct 8: 1.

46. Livnat A, Papadimitriou C, Dushoff J, Feldman MW (2008) A mixability theory for the role of sex in evolution. Proceedings of the National Academy of Sciences 105: 19803–19808.

47. Aldana M, Balleza E, Kauffman S, Resendiz O (2007) Robustness and evolvability in genetic regulatory networks. Journal of theoretical biology 245: 433–448.

48. Mengistu H, Huizinga J, Mouret JB, Clune J (2016) The evolutionary origins of hierarchy. PLOS Comput Biol 12: e1004829.

49. Arthur W (2006) Evolutionary developmental biology: developmental bias and constraint. eLS .

50. MacNeil LT, Walhout AJ (2011) Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. Genome research 21: 645–657.

51. Bishop CM, et al. (2006) Pattern recognition and machine learning, volume 1. springer New York.

52. Abu-Mostafa YS, Magdon-Ismail M, Lin HT (2012) Learning from data. AMLBook.

53. Kauffman SA (1993) The origins of order: Self-organization and selection in evolution. Oxford university press.

54. Gu X, Zhang Z, Huang W (2005) Rapid evolution of expression and regulatory divergences after yeast gene duplication. Proceedings of the National Academy of Sciences of the United States of America 102: 707–712.

55. Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov RR (2012) Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:12070580 .

56. Leclerc RD (2008) Survival of the sparsest: robust gene networks are parsimonious. Molecular systems biology 4.

57. Akaike H (1974) A new look at the statistical model identification. IEEE transactions on automatic control 19: 716–723.

58. Schwarz G, et al. (1978) Estimating the dimension of a model. The annals of statistics 6: 461–464.

59. Deng H, Runger G (2012) Feature selection via regularized trees. In: The 2012 International Joint Conference on Neural Networks (IJCNN). IEEE, pp. 1–8.

60. Soule T, Foster JA (1998) Effects of code growth and parsimony pressure on populations in genetic programming. Evolutionary Computation 6: 293–309.

61. Palmer ME, Feldman MW (2012) Survivability is more fundamental than evolvability. PloS one 7: e38025.

62. Masel J, Trotter MV (2010) Robustness and evolvability. Trends in Genetics 26: 406–414.

63. Rajon E, Masel J (2011) Evolution of molecular error rates and the consequences for evolvability. Proceedings of the National Academy of Sciences 108: 1082–1087.

64. Kashtan N, Mayo AE, Kalisky T, Alon U (2009) An analytically solvable model for rapid evolution of modular structure. PLoS computational biology 5: e1000355.

65. Dekel E, Alon U (2005) Optimality and evolutionary tuning of the expression level of a protein. Nature 436: 588–592.

66. Striedter GF (2006) Précis of principles of brain evolution. Behavioral and Brain Sciences 29: 1–12.

67. Cherniak C, Mokhtarzada Z, Rodriguez-Esteban R, Changizi K (2004) Global optimization of cerebral cortex layout. Proceedings of the National Academy of Sciences of the United States of America 101: 1081–1086.

68. Russell S, Norvig P, Intelligence A (1995) A modern approach. Artificial Intelligence Prentice-Hall, Egnlewood Cliffs 25: 27.

69. Forman G (2008) Quantifying counts and costs via classification. Data Mining and Knowledge Discovery 17: 164–206.

70. Galanti S, Jung A (1997) Low-discrepancy sequences: Monte carlo simulation of option prices. The Journal of Derivatives 5: 63–83.

71. Matoušek J (1999) Geometric discrepancy: An illustrated guide. Springer.

72. Shannon CE (2001) A mathematical theory of communication. ACM SIGMOBILE Mobile Computing and Communications Review 5: 3–55.

73. Wagner GP (1989) The biological homology concept. Annual Review of Ecology and Systematics : 51–69.

74. Wessels LF, van Someren EP, Reinders MJ, et al. (2001) A comparison of genetic network models. In: pacific Symposium on Biocomputing. volume 6, pp. 508–519.

75. Callebaut W, Rasskin-Gutman D (2005) Modularity: understanding the development and evolution of natural complex systems. MIT press.

76. Carroll SB (2001) Chance and necessity: the evolution of morphological complexity and diversity. Nature 409: 1102–1109.

77. Alon U (2006) An introduction to systems biology: design principles of biological circuits. CRC press.

78. Watson RA (2006) Compositional evolution: the impact of sex, symbiosis and modularity on the gradualist framework of evolution. Mit Press.

79. Anderson JA (1983) Cognitive and psychological computation with neural models. Systems, Man and Cybernetics, IEEE Transactions on : 799–815.