The Effect of Training Schedules on Morphological Robustness and Generalization

Edoardo Barba
University of Trento
Trento, Italy

Anil Yaman
Vrije Universiteit Amsterdam
Amsterdam, Netherlands

Giovanni Iacca
University of Trento
Trento, Italy

ABSTRACT

Robustness and generalizability are the key properties of artificial neural network (ANN)-based controllers for maintaining a reliable performance in case of changes. It is demonstrated that exposing the ANNs to variations during training processes can improve their robustness and generalization capabilities. However, the way in which this variation is introduced can have a significant impact. In this paper, we define various training schedules to specify how these variations are introduced during an evolutionary learning process. In particular, we focus on morphological robustness and generalizability concerned with finding an ANN-based controller that can provide sufficient performance on a range of physical variations. Then, we perform an extensive analysis of the effect of these training schedules on morphological generalization. Furthermore, we formalize the process of training sample selection (i.e., morphological variations) to improve generalization as a reinforcement learning problem. Overall, our results provide deeper insights into the role of variability and the ways of enhancing the generalization property of evolved ANN-based controllers.

CCS CONCEPTS

• **Theory of computation → Evolutionary algorithms**; • **Computing methodologies → Neural networks**.

KEYWORDS

Neuroevolution, Generalization, Robustness, Variability

1 INTRODUCTION

The robustness and generalizability of artificial neural network (ANN)-based controllers are crucial properties for providing reliable performance when facing changes. The physical world is, in fact, highly uncertain and is subject to variations of various kinds. To operate effectively, controllers must be robust to these variations, meaning they should be able to perform effectively in diverse conditions.

Robustness is typically defined as the ability to maintain sufficient performance in case of small perturbations relative to the training data. Whereas, generalization is the ability to handle a wider range of differences, especially concerning scenarios that are not encountered during the training process [28].

For instance, in the case of a walking robot, an ANN-based controller should be robust to varying terrains, obstacles, or changes to the robot morphology, e.g. due to damages. A controller trained to handle only one specific morphology or one specific environment would reduce its robustness and generalizability, making it more fragile to perturbations [7, 27].

Robustness and generalizability can also be vital for reducing the simulation-to-reality gap (reality gap for short), which refers to the mismatch between simulated and real-world environments [30]. The training conditions in simulation may not capture the noise and unexpected situations that can be encountered in the real world. While optimizing for *all* possible unexpected situations may not even be possible (because such situations are most likely unknown), it is plausible to think that more generalizable controllers should handle such situations better than highly specialized controllers.

Because of the importance of these properties, there is a growing interest in developing controllers that exhibit robustness across a broad spectrum of morphological variations and can generalize to unseen morphologies [4]. The main approaches to handling unexpected situations are retraining, relying on repertoires [8], or online adaptation [23, 29, 38]. However, these approaches can be costly since they rely on a form of learning or memorizing a repertoire. On the other hand, a generalist controller may be able to handle a large range of variability without the need for adaptation [7].

In various domains, it has been observed that exposing learners to more variability can be an effective strategy for enhancing their robustness and generalization capabilities [19]. More variability tends to make initial learning more challenging but eventually leads to more general and robust performance. This core principle has been discovered in many domains (e.g., visual perception [39], motor learning [18, 31], and language acquisition [26]). In supervised learning tasks, training variability can be enhanced, for instance, through techniques like data augmentation [1, 13, 17]. More variability during training is therefore essential to generalize learning, as it facilitates the formation of more abstract knowledge [26]. Variability is also essential for an "antifragile" system, a concept introduced by Taleb [34] to describe systems that can benefit from exposure to variations. Taleb distinguishes between fragile, robust, and antifragile systems, asserting that complex systems fall into the antifragile category due to their opaque nature. ANNs, with their nonlinear relationships, can benefit from environmental variations, making adaptive controllers antifragile as they navigate uncertainties. Introducing too much variability during training,

however, can be counterproductive, as high variability can make learning more difficult when learners are in the very early stages of acquiring a target behavior [19]. It is therefore crucial to understand the best way of introducing variability to improve generalization.

One approach to improve generalization by introducing variability is presented in [7], where the authors employ Evolution Strategies to optimize ANNs to control different body shapes in locomotion tasks. They enable generalization by introducing a different morphology (generated by modifying body parameters such as leg length or width) at each generation and use it as the evaluation fitness for that generation.

The present work builds upon a similar idea to study the effects of different training schedules used to evolve the ANN in depth. Our main focus is to investigate how variability influences the learning and generalization of evolved ANN-based controllers. To this aim, we perform a thorough analysis using three OpenAI Gym tasks: Bipedal Walker, Walker2D, and Ant. The training methodologies follow the algorithm presented in [7], introducing morphological variations in each generation. However, differently from [7], here we specifically focus on a thorough study of different training schedules.

In our analysis, we find that increasing the diversity of morphologies during training improves both robustness and generalization. Sampling variations from a continuous domain (i.e., modifying various morphological parameters) during training provides more robustness than a discrete set of variation possibilities. Moreover, using the Beta distribution for sampling the variations for training enhances generalization. We also confirm the observation reported in [7], that the order in which the morphological variations are introduced during training significantly impacts generalization performance. Finally, we formalize the process of online learning of training sample selection as a reinforcement learning problem and show its learning dynamics. In particular, we study if some samples (i.e., morphological variations) are preferred, and how this impacts the final generalization. Taken together, these results highlight the importance of introducing variation during training and its effects on the generalization capabilities of ANN-based controllers.

The rest of the paper is organized as follows. The next section presents the background and related works. Section 3 describes our methods. Then, Section 4 and Section 5 present the experimental setup and the numerical results. Finally, Section 6 concludes this work.

2 BACKGROUND AND RELATED WORKS

Neuroevolution through Evolution Strategies. ANN-based controllers are used in a wide range of control tasks in simulation and the real world. Often these tasks require black-box optimization approaches due to the difficulty of obtaining a closed-form objective function for which gradient-based approaches can be used. Neuro-evolution is a derivative-free approach that is used to optimize the weights (and, potentially, the structure) of ANNs using evolutionary algorithms [11, 32]. Due to the large dimensionality of the ANN search space, however, neuro-evolutionary approaches typically suffer from scalability issues. In this regard, Evolution Strategies (ES) have been shown to scale particularly well, due to their capability of estimating the gradient based on sampling [9, 35]. Therefore,

ES has been applied on a wide range of control tasks that require optimization in large dimensional parameter spaces [24, 36].

Morphological Robustness and Generalizability. As outlined by Raviv et al. [19], variability has been demonstrated to be a crucial aspect in improving the generalization capabilities of learners. In the real world, one can say that learning intrinsically requires some form of generalization. The prevailing relation between the effects of variability and generalization, observed across various domains, suggests that learning from less variable input tends to occur rapidly but may struggle to generalize to new stimuli. Conversely, learning from more varied input typically starts slower but often results in superior generalization.

Variability, however, has many dimensions. The same authors of [19] identified four distinct types: Numerosity (more or fewer examples), Heterogeneity (similarity between samples), Situational Diversity (more or less variable conditions), and Scheduling (different order of samples). The latter, in particular, is also related to the timing of variability, which can directly impact the learning process. It has been shown that the impact of variability can vary significantly, depending on the stage of the learning process [3, 19].

The common intuition is that excessive variability during the initial stages of learning can hinder learning. This is because, when individuals are just starting to acquire a new skill or behavior, they may encounter difficulties if they are exposed to too much variability too soon [16, 20]. Instead, beginning with simpler and less variable examples tends to facilitate learning. This concept is the foundation of curriculum learning, a methodology inspired by these learning patterns where skills are progressively built upon simpler concepts [25]. Eventually, exposing a learner to variability is however necessary to improve its ability to generalize learning to new contexts, as it facilitates the formation of more abstract knowledge. However, deciding when to expose the learner to variability is not trivial. In the following, we will refer to the way to decide which variations a learner should be exposed to, and when, with the term *training schedule*.

A preliminary study on training schedules was conducted in [7], aiming to improve the generalization of ANN-based controllers by exposing them to variations in each generation of an evolutionary process. Focusing in particular on the case of morphological variations, the authors showed that this approach provides better generalizability performance relative to a baseline when the average performance on all morphological variations is maximized.

In the case of morphological variations, finding ways to achieve robustness can help maintain the performance in case of unexpected changes in the physical structure of the agent without requiring retraining [7]. We refer to *morphological robustness* as the ability of a controller to maintain stable performance despite small perturbations (w.r.t. the training scenarios [27]) in the morphological parameters of the agent (e.g., a robot) controlled by that controller. Whereas, *morphological generalizability* refers to the ability to maintain a sufficient performance in case of larger changes in the morphological parameters of the agents [2].

3 METHODS

The training process of our approach is illustrated in Algorithm 1. We use a similar approach proposed by [7], as we define a set of

morphologies that are introduced during the evolutionary process. These morphologies are concerned with the parameter ranges of the physical structure of the agents. Thus, we consider a training set $M = \{m_1, m_2, \ldots, m_n\}$ that is used during training, and a validation set $V = \{v_1, v_2, \ldots, v_n\}$ that is used for validating the generalization score. M and V are initialized or sampled from the same morphological parameter ranges, but they are used differently, because V consists of a discrete and predefined set, whereas M may be defined before the evolutionary run, or can be generated automatically during the process. We then have a training schedule T that determines how the morphologies in the training set M are introduced during the evolutionary process.

Algorithm 1 Training Process

1: Initialize G as an empty set   ▷ Generalist set
2: Initialize M as the training set of morphologies
3: Initialize T as the used training schedule
4: Initialize V as the validation set of morphologies
5: **function** EVOLUTIONARYPROCESS(M, T, V)
6:      $I_{\text{best}} \leftarrow \{\}$
7:      $f \leftarrow \infty$ ▷ Assuming minimization
8:      $I \leftarrow$ initialize()   ▷ Initialize an individual
9:      **for** $i <$ maxGenerations **do**
10:         $m_i \leftarrow \Psi(M, T, i)$ ▷ Choose m_i from M based on T and i
11:         $f_{m_i} \leftarrow$ evaluate(I, m_i)   ▷ Sample solutions from I and evaluate them on m_i
12:         $I_{\text{best}} =$ individual with highest fitness on m_i
13:         $I \leftarrow$ updateParameters(I, f_{m_i})   ▷ Update I based on fitness
14:         **for each** v in V **do**
15:            $f_v \leftarrow$ evaluate(I_{best}, v)
16:         $f_{\text{best}} =$ mean fitness of I_{best} on all v in V
17:         **if** $f_{\text{best}} < f$ **then**   ▷ Assuming minimization
18:            $G \leftarrow$ add(I_{best})   ▷ Add to the generalist set
19:            $f \leftarrow f_{\text{best}}$

The training process begins with an empty "generalist set" G, followed by the generation of the initial population, where each individual represents the weights of the ANN. Then, at each generation i, the population is evaluated based on the fitness value obtained on a morphology m_i. Once each individual is evaluated, the one with the best-obtained score is selected, and its average fitness (f_{best}) across all morphologies in V is computed. The performance of a controller I_{best} on a set of different morphologies V is defined as the average performance evaluated on all morphologies in V. This is mathematically expressed as:

$$f_{\text{best}} = \frac{1}{|V|} \sum_{j=1}^{|V|} \text{evaluate}(I_{\text{best}}, v_j) \tag{1}$$

where f_{best} represents the average performance of the controller I_{best} over morphologies in V, $|V|$ denotes the total number of morphologies in set V, and evaluate(I_{best}, v_j) is the evaluation function that returns the score of the controller I_{best} on morphology v_j.

During the evolution process, every time a controller achieves a better average fitness score on the validation set V, this controller is added to the generalist set G. Therefore, at the end of the process,

the last controller added has the highest average fitness score on V. Note that in the algorithm, the Ψ function determines which morphology to select at generation t based on the type of the training schedule T.

For the evolutionary process, we use the eXponential Natural Evolution Strategy (xNES) algorithm [12] available in the EvoTorch implementation [22][1].

3.1 Training Schedules

To examine the impact of various ways of introducing variability during the evolutionary process, we define several training schedules. We refer the reader to the Supplementary Material for visualization and additional information regarding these training schedules, which are detailed next.

Discrete Random. This training schedule serves as a baseline. It involves a training set comprising 36 morphologies, generated with parameters with equal intervals within the possible morphological parameter space. At each generation, the morphology is chosen *randomly with uniform probability* from one of these morphologies.

Discrete Incremental. The discrete incremental training schedule uses the same morphology set as the discrete random but it selects the morphological variations in an incremental order on 2-dimensional parameter space starting from the initial x and y morphological parameters and incrementing the parameters along the x-axis first, followed by the increments on the parameters along the y-axis. Morphologies therefore are defined first with small body parameters (e.g., leg length) and then such parameters gradually increase. When the maximum parameter for x and y is reached, the process returns to the initial state. Defining morphologies in this way ensures a gradual transition between morphologies, minimizing large leaps from one morphology to another, and thus facilitating the learning process.

Uniform. This schedule samples morphology parameters randomly with uniform probability from the 2-dimensional continuous space at each generation. This differs from the discrete case in providing more variability.

Gaussian. Gaussian training schedule uses a Gaussian distribution that is centered in the center of the 2-dimensional morphology parameter space with $\Sigma = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$. The rationale behind this choice is to moderate the extent of variation introduced during the sampling focused in the center of the morphology space.

Cauchy. To introduce more variability, but still centered around the center of the morphology space, we used the Cauchy distribution. By leveraging a fat-tailed distribution like Cauchy, there is more probability of selection of variation far from the center.

Beta. We use the Beta distribution to sample the morphology parameters with α and β parameters set both to 0.1. This schedule primarily samples from the outer edges of the parameter space, resulting in significant variability during training. Each generation encounters markedly different morphologies, potentially leading to diverse behaviors required for effective control. This strategy aims

---

[1]The code of our algorithm is publicly accessible at: https://github.com/edoardobarba/evolving-generalist-controllers

to enhance generalization capability by exposing the trained ANN to a wide range of morphologies.

3.2 Multi-armed Bandit Approach

This section presents a method to learn to dynamically select training morphologies during the evolutionary process. Indeed, certain morphologies may be more useful than others for improving the generalizability of the controllers. Thus, to achieve that, we formalize the learning process as a reinforcement learning approach on a Multi-Armed Bandit (MAB) problem [15, 33, 37]. Based on this formalization, we can keep a utility estimate of each morphological variation. Then, we can make a selection based on the utilities. After this selection and evaluation, we can update the estimate of the utility of the selected variation based on its reward, that is whether or not this selection provided an improvement in generalization.

Formalization. We have a N number of arms $\{a_1, \ldots, a_N\}$ corresponding to the set of morphological variations $\{D_1, \ldots, D_N\}$. At each generation, we choose a morphology to evaluate, and subsequently update its utility based on the observed reward r. We characterize the utility of each morphological variation k with a value θ_k ranging between 0 and 1. The agent initially holds a prior probability distribution, which is subsequently updated. The distribution is represented by Beta distributions, characterized by parameters α_k and β_k. Therefore each morphology k is associated with its own α_k and β_k parameters. The probability density function is given by:

$$f(\theta_k | \alpha_k, \beta_k) = \frac{1}{B(\alpha_k, \beta_k)} \theta_k^{\alpha_k - 1} (1 - \theta_k)^{\beta_k - 1}. \qquad (2)$$

Reward and Update. After each generation t of the training process, the α and β parameters of all morphologies are updated based on a reward. The generalization capability f, defined as the average score across all morphologies in the validation set, serves as the basis for determining the reward. Specifically, a reward r_t is granted if the chosen morphology yields an improvement in the generalization capability compared to the moving average of the capability over the preceding s generations:

$$r_t = \begin{cases} 1, & \text{if } f_t > \text{average of } f \text{ over the last } s \text{ generations} \\ 0, & \text{otherwise} \end{cases} \qquad (3)$$

where s represents a hyper-parameter, set to 10 for our experiments.

We can then leverage the "conjugacy" property of the Beta distribution, which states that the posterior distribution of a Beta distribution is also a Beta distribution. This property simplifies the updating process of the parameters α_k and β_k after each training iteration. Specifically, after each generation, the parameters α_k and β_k of all the morphologies in the validation set can be updated with the following rule [10]:

$$(\alpha_k, \beta_k) = \begin{cases} (\hat{\alpha}_k, \hat{\beta}_k), & \text{if } k \neq x^t \\ (\hat{\alpha}_k, \hat{\beta}_k) + (r_{tb}, 1 - r_{tb}), & \text{if } k = x^t \end{cases} \qquad (4)$$

where

$$\hat{\alpha}_k = (1 - \gamma)\alpha_k + \gamma\alpha_0 \quad \text{and} \quad \hat{\beta}_k = (1 - \gamma)\beta_k + \gamma\beta_0,$$

and γ controls how quickly uncertainty is injected into the system. Moreover, x^t denotes the morphology chosen at time step t (i.e., at

generation t), whereas α_0 and β_0 represent the prior probabilities. Since the utility of each morphology varies during the training process (e.g., some morphologies can be more useful at the beginning of the training process while others at the end), we can consider the system as non-stationary. To address this system property, inspired by the methodology proposed in [15], one simple approach involves diminishing the significance of past observations at each step with the introduction of a tunable decay ratio γ.

Morphological Variation Choice. At each generation, the choice of which morphology to use relies on Thompson Sampling [10], namely, We choose the morphology corresponding to the highest θ^k value (i.e., $\text{argmax}_k \theta^k$). In essence, Thompson Sampling introduces a balance between exploration (trying out new morphologies) and exploitation (sticking with the best-known ones), allowing the model to learn and adapt based on the outcomes of its actions, leading to more informed and effective decisions over time.

4 EXPERIMENTAL SETUP

Environments. Experiments were conducted in various OpenAI Gym continuous control problems [14], typically employed for testing reinforcement learning algorithms. Specifically, we employed the *Bipedal Walker* environment from *Box2D* [5], along with *Walker2D* and *Ant* from *MuJoCo* [36]. In each of these environments, two body parameters were varied to train and test the ANNs on diverse morphologies. These body parameters are leg width and leg length for Bipedal Walker, and upper leg length and lower leg length for both Walker2D and Ant. The parameter ranges for the training and test sets differ across control tasks and are reported in Table 1.

The ANN-based Controllers. The controller used in this work is a fully connected feedforward ANN with the same topology across all tasks. This topology comprises input, hidden, and output layers, with the variation among tasks lying in the number of neurons. The network structures utilized in this paper are derived from [7], which, in turn, draws from prior investigations [24]. The hidden layer comprises 20 neurons, including one bias neuron, across all three tasks. The input layer consists of 24 neurons for the Bipedal Walker task, 17 neurons for the Walker2D task, and 27 neurons for the Ant task, respectively. Conversely, the output layer consists of 4 neurons for the Bipedal Walker task, 6 neurons for the Walker2D task, and 8 neurons for the Ant task, respectively. Finally, the tangent hyperbolic (tanh) activation function is employed for all neurons in the network.

Validation Sets. A validation set is utilized to test the generalizability of controllers during training. For all three tasks, the validation set encompasses 36 morphologies, with the parameter values within the ranges outlined in Table 1, under the "Training set" column. For visual representations of the validation sets for the three tasks, please consult the Supplementary Material.

Experimental parameters. For the Bipedal Walker task, the experiment employs 3000 generations, with each episode consisting of 1000 steps. For the Walker2D and Ant tasks, the number of generations is set to 5000, with each episode comprising 1000 steps.

Table 1: Ranges of parameters for each task.

Task	Parameter Set	Training set	Testing Set	Step
Bipedal walker	Leg Length	[7, 17]	[3, 6] ∪ [18, 21]	2
	Leg Width	[24, 44]	[16, 23] ∪ [45, 52]	4
Walker2d	Lower leg length	[0.3, 0.425]	[0.225, 0.25] ∪ [0.45, 0.5]	0.025
	Upper leg length	[0.4, 0.65]	[0.25, 0.35] ∪ [0.7, 0.8]	0.05
Ant	Lower leg length	[0.5, 1.5]	[0.2, 0.4] ∪ [1.6, 1.9]	0.1
	Upper leg length	[0.7, 1.7]	[0.4, 0.6] ∪ [1.8, 2.1]	0.1

Evaluation Metrics. The evaluation phase aims to measure the robustness and generalization capabilities of the ANN. As mentioned, we refer to robustness as the ability to control morphologies within the parameter ranges of the training set, while to generalization as the ability to control morphologies outside these parameter ranges, which have not been encountered during the training process. The evaluation approach draws inspiration from the methodology introduced by Packer et al. [6] for assessing the generalization of deep reinforcement learning controllers. Namely, we test robustness to in-distribution samples (interpolation) and generalization to out-of-distribution samples (extrapolation). Interpolation implies that agents should perform well in controlling morphologies where parameters are similar to those seen during training. Extrapolation requires agents to perform well in morphologies where parameters are different from those seen during training. We utilize two distinct evaluation sets to address these aspects:

(1) **Training morphologies**: a set of morphologies with the same parameter ranges used for training allows us to assess the robustness of the controller on *in-distribution samples*.
(2) **Testing morphologies**: a set of morphologies with parameters ranges outside those used for training, to evaluate the ability to generalize on *out-of-distribution samples*.

We can therefore define the average performance obtained on morphologies on the training set as a measure of *robustness* within the context of the data the controller has been exposed to during training. Meanwhile, the average scores on morphologies on the test set provide insight into its ability to *generalize* beyond the data it has seen during training. For a visual representation of the training and testing sets, refer to the Supplementary Material.

Statistical Analysis. The results presented in the following Section represent the median values obtained from 30 independent evolutionary training runs. To determine the statistical significance of the observed differences, the Mann-Whitney U test [21] was utilized with confidence level $\alpha = 0.05$. In the figures presenting the results, the statistical test outcomes are depicted using bars with red asterisks placed above the box plots, denoting groups that exhibit significant differences.

5 RESULTS

We present our results separately analyzing first the effect of the heterogeneity and of the scheduling of the training morphologies. Then, we present the results of the MAB approach. Finally, we analyze the distribution of morphologies selected during training.

5.1 Effect of the Heterogeneity of Morphologies

For this analysis, we consider only the Bipedal Walker task. Figure 1 illustrates the performance results on Bipedal Walker of the different training schedules based on sampling distributions on the testing morphologies that are not introduced during training.
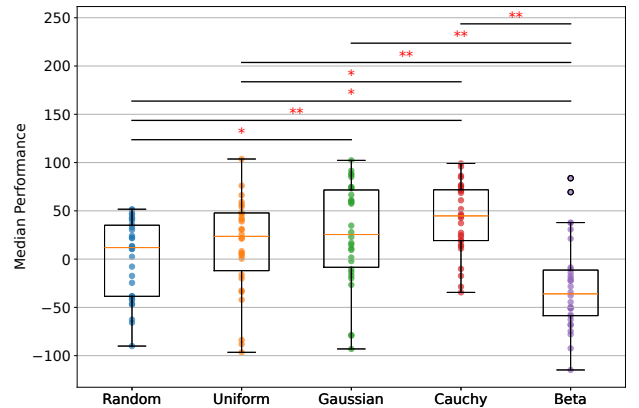

Figure 1: Performance comparison of training schedules on the testing morphologies on Bipedal Walker. * indicates $p < 0.05$, ** indicates $p < 0.01$, denoting statistical significance levels.

The figure shows the average, lower, and upper quartile obtained by 30 independent evolutionary runs. The Beta training schedule obtains better performance on the testing morphologies (assuming minimization), with a median score of -35.99. The p-values show that there is a statistically significant difference between Beta and all the other distributions.

Analyzing the performance of the morphologies on the training morphologies, we observe a different trend where the Uniform distribution stands out as the most effective, with a median performance of -254.40, and Beta as the worst one, with a median performance of -218.66. However, a statistically significant difference exists only between the Uniform and Beta distributions, evidenced by a p-value of 0.004.

While the differences between these training schedules on the training morphologies are not as pronounced, we see that on the testing morphologies, the Beta training schedule leads to more generalization. This can be attributed to the fact that morphologies sampled with the Beta distribution lie closer to the "border" of the

training parameter space, thereby resembling the morphologies encountered during testing.

Similarly, if we compare the results on all morphologies combining the training and test sets, the Beta training schedule still demonstrates superior performance, achieving a score of -123.55, with a statistically significant difference from all other approaches. This aligns with our earlier discussion: introducing increased variability during training, especially in the bordering cases, tends to improve the controller's ability to extend its performance to unfamiliar scenarios.

5.2 Effect of Scheduling of Morphologies

This section shows the results of the investigation on whether altering the order in which morphology variations are introduced during the evolutionary process yields different outcomes in terms of robustness and generalization. The central question we aim to address through this experiment is: "Does modifying the order of samples impact learning and generalization?" In order to answer this question, we conducted a comparative analysis between two training methodologies, namely the Random and the Incremental training schedules. In terms of heterogeneity, the training morphologies utilized are the same for both schedules. For this analysis, we consider Bipedal Walker and Walker2D.

Bipedal Walker. On the Bipedal Walker, the controllers trained using the *Incremental* training schedule achieved a median score of -248.94 on the training morphologies while scoring a median of -2.21 on the testing morphologies. In contrast, the controllers trained with the Random schedule achieved scores of -217.25 and 11.92 on the training and testing morphologies, respectively. Although the performances on the testing morphologies are comparable, with no significant difference, a notable disparity emerges on the training morphologies, further evidenced by a p-value of 0.01. Figure 2 illustrates these results. As can be seen from the figure, the controllers obtained with the Incremental training morphologies exhibit superior performance on the training morphologies compared to the one obtained with the Random schedule.
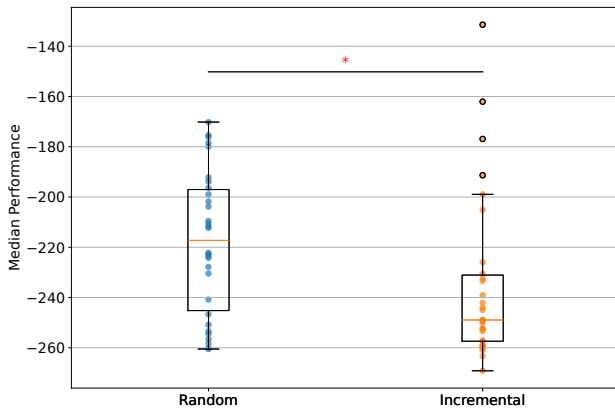

Figure 2: Performance comparison on the training morphologies on Bipedal Walker. * indicates $p < 0.05$, ** indicates $p < 0.01$, denoting statistical significance levels.

Walker2D. We observe similar trends on the Walker2D task. Notably, the controllers trained with the Incremental training schedule led to a better performance compared to those trained with the Random schedule on both the training and testing morphologies. Specifically, they achieved a median score of -1450.41 and -991.89 on the training and testing morphologies, respectively, while the Random schedule yielded scores of -1094.39 and -887.98. However, no statistically significant difference is observed in both sets.

These results highlight the significant impact of the order in which morphologies are presented during training. It is worth emphasizing once more that the set of morphologies remains identical; only the selection order differs. It is very interesting to observe the significant impact on performance that arises simply from altering the order in which morphologies are presented. Incrementally increasing the values of body parameters fosters the development of controllers that exhibit greater resilience to changes in morphologies, as also found in [7].

5.3 Multi-armed Bandit Approach

This section presents the results of the MAB approach for morphology selection during training, taking the Random schedule as the baseline. It is important to highlight that all the training morphologies are the same; the only difference lies in the order in which morphologies are selected. In this set of experiments, we consider the three tasks, namely Bipedal Walker, Walker2D, and Ant. Figure 3 visualizes these results.

Bipedal Walker. Although the difference was not statistically significant, the MAB approach achieved a better median score of -245.41 on training (relative to -217.26 for Random) and a median score of 5.52 on testing morphologies (relative to 11.92 for Random) On the training morphologies, the ANN controllers trained with the MAB approach demonstrated performance superior to the Random schedule, meaning that the MAB strategy used for training was effective. On the testing morphologies instead, it reached comparable performances with Random.

Walker2D. In the Walker2D task, the MAB approach was significantly better relative to Random. It achieved a median score of -1433.21 on the training morphologies, compared to -1325.01 for the Random schedule. Conversely, on the testing morphologies, it reached a median score of -756.76, whereas the Random schedule scored -562.05. Remarkably, on both the training and testing morphologies, the network trained with the MAB approach outperformed the Random training schedule, with a statistically significant difference on the test set.

Ant. Finally, on the Ant task, the MAB approach showed superior performance when compared to the Random schedule, both in the training and testing morphologies, although statistical significance was not achieved. On the training morphologies, the MAB approach achieved a median score of -3824.95, compared to -3779.36 obtained by the Random schedule. On the testing morphologies, instead, the MAB approach yields a median score of -2169.75, while the Random schedule achieves -2139.04.
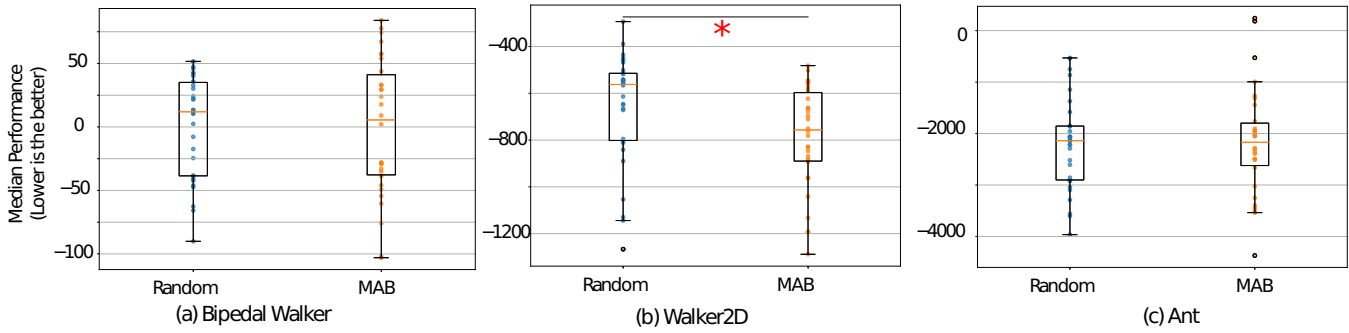
Figure 3: Performance comparison of Random schedule and the Multi-Armed Bandit approach on the testing morphologies on Bipedal Walker, Walker2D, and Ant. * indicates $p < 0.05$, ** indicates $p < 0.01$, denoting statistical significance levels.

Figure 4: Heatmap illustrating the frequency of selection for each morphology during training on Walker2D. Each cell in the heatmap represents a unique morphology, and the numerical value within the cell indicates the frequency with which that morphology was selected during training. This frequency is computed as the average occurrence across 30 independent training runs.

5.4 Analysis on Morphology Selection

An interesting analysis to perform is to look at which morphologies are selected most frequently, averaged over the 30 runs, during training.

Walker2D. Figure 4 presents a heatmap illustrating the frequency of each morphology's usage for Walker2D. The morphologies selected the most are represented by darker colors. Analysing the heatmap we can see that the morphologies selected the most are the ones in the upper-right region of the morphological parameter space. The MAB approach exhibits a non-uniform selection pattern for morphologies, particularly focusing on the upper-right quadrant (morphologies with high upper leg length and high lower leg length). The morphology selected the most is the one with an upper leg length of 0.4 and lower leg length of 0.6, used 180 times, while the least used was used 114 times. The MAB approach identifies

this region as prone to improvements, while other regions (such as the lower-left quadrant) may present challenges, resulting in limited enhancements.

Figure 5 presents heatmaps illustrating the average performances across all morphologies obtained using two different scheduling strategies: the Random schedule (Figure 5a) and the MAB approach (Figure 5b). The same analyses for Bipedal Walker and Ant are reported in the Supplementary Material. Each cell in the heatmaps corresponds to a specific morphology. Unlike the frequency heatmap, this visualization incorporates morphologies from both the training (inside the red square) and testing morphologies. The shading of each cell indicates the average performance of the ANN controllers on that particular morphology, with darker shades indicating better scores. These averages are computed over all controllers obtained from 30 runs. Observing the performance on the training morphologies, in the case of the MAB approach, the most significant improvement is notably concentrated in the upper-right region of the space (within the red square), coinciding with the most frequently chosen morphologies. This indicates that the MAB approach has effectively enhanced performance in a specific area of the space where improvement was feasible. Selecting more morphologies in that region of the space allowed to evolve ANN-based controllers with a greater capability to control those types of morphologies, without affecting performances in other parts of the morphological space.

Bipedal Walker. In Bipedal Walker (heatmap reported in the Supplementary Material), the MAB approach favors morphologies located in the lower-right region of the morphology space (high leg width and low leg length). Morphologies in this region are utilized roughly twice as often as those in the left upper quadrant. For instance, the morphology with leg width of 17 and leg height of 24 was selected an average of 135 times, whereas the morphology with leg width of 7 and leg height of 44 was chosen an average of 61 times. While it may seem logical to assume that the lower-right part was prioritized because it is the more complex part of the morphology space to control, this interpretation is not entirely accurate. The most challenging area within the morphology space is not the lower-right part, but rather the upper-right section appears to pose greater difficulty for the ANN-based controller. This is also evident from the heatmaps provided in the Supplementary Material, where the controller achieves higher scores in the upper

(a) Random Schedule

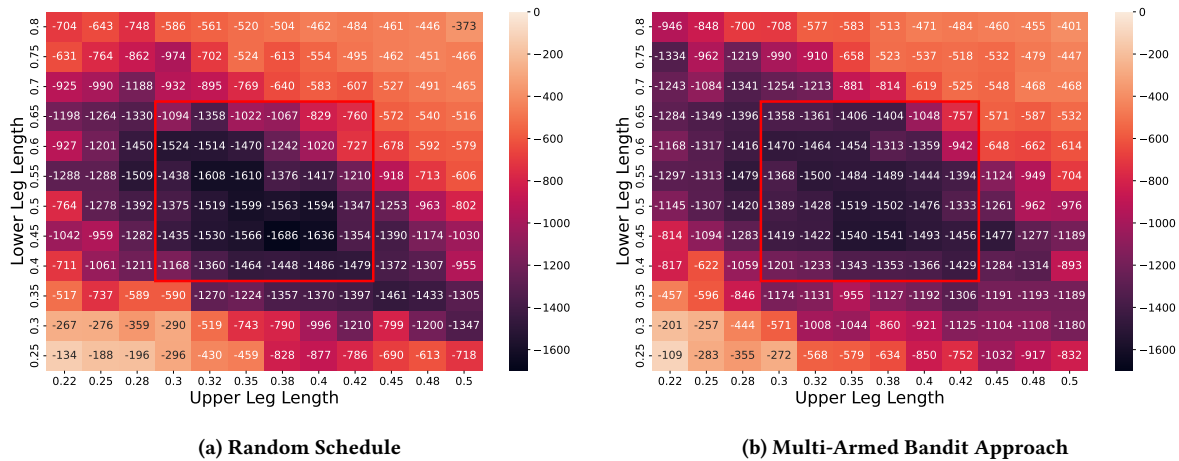(b) Multi-Armed Bandit Approach

Figure 5: Heatmap illustrating the average performance achieved on all morphological variations on Walker2D using the Random and MAB training schedules. The cells that are within the square with red borders show the training morphologies. The morphological parameters that are beyond these borders are used for the testing morphologies.

right region, indicating poorer performance. The focus on training with morphologies in the lower right section could be attributed to the potential for performance improvement in this area. In contrast, the upper right section may have reached a point where further improvements are no longer possible. Therefore it may be beneficial to focus on that part because improvements are more readily achievable. Overall, on this task the strategy employed by the MAB approach improves the performances compared to the Random training schedule, focusing on a region of the morphology space in which performances can be increased without negatively affecting performances in other parts of the space.

Ant. Upon analyzing the frequency of morphology selection during training on Ant, we observed an imbalance in the utilization of morphologies also in this case, particularly skewed towards the left-central region of morphology space (heatmap reported in the Supplementary Material). The focus of the MAB approach on this specific region of the space leads to an improvement in performance within that area. The controllers trained with the Random schedule achieved higher scores predominantly on the right side of the morphology space, whereas the networks trained with the MAB schedule slightly enhanced performance towards the left region. Nonetheless, this improvement comes at the expense of slightly degraded performance on the right side of the space. While we may not observe a significant difference in results, it is interesting to observe that even in this case, the MAB approach chooses to focus on a specific region of the morphology space rather than uniformly utilizing all available morphologies.

6 CONCLUSIONS

In this study, we examined the impact of variability on the learning and generalization of ANN-based controllers. Alongside several predefined training schedules to select morphologies to train on, we employed a reinforcement learning-based approach to the automatic selection of these morphologies formalizing the selection as

a multi-armed bandit problem. Our analysis focused on two types of variability: heterogeneity of samples and scheduling of samples. The analysis of different training schedules highlighted the importance of increasing heterogeneity for improved robustness and generalization. Notably, using a uniform distribution for training morphologies generation led to better performance on the training morphologies compared to using discrete morphologies. The Beta training schedule showed the best approach to enhance generalization to previously unseen morphologies. Furthermore, we explored how altering the scheduling of the training morphologies affects the results. In this case, the Incremental training schedule showed the best performance. The MAB approach for automatic morphology selection during training revealed a non-uniform selection strategy by focusing on certain regions, which in turn led to significant improvements relative to Random selection.

Future research can deepen our understanding of how variability influences generalization. Investigating discriminative and non-discriminative features within tasks may provide valuable insights; for instance, we found that modifying the leg length in Bipedal Walker had a greater impact on performance than the leg width, which suggests that this parameter is more discriminative. One approach could be extended to prioritize variability introduction on non-discriminative features before moving to discriminative ones, a strategy that has been shown to enhance learning [19]. Exploring automated scheduling through meta-learning approaches like (Deep) Q-learning is another avenue for investigation. Additionally, evolutionary algorithms hold the potential for optimizing the order of samples, though computational challenges must be addressed. Furthermore, future studies could compare the performance of non-stationary MAB approaches against stationary ones, shedding light on their relative effectiveness. These research directions offer promising avenues for advancing our understanding of variability and optimization in complex tasks.

REFERENCES

[1] Alex Hernández-García and Peter König. 2018. Data augmentation instead of explicit regularization. arXiv:1806.03852.

[2] Aravind Rajeswaran and Kendall Lowrey and Emanuel Todorov and Sham M. Kakade. 2017. Towards Generalization and Simplicity in Continuous Control. arXiv:1703.02660.

[3] Boyce, B. A. and others. 2006. Implications for variability of practice from pedagogy and motor learning perspectives: finding a common ground. *Quest* 58 (2006), 330–343.

[4] Carvalho, Jonata Tyska and Nolfi, Stefano. 2023. The role of morphological variation in evolutionary robotics: Maximizing performance and robustness. *Evolutionary Computation* Early access (2023), 1–18.

[5] Catto, Erin. 2011. Box2D: A 2D physics engine for games.

[6] Charles Packer and Katelyn Gao and Jernej Kos and Philipp Krähenbühl and Vladlen Koltun and Dawn Song. 2018. Assessing Generalization in Deep Reinforcement Learning. arXiv:1810.12282.

[7] Corinna Triebold and Anil Yaman. 2024. Evolving generalist controllers to handle a wide range of morphological variations. arXiv:2309.10201.

[8] Antoine Cully, Jeff Clune, Danesh Tarapore, and Jean-Baptiste Mouret. 2015. Robots that can adapt like animals. *Nature* 521, 7553 (2015), 503–507.

[9] Daan Wierstra and Tom Schaul and Tobias Glasmachers and Yi Sun and Jürgen Schmidhuber. 2011. Natural Evolution Strategies. arXiv:1106.4487.

[10] Daniel Russo and Benjamin Van Roy and Abbas Kazerouni and Ian Osband. 2017. A Tutorial on Thompson Sampling. arXiv:1707.02038.

[11] Floreano, Dario and Dürr, Peter and Mattiussi, Claudio. 2008. Neuroevolution: from architectures to learning. *Evolutionary Intelligence* 1, 1 (2008), 47–62.

[12] Glasmachers, Tobias and Schaul, Tom. 2010. Exponential natural evolution strategies. *Evolutionary Computation* 18, 4 (2010), 553–572.

[13] Gliozzi, Valentina and Plunkett, Kim. 2019. Grounding Bayesian accounts of numerosity and variability effects in a similarity-based framework: the case of self-organising maps. *Cognitive Psychology* 31 (2019), 605–618.

[14] Greg Brockman and Vicki Cheung and Ludwig Pettersson and Jonas Schneider and John Schulman and Jie Tang and Wojciech Zaremba. 2016. OpenAI Gym. arXiv:1606.01540.

[15] Han Guo and Ramakanth Pasunuru and Mohit Bansal. 2019. AutoSeM: Automatic Task Selection and Mixing in Multi-Task Learning. arXiv:1904.04153.

[16] Hebert, E.P. and others. 1996. Practice schedule effects on the performance and learning of low- and high-skilled students: an applied study. *Research Quarterly for Exercise and Sport* 67 (1996), 52–58.

[17] Hill, Felix and others. 2019. Environmental drivers of systematicity and generalization in a situated agent. arXiv:1910.00571.

[18] Huet, Michaël and Jacobs, David and Camachon, Cyril and Missenard, Olivier and Gray, Rob and Montagne, Gilles. 2011. The Education of Attention as Explanation of Variability of Practice Effects: Learning the Final Approach Phase in a Flight Simulator. *Journal of experimental psychology. Human perception and performance* 37 (07 2011), 1841–54.

[19] Limor Raviv and Gary Lupyan and Shawn C. Green. 2022. How variability shapes learning and generalization. *Trends in Cognitive Sciences* 26, 6 (2022), 462–483.

[20] Magill, R.A. and Hall, K.G. 1990. A review of the contextual interference effect in motor skill acquisition. *Human Movement Science* 9 (1990), 241–289.

[21] Mann, H B and Whitney, D R. 1947. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics* 18, 1 (1947), 50–60.

[22] Nihat Engin Toklu and Timothy Atkinson and Vojtěch Micka and Paweł Liskowski and Rupesh Kumar Srivastava. 2023. EvoTorch: Scalable Evolutionary Computation in Python. arXiv:2302.12600.

[23] Tønnes F Nygaard, Charles P Martin, Jim Torresen, Kyrre Glette, and David Howard. 2021. Real-world embodied AI through a morphologically adaptive quadruped robot. *Nature Machine Intelligence* 3, 5 (2021), 410–419.

[24] Pagliuca, Paolo and Milano, Nicola and Nolfi, Stefano. 2020. Efficacy of modern neuro-evolutionary strategies for continuous control optimization. *Frontiers in Robotics and AI* 7 (2020), 98.

[25] Petru Soviany and Radu Tudor Ionescu and Paolo Rota and Nicu Sebe. 2021. Curriculum Learning: A Survey. arXiv:2101.10382.

[26] Posner, M.I. and Keele, S.W. 1968. On the genesis of abstract ideas. *Journal of Experimental Psychology* 77 (1968), 353–363.

[27] Putter, Ruben and Nitschke, Geoff. 2017. Evolving morphological robustness for collective robotics. In *IEEE Symposium Series on Computational Intelligence*. IEEE, New York, US, 1–8.

[28] Ravi Mangal and Aditya V. Nori and Alessandro Orso. 2019. Robustness of Neural Networks: A Probabilistic and Practical Approach. arXiv:1902.05983.

[29] Sebastian Risi and Kenneth O Stanley. 2013. Confronting the challenge of learning a flexible neural controller for a diversity of morphologies. In *Proceedings of the 15th annual conference on Genetic and evolutionary computation*. ACM, New York, NY, US, 255–262.

[30] Erica Salvato, Gianfranco Fenu, Eric Medvet, and Felice Andrea Pellegrino. 2021. Crossing the reality gap: A survey on sim-to-real transferability of robot controllers in reinforcement learning. *IEEE Access* 9 (2021), 153171–153187.

[31] Schmidt, Richard. 1997. Variability of Practice and Implicit Motor Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23 (07 1997), 987–1006.

[32] Stanley, Kenneth and Clune, Jeff and Lehman, Joel and Miikkulainen, Risto. 2019. Designing neural networks through neuroevolution. *Nature Machine Intelligence* 1, 1 (01 2019), 24–35.

[33] Sébastien Bubeck and Nicolò Cesa-Bianchi. 2012. Regret Analysis of Stochastic and Nonstochastic Multi-armed Bandit Problems. arXiv:1204.5721.

[34] Taleb, Nassim Nicholas. 2013. *Antifragile*. Random House, New York.

[35] Tim Salimans and Jonathan Ho and Xi Chen and Szymon Sidor and Ilya Sutskever. 2017. Evolution Strategies as a Scalable Alternative to Reinforcement Learning. arXiv:1703.03864.

[36] Todorov, Emanuel and Erez, Tom and Tassa, Yuval. 2012. MuJoCo: A physics engine for model-based control. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, New York, NY, US, 5026–5033.

[37] Vishnu Raj and Sheetal Kalyani. 2017. Taming Non-stationary Bandits: A Bayesian Approach. arXiv:1707.09727.

[38] Anil Yaman, Giovanni Iacca, Decebal Constantin Mocanu, Matt Coler, George Fletcher, and Mykola Pechenizkiy. 2021. Evolving plasticity for autonomous learning under changing environmental conditions. *Evolutionary computation* 29, 3 (2021), 391–414.

[39] Zahra Hussain and Patrick J. Bennett and Allison B. Sekuler. 2012. Versatile perceptual learning of textures after variable exposures. *Vision Research* 61 (2012), 89–94. Perceptual Learning - mechanisms and manifestations.

The Effect of Training Schedules in Morphological Generalization (Supplementary Material)

EDOARDO BARBA, University of Trento, Italy
ANIL YAMAN, Vrije Universiteit Amsterdam, Netherlands
GIOVANNI IACCA, University of Trento, Italy

A   DATASET VISUALIZATION

A.1   Training and testing sets

Figure 1 presents a visual depiction of the training and testing sets for the Bipedal Walker task. The dots representing the training set are depicted with a red background, while those representing the testing set have a blue background. Each dot on the plot represents a specific morphology characterized by a combination of leg width and leg length parameters.

To assess the performance of an individual on the training set, its score is computed for each morphology within the training set. These individual scores are then averaged to derive a single metric representing the individual's performance on the training set. The same methodology is applied to compute the individual's score on the testing set. This systematic approach facilitates a thorough evaluation of the individual's ability to generalize across a diverse range of morphologies present in both the training and testing sets.
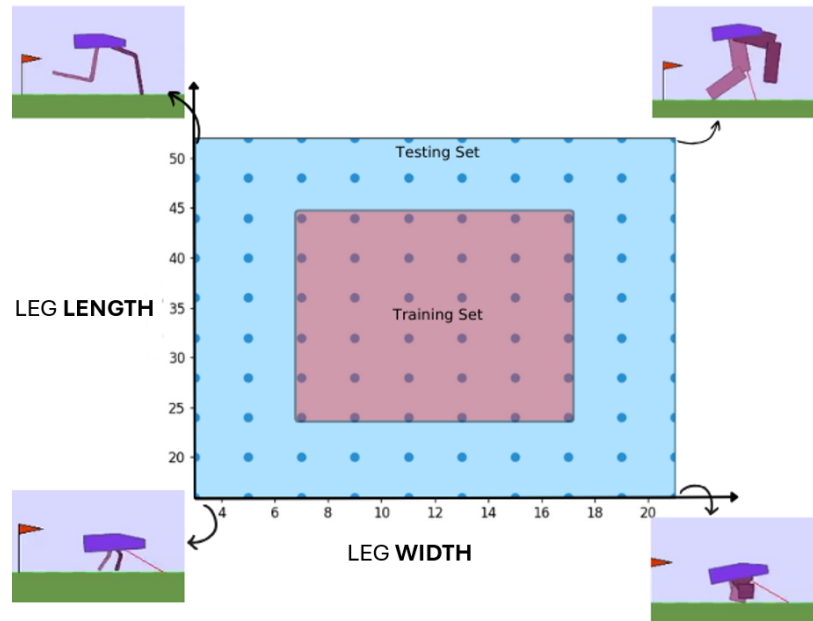
Fig. 1. Visual representation of the Training and Testing sets for the Bipedal Walker task. Each point represents a specific morphology characterized by a combination of leg width and leg length parameters. The training set consists of 36 morphologies, while the testing set consists of 64 morphologies.

A.2 Validation sets

Figures 2a-2c illustrate the validation sets utilized during the training phase for the respective tasks. Each point in these plots represents a morphology characterized by specific parameter combinations. During training, the validation set serves to evaluate the generalization ability of the top-performing individual on the morphology presented in that generation. This process entails computing the score achieved by the individual on each of these 36 morphologies separately and then averaging these scores to obtain a single metric.

(a) Bipedal Walker   (b) Walker2D   (c) Ant
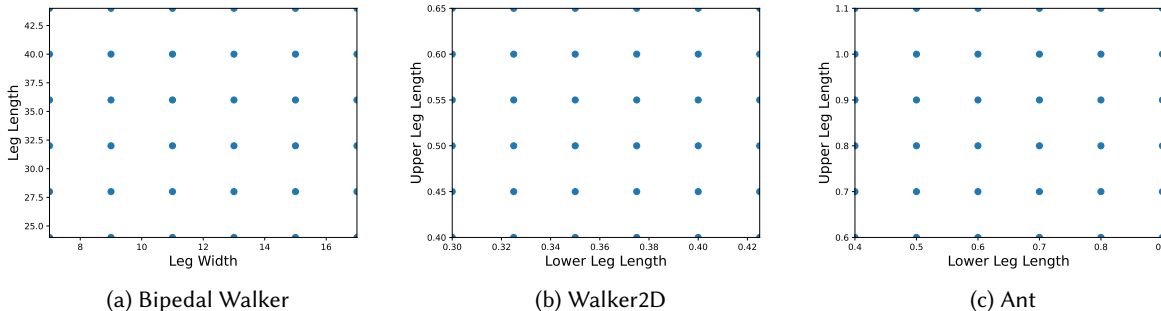
Fig. 2. Validation sets on the three tasks.

B DISTRIBUTION VISUALIZATION

Figures 3a-3d provide visual representations of the training sets utilized on the Ant task by the Uniform, Gaussian, Beta, and Cauchy training schedules, respectively. In these plots, each point signifies a morphology, with parameters sampled independently from the given distribution. The background colors serve to reflect the density of the samples, providing insights into the distribution of morphologies across the parameter space.

C PERFORMANCE ANALYSIS

Figures 4a and 4b present boxplots representing the performance of artificial neural networks (ANNs) trained using various training schedules on Bipedal Walker, respectively on the training and testing sets.

C.1 Scheduling

Figures 5a and 5b show the performance of ANNs trained using Random schedule and Incremental schedule on the Ant task, respectively on the training and testing sets. The network trained with the Random schedule achieved a median performance of -3779.36 on the training set and -2139.04 on the testing set. Conversely, the network trained with the Incremental schedule reached -3590.29 and -2375.35 on the respective sets.

D MULTI-ARMED BANDIT SCHEDULE

D.1 Performance comparison

Figures 6a and 6b illustrate a performance comparison between the Random and Multi-Armed Bandit (MAB) schedules on the testing set, specifically for the Bipedal Walker and Ant tasks, respectively. Figure 7 and 8 presents, respectively for Bipedal Walker and Ant, heatmaps illustrating the average performances of ANNs (across all the morphologies) trained using two different scheduling strategies: Random and MAB. Each cell in the heatmaps corresponds to a specific morphology. This visualization incorporates morphologies from both the

(a) Uniform distribution

(b) Gaussian distribution

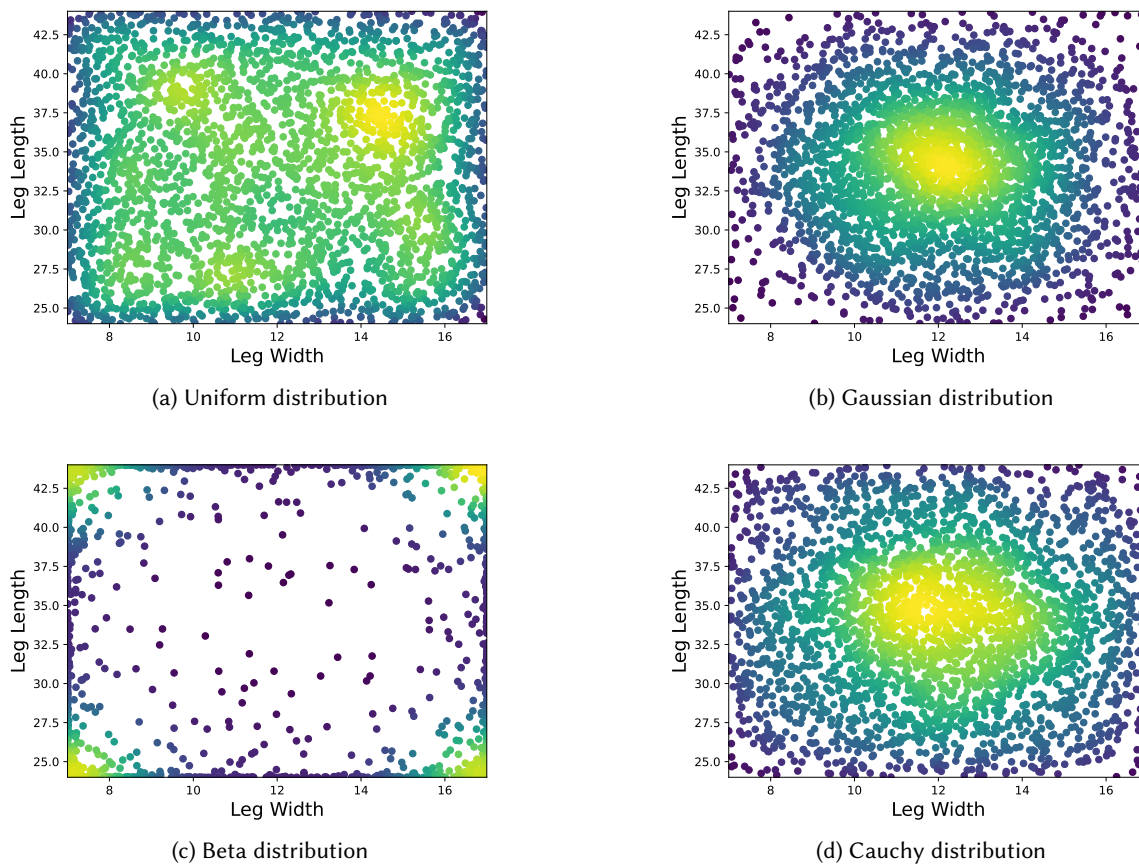(c) Beta distribution

(d) Cauchy distribution

Fig. 3. Distribution on Ant.

training (inside the red square) and testing sets. The shading of each cell indicates the average performance of the ANN controller on that particular morphology, with darker shades indicating better scores. These averages are computed over all controllers obtained from 30 runs.

D.2 Morphology frequency

Figures 9a and 9b display heatmaps showing the frequency of each morphology's selection during training for the Bipedal Walker and Ant tasks, respectively. Each cell in the heatmap represents a unique morphology, and the numerical value within the cell indicates the frequency with which that morphology was selected during training. This frequency is computed as the average occurrence across 30 independent training runs.
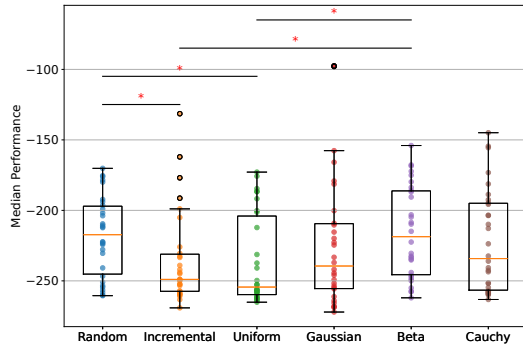
(a) Training set
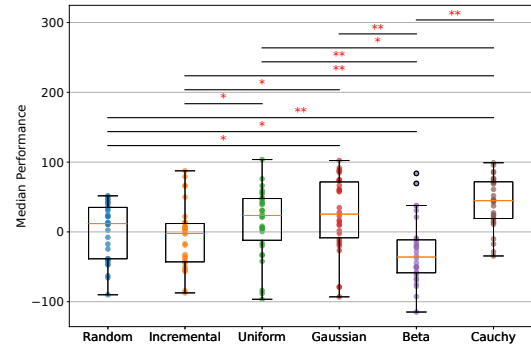
(b) Testing set

Fig. 4. Performance of all the training schedules on Bipedal Walker. * and ** indicate statistical significance with $p < 0.05$ and $p < 0.01$, respectively.
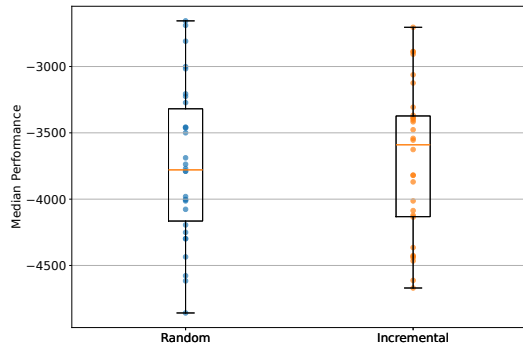
(a) Training set
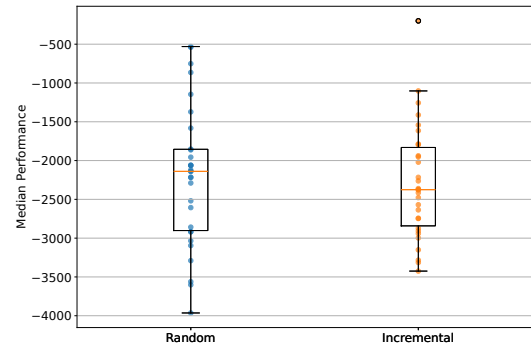
(b) Testing set

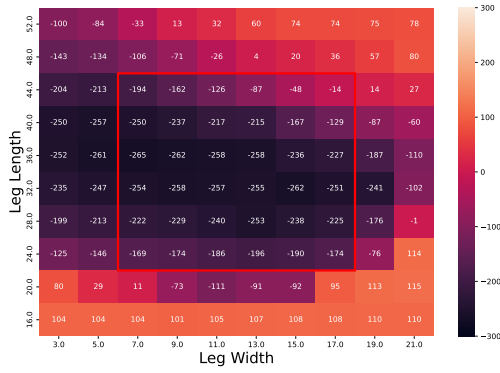Fig. 5. Performance comparison between Random and Incremental schedules on Ant.
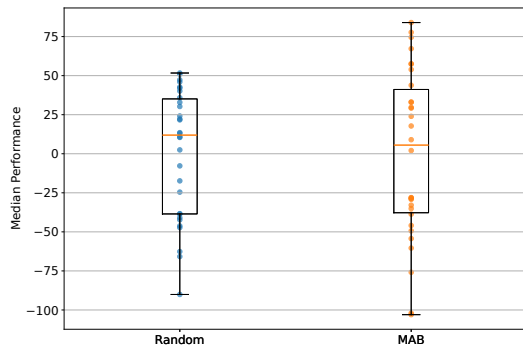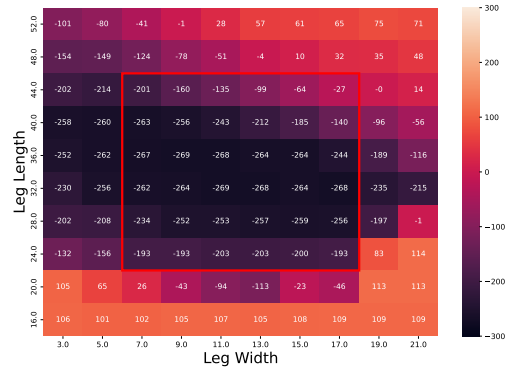

(a) Bipedal Walker
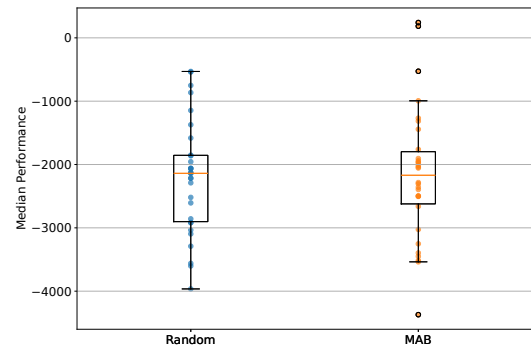
(b) Ant

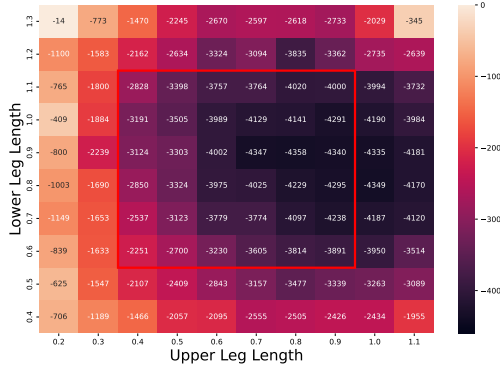Fig. 6. Performance comparison between Random and Multi-Armed Bandit on the test morphologies.
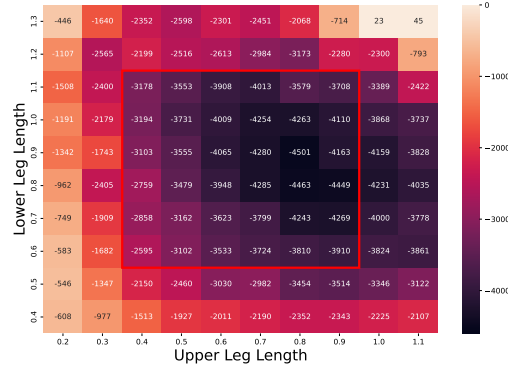
4

(a) Random

(b) Multi-Armed Bandit

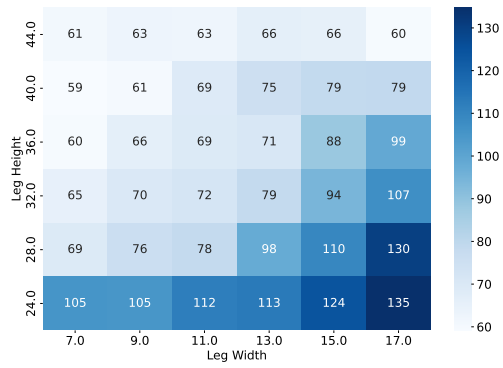Fig. 7. Average performance across morphologies for Bipedal Walker.
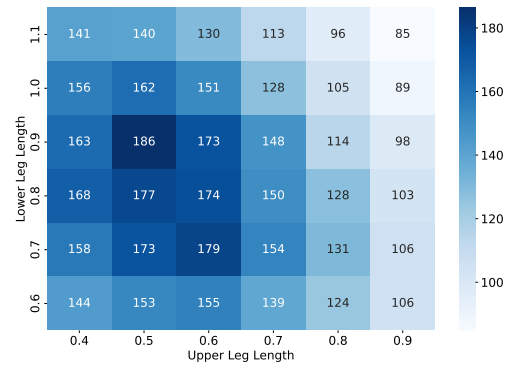

(a) Random

(b) Multi-Armed Bandit

Fig. 8. Average performance across morphologies for Ant.

(a) Bipedal Walker   (b) Ant

Fig. 9. Frequency of each morphology's selection during training.