Moderate Environmental Variation Across Generations Promotes the Evolution of Robust Solutions

Nicola Milano*
Institute of Cognitive Sciences and Technologies
National Research Council
nicola.milano@istc.cnr.it

Jônata Tyska Carvalho
Institute of Cognitive Sciences and Technologies
National Research Council
jonata.carvalho@istc.cnr.it

Federal University of Rio Grande Center for Computational Sciences

Stefano Nolfi
Institute of Cognitive Sciences and Technologies
National Research Council

Keywords
Environmental variations, evolvability, stability, artificial evolution

Abstract
Previous evolutionary studies demonstrated how robust solutions can be obtained by evaluating agents multiple times in variable environmental conditions. Here we demonstrate how agents evolved in environments that vary across generations outperform agents evolved in environments that remain fixed. Moreover, we demonstrate that best performance is obtained when the environment varies at a moderate rate across generations, that is, when the environment does not vary every generation but every $N$ generations. The advantage of exposing evolving agents to environments that vary across generations at a moderate rate is due, at least in part, to the fact that this condition maximizes the retention of changes that alter the behavior of the agents, which in turn facilitates the discovery of better solutions. Finally, we demonstrate that moderate environmental variations are advantageous also from an evolutionary computation perspective, that is, from the perspective of maximizing the performance that can be achieved within a limited computational budget.

1 Introduction

The last two decades have seen increasing recognition of the role of environmental variations in evolution.

The interaction between environmental conditions and the expression of genetic variation influences the evolutionary dynamics. Genes influencing a trait in one environment may not be important in a different one [30]. Mutations often have environment-dependent effects [13, 28]. The environmental conditions influence the genetic interactions among traits, that is, the correlation between the genetic influences on a trait and the genetic influences of another trait, which are known to influence the evolutionary dynamics [27]. For instance, the genetic correlations among certain traits can be positive in one environment and negative in another. Consequently, environmental variations influence evolutionary trajectories in populations [27].

Moreover, as stressed by [31], phenotypic variation arises not only as a result of genetic variations but also as a result of environmental variations. “Environmentally induced phenotypic changes can give rise to adaptive evolution as readily as mutational induced changes; both are equally subject to genetic accommodation” [31, p. 498].

* Corresponding author.

In this article we analyze the impact of environmental variations on the evolution of neuro-controlled agents situated in an external environment and the conditions that promote the evolution of agents that are robust with respect to environmental variations.

Previous research demonstrated how robust solutions can be obtained by evaluating evolving agents multiple times in varying environmental conditions and by selecting individuals on the basis of the average or worst fitness. This method has been applied to evolve electronic circuits robust to temperature variations [29], fault-tolerant neural networks [26], job shop scheduling systems [25], flight control systems operating under changing conditions [1], and robots operating in varying environmental conditions [2, 9, 21, 23].

In this work we analyze whether the occurrence of environmental variations across generations promotes the evolution of robust solutions. The obtained results demonstrated that indeed agents evolved in environmental conditions that vary across generations outperform agents evolved in environments that remain stable. By the term performance we mean the ability of individuals to solve their adaptive problem in all possible environmental conditions. Performance can be estimated on the basis of the averaged fitness obtained by the individuals during multiple post-evaluation tests carried out in randomly different environmental conditions. The precision with which the fitness and/or the performance is estimated depends on the number of evaluations. Indeed, the higher the number of evaluations is, the higher the precision of the estimation is. Moreover, our results indicate that the performance is maximized when the environment varies across generations at a moderate rate (i.e., when the environment does not vary every generation, but only every $N$ generations).

As we will see, the advantage of exposing evolving agents to environments that vary across generations at a moderate rate is due, at least in part, to the fact that this condition maximizes the retention of changes that alter the behavior of the agents, which in turn facilitates the discovery of better solutions.

This study is related to evolutionary dynamic optimization [4, 10, 11, 12, 18, 22], namely, the study of how an optimization algorithm can solve an optimization problem in a given period $[t_{\text{begin}}$, $t_{\text{end}}]$ in which the underlying fitness landscape changes and in which the optimization algorithm reacts to that change by providing new optimal solutions (adapted from Definition 1 in [18]). However, the objectives of robust and dynamics optimization differ in that the former aims to develop solutions capable of operating as effectively as possible in new environmental conditions without further adaptation, and the latter, instead, aims to develop solutions that are not necessarily effective after the environment changes but that can adapt to the new environmental conditions as readily as possible. Consequently, the methodological issues also differ. For example, the formalization of a method for measuring the speed with which agents adapt to the new environmental conditions is crucial for studying dynamic optimization [4, 18], but is not relevant for the study of robust agents.

Moreover, this research is related to the evolution of plasticity, that is, the ability of agents to react to internal or external environmental states with a change in form, state, movement, or rate of activity [31]. Indeed, the agents described in this article have the ability to detect the current environmental conditions and to vary their behavior in a context-dependent manner. For an analysis on the role of plasticity in artificial evolving agents see [3, 5, 20]. In this article, however, we focus on the impact that environmental variation has on the course of the evolutionary process and on the effect of the rate at which the environment varies.

2 Method

To investigate the evolution of robust solutions we evolved a population of neuro-controlled agents for the ability to balance two poles attached to their tops through passive joints in varying environmental conditions (see Figure 1). The characteristics of the environment that vary are the inclination and the friction of the plane and the initial state of the cart and the poles. Consequently, the evolving agents should display an ability to balance the poles irrespective of the characteristics of the plane and of the initial state of the cart and the poles.
This problem constitutes an extended version of the non-Markovian version of the double-pole balancing problem introduced by [33] that has become a commonly recognized benchmark for nonlinear control [8, 14]. In the standard version of the problem the plane is always parallel to the ground and the friction between the plane and the cart is always null. For more details, see Appendix A.1.

Agents are provided with a three-layer neural network with five sensory neurons, ten internal neurons with recurrent connections, and one motor neuron (Figure 1, right). The sensory neurons encode the position of the cart ($x$), the angular position of the two poles ($\theta_1$ and $\theta_2$), the inclination of the plane ($\alpha$), and the friction coefficient of the plane ($\mu$). The activation state of the motor neuron is normalized in the range $[-10.0, 10.0]$ N and is used to set the force ($F$) applied to the cart along the $x$ axis. For more details, see Appendix A.2.

The connection weights of the neural network, which determine the behavior of the agents, are encoded in artificial genotypes and evolved through a steady state evolutionary algorithm, a method widely used to evolve embodied agents [19, 22]. For more details see Appendix A.3.

Each agent is evaluated for NT trials that vary with respect to the characteristics of the plane and with respect to the initial state of the cart and the poles. More specifically, at the beginning of each trial the inclination of the plane ($\alpha$), the friction coefficient between cart and plane ($\mu$), the initial position of the plane on the plane ($\phi$), the velocity of the cart ($\dot{x}$), the angular positions of the two poles ($\theta_1$ and $\theta_2$), and the angular velocities of the two poles ($\dot{\theta}_1$ and $\dot{\theta}_2$) are set to values selected randomly with a uniform distribution within the following ranges: $[0.0, 0.2617]$, $[0.0, 0.30]$, $[-1.5, 1.5]$, $[-1.2, 1.2]$, $[-0.1047, 0.1047]$, $[-0.1350, 0.1350]$.

Trials terminated after 1,000 steps or when the angular position of one of the two poles exceeded the range $[-0.628319, 0.628319]$ rad and/or the position of the cart exceeded the range $[-2.4, 2.4]$ m.

The fitness of the agent during a trial is given by the fraction of time steps in which the agent maintains the cart and the poles within the allowed position and orientation ranges. The total fitness is obtained by averaging the fitnesses obtained in the NT trials. The higher NT is, the higher is the accuracy of the estimation of the agent’s performance in varying environmental conditions.

Figure 1. Left: Schematization of the extended double-pole balancing problem. See text for explanation. Right: The architecture of the neural network controller. The circles shown in the bottom, middle, and top parts of the figure represent the sensory, internal, and motor neurons, respectively. Red circles represent the biases. The arrows represent connections. For the sake of clarity, only the connections departing from the first sensory and the first internal neuron are displayed.
The performance of evolved agents, that is, the ability of an evolved agent to solve the problem in variable environmental conditions, is measured by post-evaluating the agents for 1,000 trials in which the characteristics of the environment and the initial state of the cart are set randomly with a uniform distribution in one of the ranges described above. Although the number of different environmental conditions that an agent can encounter is practically infinite, measuring the performance of the agents on 1,000 randomly different environmental conditions provides a good estimate of their performance, that is, of their ability to solve the problem in all possible environmental conditions.

The experiments have been replicated in a fixed condition in which the environmental conditions do not vary across generations, in a varying condition in which the environmental conditions vary every generation, and in an intermediate condition in which the environmental conditions vary every \( N \) generations. The environmental conditions experienced by the agents are stored in an \( N \times T \times 8 \) matrix that encodes for each trial the inclination and the friction coefficient of the plane, the initial position and velocity of the cart, and the initial position and velocity of the poles. The values of the matrix are generated randomly with a uniform distribution in the range described above. In the fixed condition the \( N \times T \times 8 \) matrix is generated randomly once and is maintained constant during the entire evolutionary process. In the varying condition the matrix is regenerated randomly every generation. Finally, in the intermediate condition, the matrix is regenerated randomly every \( N \) generations.


3 Moderate Environmental Variation across Generations Promotes the Evolution of Better Solutions

Figure 2 displays the results of a series of experiments in which the environmental conditions either remained fixed or varied every 5,000, 2,500, 1,000, 500, 200, 100, 50, 25, 10, or 1 generation(s). In all experiments the evolutionary process was continued for 50,000 generations. Evolving agents were evaluated for 50 trials.

As can be seen, the best performance is achieved by agents evolved in environments that vary at a moderate rate across generations. Indeed, the agents evolved in the experiments in which the environment changes every 100 generations achieve significantly better performance than the agents evolved in the experiments in which the environment remains fixed or changes every generation (Kruskal-Wallis, \( p \)-value < 0.001; Wilcoxon test, \( p \)-value < 0.05 with Bonferroni correction in all cases; effect size: Cohen’s \( d > 0.8 \) in all cases).

The agents evolved in environments that vary across generations outperform the agents evolved in nonvarying environments. Indeed, the performance obtained in the fixed condition is significantly lower than the performance achieved in all other conditions (Kruskal-Wallis, \( p \)-value < 0.001; Wilcoxon test, \( p \)-value < 0.05 with Bonferroni correction in all cases; effect size: Cohen’s \( d > 0.8 \) in all cases).

In the rest of the article, we will use the term intermediate condition to refer to the experiments in which the environment varies every 100 generations, since that corresponds approximately to the average value for the best conditions.

The beneficial effect of moderate environmental variations across generations persists in the long term, as demonstrated by the fact that the agents evolved in the intermediate condition outperform agents evolved in fixed or in always-varying conditions even after 100,000 generations (Figure 3, Kruskal-Wallis, \( p \)-value < 0.001; Wilcoxon test, \( p \)-value < 0.05 with Bonferroni correction; effect size > 0.8).

Some of these data were previously reported in [17].

In the following two sections we analyze why moderate environmental variations promote the evolution of better agents. In Section 6 we analyze whether moderate environmental variations are
also advantageous from an evolutionary computational perspective, that is, from the point of view of optimizing the performance that can be achieved by using a limited computational budget. Finally, in Section 7 we draw our conclusions.

4 On the Role of Fortunate Specific Environmental Conditions

A hypothesis that could explain why agents evolved in varying environmental conditions achieve better performance than agents evolved in fixed conditions is that environmental variations enable evolving agents to encounter, sooner or later, favorable environmental conditions that promote the evolution of better solutions.

This hypothesis is based on evidence collected in incremental evolutionary experiments in which the problem and/or the environment become progressively more challenging throughout generations [16, 20]. The possibility of experiencing easier conditions during the first evolutionary phase leads to the synthesis of solutions for simple conditions that represent a stepping-stone toward the
synthesis of solutions for harder conditions. Moreover, the hypothesis is based on evidence, collected in the neural network learning literature, that indicates that the relative distribution of qualitatively different stimuli in the training set affects the outcome of the learning process [7, 34]. Environmental variations in our experiments are stochastic and consequently cannot lead to a progressive complexification of the adaptive problem. On the other hand, one could hypothesize that continuous variation of the environmental conditions can generate, sooner or later, sufficiently easy environmental conditions or favorable environmental conditions that can boost evolution.

To verify this hypothesis, we analyzed the impact of different environmental conditions on the evolution of a population of evolving agents. The analysis was conducted on 30 populations of agents evolved in the intermediate experimental condition in which the environment varied every 100 generations after one, five, and then ten thousand generations. One hundred copies of these populations were evolved for 500 generations in 100 different corresponding environments for 50 trials. The performance of these populations was post-evaluated every 100 generations on 729 trials in which agents were exposed to systematically varied environmental conditions. Agents were evaluated...
for $3^6 = 729$ trials, during which the state of the six variables that encode the characteristics of the plane and the most important initial characteristics of the cart assumed one of the following values: $\eta [-0.1385, 0.0, 0.1385]$, $\mu [-0.15, 0.0, 0.15]$, $\chi [-0.75, 0.0, 0.75]$, $\delta [-0.6, 0.0, 0.6]$, $\theta_1 [-0.05235, 0.0, 0.05235]$, $\theta_2 [-0.0675, 0.0, 0.0675]$. The angular position and velocity of the shorter pole ($\theta_2$ and $\theta_2'$), which are easier to control, were always set to 0.0.

As can be seen, the performance of the agents after 1,000 and 5,000 generations increases by a similar amount in all environmental conditions (Figure 4, top and middle). Later on (viz., after 10,000 generations), performance variations become much smaller and lead both to increase and to decrease of performance (Figure 4 bottom; notice that the scale used for displaying variation in performance is one order of magnitude smaller in the case of the bottom picture).

Overall, these data do not show evidence of fortunate specific environmental conditions capable of boosting the evolutionary process. During the initial evolutionary phase, the evolving agents improve their ability by a similar amount in the large majority of environmental conditions. Later on, the evolving agents improve their ability by a similar small amount in the majority of the environmental conditions and worsen their performance by a similar small amount in the remaining conditions.

5 Environmental Variation Increases the Rate at Which Evolving Agents Change Across Generations

Another possible reason that could explain why moderate environmental variation promotes the evolution of better solutions is that moderate environmental variation promotes evolutionary change.

The rationale behind this hypothesis is that adaptation depends on the generation of phenotypic changes and on the retention of the changes that are adaptive. As stressed by [31], phenotypic variations arise not only as a result of genetic variations, but also as a result of environmental variations. Indeed, adaptations can arise both as a result of genetically induced changes and as a result of environmentally induced changes, since both are subjected to genetic accommodation [31]. The sum of environmentally induced changes and genetically induced changes, therefore, produce more variation than genetically induced changes alone and can thus facilitate the discovery of better solutions. The fact that moderate environmental variation promotes the evolution of better solutions than frequent environmental variation is due to the fact that it represents an optimal tradeoff between the contrasting needs of variation and stability. Indeed, too infrequent environmental variation provides only limited opportunity for change. On the other hand, too frequent environmental variation prevents the stability that is necessary to enable genetic accommodation. This hypothesis is in line with West-Eberhard’s claim that the most important reason that explains why environmentally induced changes are evolutionarily important is indeed their time persistence:

Perhaps the most compelling argument for the superiority of environmental induction over mutations in terms of recurrence and persistence has to do with the inexorable persistence of an environment immune to natural selection: environmental inducers might be not only immediately widespread without necessity for positive effects on fitness sufficient to spread them due to differential reproduction of their bearers (selection), but they are inexorably present. [31, p. 504]

This hypothesis suggests that agents evolved in environmental conditions that vary at a moderate rate can accumulate more phenotypic variations across generations than agents evolved in nonvarying environments and agents evolved in environments that vary every generation.

To verify how the rate of variation of the environment affects the rate at which the agents vary across generations, at the behavioral level, we compared, every 100 generations, the best evolved agent and its ancestor of 100 generations before. The comparison was made by post-evaluating the agents for 729 trials, during which they were exposed to systematically varied environmental
Figure 4. Fraction of additional trials solved every 100 generations by populations of agents evolved for 500 generations in 100 different environments. The top, middle, and bottom pictures show the results obtained by analyzing the agents evolved in the intermediate condition after 1,000, 5,000, and 10,000 generations, respectively. Data averaged over 10 replications of the experiment. Columns correspond to data after 100, 200, 300, 400, and 500 generations. Lines indicate performance variations observed in each of the 100 different environments.
conditions as described above, and by counting the number of trials in which the agent manages to balance the poles while its ancestor fails or vice versa. The evolutionary process was continued for 5,000 generations in all cases. As in the case of the analysis reported in Figure 2, NT was set to 50, the mutation rate to 1%, and the stochasticity level to 0%.

As expected, the rate of variation decreases across generations as a result of the evolution of better and better agents (Figure 5). Interestingly, the behavior of agents that were evolved in the intermediate experimental condition, in which the environment varies every 100 generations (blue line), varies more across generations than the behavior of agents that were evolved in the nonvarying environment (black line) and in the always varying environment (red line) (Figure 5, Wilcoxon test, \( p \)-value < 0.05 with Bonferroni correction, effect size > 0.8). The rate of variation of the behavior of the agents evolved in the fixed and always varying experimental conditions, instead, does not differ statistically (Figure 5, Wilcoxon test, \( p \)-value > 0.5 with Bonferroni correction).

To verify how the rate of variation of the environment affects the rate at which the agents vary across generations, at the genetic level, we compared the fraction of genes that differ between the best evolved agent and its ancestor of 500 generations before, every 500 generations. Also, in this case the analysis was performed on experiments continued for 5,000 generations. The agents evolved in the intermediate condition accumulate more genetic variations across generations than the agents evolved in the always varying and fixed experimental conditions (Figure 6, Wilcoxon test, \( p \)-value...
0.05 with Bonferroni correction in both cases, effect size = 0.816). The agents evolved in the always varying condition accumulate more changes than the agents evolved in the fixed condition (Figure 6, Wilcoxon test, p-value < 0.05 with Bonferroni correction, effect size = 0.744).

Clearly, the rate at which agents change throughout generations at the genetic level is influenced by the mutation rate, which, however, is always 1% in the experiments reported in this section.

The analysis of these data also indicates a significant correlation between the performance of evolving agents and the rate with which agents vary throughout generations at the behavioral (Spearman, \( r = 0.52, p\)-value < \(10^{-7}\)) and genetic level (Spearman, \( r = 0.53, p\)-value < \(10^{-8}\)).

Overall these results indicate that moderate environmental variation promotes the retention of genetic variations that alter the behavior of the evolving agents, which in turn favors the discovery of better solutions.

6 On the Advantage of Moderate Environmental Variations from an Evolutionary Computation Perspective

In this section we analyze whether moderate environmental variations are advantageous also from an evolutionary computation perspective, that is, from the point of view of maximizing the performance that can be achieved with a limited computational budget.
Since the major computation cost in these experiments is that of the evaluation of candidate solutions, in this section we compare experiments in which the evolutionary process is continued until a maximum total number of evaluations is performed.

To identify the best value of the parameters, we ran three series of experiments in the fixed, always varying, and intermediate experimental conditions in which we systematically varied the number of trials experienced by evolving agents (NT), the mutation rate, and the stochasticity level. The total number of evaluations was set to 50 million in all cases. The analysis of these data indicates that the best performance is achieved by setting NT to 200, 25, and 25 in the fixed, always variable, and intermediate conditions. Performance with the best parameter differs statistically in all cases: fixed condition (Kruskal-Wallis, \(p\)-value < 0.001; Wilcoxon test, \(p\)-value < 0.05 with Bonferroni correction; effect size > 0.7); always variable condition (Appendix A.4, Table 2: Kruskal-Wallis, \(p\)-value < 0.001; Wilcoxon test, \(p\)-value < 0.05 with Bonferroni correction; effect size > 0.7); intermediate condition (Appendix A.4, Table 3, Kruskal-Wallis, \(p\)-value < 0.001; Wilcoxon test, \(p\)-value < 0.05 with Bonferroni correction; effect size > 0.7). Interestingly, therefore, the value of NT that leads to the best performance is relatively small in the varying and intermediate experimental conditions. This implies that synthesizing solutions capable of operating effectively in widely varied environmental conditions

![Figure 7](http://direct.mit.edu/artl/article-pdf/24/4/277/1667178/artl_a_00274.pdf)
does not necessarily cause an explosion of the computational cost required to evaluate candidate solutions.

The best values for the mutation rate are 1% in all conditions (Appendix A.4, Tables 1–3). The best value of the stochasticity parameter is 30%, 0%, and 20% in the fixed, intermediate, and variable experimental conditions, respectively (Appendix A.4, Tables 1–3).

The comparison of the results obtained in the three conditions by using the best parameters for each condition shows that the agents evolved in the moderate condition outperform the agents evolved in the other two conditions both after 50 and 100 million evaluations (Figure 7, Kruskal-Wallis, \( p \)-value < 0.001; Wilcoxon test, \( p \)-value < 0.05 with Bonferroni correction; effect size > 0.7).

The agents evolved in the moderate environmental condition outperform the agents evolved in the fixed environmental conditions independently of the duration of the evolutionary process. Indeed, the agents evolved in the intermediate condition with the best parameters for 50 million evaluations outperform the agents evolved in the fixed environmental condition with the best parameters for 400 million evaluations (Figure 7 right: Kruskal-Wallis, \( p \)-value < 0.001; Wilcoxon test, \( p \)-value < 0.05 with Bonferroni correction; effect size = 0.850).

7 Conclusions

Previous research reviewed in the introduction section demonstrated how robust solutions can be obtained by evaluating evolving agents multiple times in variable environmental conditions. Our results demonstrate how the occurrence of variations in the environmental conditions across generations leads to better solutions. Moreover, our results demonstrate that performance is maximized when the environment varies at a moderate rate across generations, that is, when the environment does not vary every generation but every \( N \) generations.

These results have been obtained on a standard problem commonly used as a benchmark in artificial evolution. In principle we can expect that these results will generalize to all problem domains that require solutions robust to environmental variations. The generality of the effect, however, should be verified in future studies conducted on different problem domains.

Lineages of agents evolved in environments that vary across generations change more, at the behavioral level, than lineages of agents evolved in nonvarying environments. Moreover, lineages of agents evolved in environments that vary across generations at a moderate rate change more, at the behavioral level, than lineages of agents evolved in nonvarying or always varying environments. Consequently, the advantage provided by moderate environmental variation is due, at least in part, to the fact that it promotes evolutionary change. The fact that the amount of evolutionary change is maximized at an intermediate rate of environmental variation is due to the fact that it represents an optimal tradeoff between the contrasting needs of variation and stability. These observations confirm the importance of environmentally induced changes in natural evolution stressed by [31]. Moreover, these observations confirm the hypothesis that a key reason for the importance of environmentally induced change is its persistence over time [31].

Finally, we demonstrate that moderate environmental variations are advantageous also from an evolutionary computation perspective, that is, from the perspective of maximizing the performance that can be achieved within a limited computational budget. Indeed, the experiments carried out by systematically varying the number of trials during which evolving agents are evaluated (NT) and the rate at which the environment changes across generations while maintaining the total number of evaluations constant indicate that performance is maximized when the environment varies at a moderate rate and the evolving agents are evaluated for a small number of trials. Experiencing variable environmental conditions across generations permits one to select solutions that are robust with respect to environmental variations while minimizing the computational cost required to evaluate each candidate solution and maximizing the number of generations.
Acknowledgment
Work was partially funded by CAPES through the Brazilian program Science Without Borders.

References


**Appendix**

### A.1 The Agent and the Environment

The cart has a mass of 1 kg. The long pole and the short pole have masses of 1.0 and 0.1 kg and lengths of 0.5 and 0.05 m, respectively. The cart can move along one dimension within a track of length 4.8 m. It is provided with five sensors that encode the current position of the cart on the track (\(x\)), the current angles of the two poles (\(\theta_L\) and \(\theta_S\)) with respect to the cart, the angle of the inclined plane (\(a\)), and the friction coefficient (\(\mu\)). The motor controls the force (\(F\)) applied to the cart along the \(x\) axis. The goal of the agent is to move so as to maintain the angles of the poles and the position of the cart within a viable range (see below).

The behavior of the agent has been simulated on the basis of Equations 1–5 below. This is an extended version of the equations proposed by [6] for the standard problem, in which the inclination of the plane and the friction between the cart and the plane were not considered:

\[
\ddot{x} = \frac{F + \mu_c \ddot{M}g + M_c g \sin a + \sum_{i=1}^{n} \ddot{F}_i}{M_c + \sum_{i=1}^{n} m_i} \tag{1}
\]

\[
\ddot{\theta}_i = -\frac{3}{4 \dot{h}_i} \left( \ddot{x} \cos \theta_i - g \sin \theta_i + \frac{\mu_c \dot{\theta}_i}{m_i h_i} \right) \tag{2}
\]
\[ F_i = \mu \left[ \frac{3}{4} m_i g \sin^2 \theta_i - \frac{3 \mu_i l_i}{4} \sin \theta_i \dot{\theta}_i + m_i l_i \dot{\theta}_i^2 \cos \theta_i \right] - \frac{3}{4} \left[ m_i g \sin \theta_i \cos \theta_i + \mu \dot{\theta}_i \cos \theta_i \right] \]

\[ \bar{m}_i = \frac{3}{4} \left[ \cos^2 \theta_i - \mu_i \cos \theta_i \sin \theta_i \right] \]

\[ \bar{M} = M_i \cos \alpha + \sum_{i=1}^{n} m_i \]

where \( n \) is the number of poles on the cart, \( g \) is the acceleration due to gravity, \( m_i \) and \( l_i \) are the mass and the half length of the \( i \)th pole, \( M_i \) is the mass of the cart, \( \mu_i \) is the coefficient of friction of the cart on the track, \( \mu_p \) is the coefficient of friction for the \( i \)th hinge, \( F \) is the force applied to the cart, \( \bar{F}_i \) is the effective force from the \( i \)th pole on the cart, \( \bar{m}_i \) is the effective mass of the \( i \)th pole, and \( \bar{M} \) is the effective mass of the cart.

**A.2 The Neural Network Controller of the Agent**

The controller of the agent is constituted by a neural network with five sensory neurons, ten internal neurons with recurrent connections, and one motor neuron. The sensory neurons are fully connected with the internal neurons, and the internal neurons are fully connected with the motor neurons and the internal neurons.

The sensory neurons encode the position of the cart (\( x \)) expressed in meters, the angular position of the two poles (\( \theta_1 \) and \( \theta_2 \)) in radians, the inclination of the plane (\( \alpha \)) in radians, and the friction coefficient of the plane and cart (\( \mu_c \)). The state of all sensors is normalized in the range \([-0.5, 0.5]\]. The activation state of the motor neuron is normalized in the range \([-10.0, 10.0]\] N and is used to set the force applied to the cart. The state of the sensors, the activation of the neural network, the force applied to the cart, and the position and velocity of the cart and of the poles are updated every 0.01 s.

The neural network’s architecture is fixed. The activation state of the internal and motor neurons is updated on the basis of the logistic function. The connection weights and the biases of the network are encoded in the agent’s genome and evolved. More specifically, each genome consists of a vector of \( 171 \times 8 = 1368 \) bits that encode the 160 connection weights and the 11 biases of the corresponding neural network controller.

**A.3 The Evolutionary Method**

The evolutionary algorithm consists of a simple (\( \mu + 1 \)) evolutionary strategy [24] that operates on the basis of populations formed by \( \mu \) parents. During each generation, each parent generates one offspring, the parent and the offspring are evaluated, and the best \( \mu \) individuals are selected as new parents. When the environmental conditions do not change with respect to the previous generation, the fitness of the parent is set equal to the fitness measured during previous evaluations and the evaluation of the parents is skipped. The genome of the initial population is composed by a \( \mu \times 1368 \) matrix of bits that are initialized randomly. Each block of 8 bits is converted into a floating-point number in the range \([-5.0, 5.0]\] that is used to set the weight of the corresponding connection or bias of the neural network controller. Offspring are generated by creating a copy of the genotype of the parent and by subjecting each bit to mutation with a probability MutRate. Mutations are realized by flipping the mutated bit.
The selection pressure is regulated by adding to the fitness of individuals a value randomly selected in the range $[-\text{Noise}, \text{Noise}]$ with a uniform distribution [11], where Noise corresponds to the theoretical maximum fitness multiplied by the value of the parameter Stochasticity. When Stochasticity is set to 0.0, only the best $\mu$ individuals are allowed to reproduce. The higher the level of Stochasticity, the higher is the probability that the worse individuals reproduce.

This method requires us to set two parameters: MutRate and Stochasticity. To identify the optimal values of the parameters we carried out a series of control experiments in which the two parameters were varied systematically (see the following sections). The method operates well on small populations, for example, populations formed by 100 individuals [32].

### A.4 Performance Achieved with Systematically Varied Parameters

Tables 1, 2, and 3 report the results obtained by systematically varying the number of trials, the mutation rate, and the stochasticity level in experiments carried out in the fixed, varying, and intermediate experimental conditions in which the environment remains stable, varies every generation, and varies every 100 generations, respectively. The population size is always set to 100. The evolutionary process was continued for 50 million evaluations. Each number indicates the average results of 10 replications.
Table 1. Performance of the best agents evolved in the fixed environmental condition obtained by systematically varying the number of evaluation trials, the mutation rate, and the level of stochasticity. Each number indicates the average results of 10 replications. The evolutionary process was continued for 50 million evaluations. Data were obtained by post-evaluating evolved agents for 1,000 trials.

<table>
<thead>
<tr>
<th></th>
<th>Stoch 0%</th>
<th>Stoch 10%</th>
<th>Stoch 20%</th>
<th>Stoch 30%</th>
<th>Stoch 40%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 trial Mut 1%</td>
<td>78.9</td>
<td>85.6</td>
<td>95.3</td>
<td>86.3</td>
<td>75.3</td>
</tr>
<tr>
<td>Mut 2%</td>
<td>68.3</td>
<td>75.3</td>
<td>81.3</td>
<td>74.2</td>
<td>70.2</td>
</tr>
<tr>
<td>Mut 4%</td>
<td>55.3</td>
<td>63.2</td>
<td>53.3</td>
<td>56.8</td>
<td>55.2</td>
</tr>
<tr>
<td>50 trials Mut 1%</td>
<td>562.9</td>
<td>578.6</td>
<td>629.8</td>
<td>613.4</td>
<td>646.4</td>
</tr>
<tr>
<td>Mut 2%</td>
<td>606.9</td>
<td>628.4</td>
<td>625.7</td>
<td>645.2</td>
<td>619.6</td>
</tr>
<tr>
<td>Mut 4%</td>
<td>537.2</td>
<td>504.0</td>
<td>537.3</td>
<td>506.8</td>
<td>525.2</td>
</tr>
<tr>
<td>100 trials Mut 1%</td>
<td>625.6</td>
<td>595.9</td>
<td>622.0</td>
<td>615.9</td>
<td>634.1</td>
</tr>
<tr>
<td>Mut 2%</td>
<td>658.1</td>
<td>629.9</td>
<td>623.3</td>
<td>574.5</td>
<td>606.5</td>
</tr>
<tr>
<td>Mut 4%</td>
<td>479.2</td>
<td>442.6</td>
<td>506.1</td>
<td>477.0</td>
<td>463.0</td>
</tr>
<tr>
<td>150 trials Mut 1%</td>
<td>649.9</td>
<td>639.7</td>
<td>653.9</td>
<td>644.1</td>
<td>643.9</td>
</tr>
<tr>
<td>Mut 2%</td>
<td>601.6</td>
<td>605.3</td>
<td>624.2</td>
<td>629.8</td>
<td>602.9</td>
</tr>
<tr>
<td>Mut 4%</td>
<td>431.8</td>
<td>461.8</td>
<td>466.1</td>
<td>462.0</td>
<td>427.0</td>
</tr>
<tr>
<td>200 trials Mut 1%</td>
<td>641.6</td>
<td>644.2</td>
<td>634.0</td>
<td><strong>658.7</strong></td>
<td>636.8</td>
</tr>
<tr>
<td>Mut 2%</td>
<td>584.5</td>
<td>592.7</td>
<td>594.2</td>
<td>572.1</td>
<td>584.8</td>
</tr>
<tr>
<td>Mut 4%</td>
<td>424.5</td>
<td>442.7</td>
<td>444.2</td>
<td>432.1</td>
<td>427.8</td>
</tr>
<tr>
<td>300 trials Mut 1%</td>
<td>612.8</td>
<td>600.9</td>
<td>625.5</td>
<td>606.4</td>
<td>608.0</td>
</tr>
<tr>
<td>Mut 2%</td>
<td>571.6</td>
<td>546.9</td>
<td>551.9</td>
<td>568.7</td>
<td>539.5</td>
</tr>
<tr>
<td>Mut 4%</td>
<td>391.6</td>
<td>406.9</td>
<td>401.9</td>
<td>408.7</td>
<td>409.5</td>
</tr>
</tbody>
</table>
Table 2. Performance of the best agents evolved in the always varying environmental condition obtained by systematically varying the number of evaluation trials, the mutation rate, and the level of stochasticity. Each number indicates the average results of 10 replications. The evolutionary process was continued for 50 million evaluations. Data were obtained by post-evaluating evolved agents for 1,000 trials.

<table>
<thead>
<tr>
<th>Trials</th>
<th>Stoch 0%</th>
<th>Stoch 10%</th>
<th>Stoch 20%</th>
<th>Stoch 30%</th>
<th>Stoch 40%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mut 0.5</td>
<td>91.9</td>
<td>108.6</td>
<td>102.9</td>
<td>105.8</td>
<td>90.7</td>
</tr>
<tr>
<td>Mut 1%</td>
<td>119.9</td>
<td>130.2</td>
<td>118.9</td>
<td>111.8</td>
<td>103.7</td>
</tr>
<tr>
<td>Mut 2%</td>
<td>103.9</td>
<td>113.2</td>
<td>109.9</td>
<td>101.8</td>
<td>92.7</td>
</tr>
<tr>
<td>15 Trials</td>
<td>Stoch 0%</td>
<td>Stoch 10%</td>
<td>Stoch 20%</td>
<td>Stoch 30%</td>
<td>Stoch 40%</td>
</tr>
<tr>
<td>Mut 0.5</td>
<td>581.9</td>
<td>589.2</td>
<td>592.9</td>
<td>595.8</td>
<td>590.7</td>
</tr>
<tr>
<td>Mut 1%</td>
<td>669.9</td>
<td>670.2</td>
<td>678.9</td>
<td>681.8</td>
<td>673.7</td>
</tr>
<tr>
<td>Mut 2%</td>
<td>623.9</td>
<td>643.2</td>
<td>639.9</td>
<td>661.8</td>
<td>662.7</td>
</tr>
<tr>
<td>25 Trials</td>
<td>Stoch 0%</td>
<td>Stoch 10%</td>
<td>Stoch 20%</td>
<td>Stoch 30%</td>
<td>Stoch 40%</td>
</tr>
<tr>
<td>Mut 0.5</td>
<td>594.6</td>
<td>606.3</td>
<td>608.8</td>
<td>601.6</td>
<td>588.2</td>
</tr>
<tr>
<td>Mut 1%</td>
<td>696.9</td>
<td>707.6</td>
<td>709.5</td>
<td>700.5</td>
<td>698.5</td>
</tr>
<tr>
<td>Mut 2%</td>
<td>631.9</td>
<td>656.3</td>
<td>658.8</td>
<td>681.6</td>
<td>668.2</td>
</tr>
<tr>
<td>50 Trials</td>
<td>Stoch 0%</td>
<td>Stoch 10%</td>
<td>Stoch 20%</td>
<td>Stoch 30%</td>
<td>Stoch 40%</td>
</tr>
<tr>
<td>Mut 0.5</td>
<td>570.4</td>
<td>575.7</td>
<td>598.8</td>
<td>591.6</td>
<td>588.2</td>
</tr>
<tr>
<td>Mut 1%</td>
<td>665.6</td>
<td>675.6</td>
<td>698.9</td>
<td>697.8</td>
<td>675.7</td>
</tr>
<tr>
<td>Mut 2%</td>
<td>593.5</td>
<td>624.7</td>
<td>628.8</td>
<td>627.3</td>
<td>626.2</td>
</tr>
<tr>
<td>200 Trials</td>
<td>Stoch 0%</td>
<td>Stoch 10%</td>
<td>Stoch 20%</td>
<td>Stoch 30%</td>
<td>Stoch 40%</td>
</tr>
<tr>
<td>Mut 0.5</td>
<td>521.2</td>
<td>525.3</td>
<td>535.3</td>
<td>530.5</td>
<td>529.3</td>
</tr>
<tr>
<td>Mut 1%</td>
<td>604.2</td>
<td>615.2</td>
<td>617.2</td>
<td>613.5</td>
<td>611.6</td>
</tr>
<tr>
<td>Mut 2%</td>
<td>503.4</td>
<td>504.2</td>
<td>514.5</td>
<td>517.7</td>
<td>513.1</td>
</tr>
</tbody>
</table>
Table 3. Performance of the best agents evolved in the intermediate experimental condition obtained by systematically varying the number of evaluation trials, the mutation rate, and the level of stochasticity. The evolutionary process was continued for 50 million evaluations. Each number indicates the average results of 10 replications of each experiment. Data were obtained by post-evaluating evolved agents for 1,000 trials.

<table>
<thead>
<tr>
<th>Trials</th>
<th>Stoch 0%</th>
<th>Stoch 10%</th>
<th>Stoch 20%</th>
<th>Stoch 30%</th>
<th>Stoch 40%</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mut 1%</td>
<td>760.2000</td>
<td>755.0000</td>
<td>731.3000</td>
<td>739.5000</td>
<td>725.5000</td>
</tr>
<tr>
<td>Mut 2%</td>
<td>714.4000</td>
<td>704.2000</td>
<td>680.0000</td>
<td>685.9000</td>
<td>673.9000</td>
</tr>
<tr>
<td>Mut 4%</td>
<td>582.4000</td>
<td>578.4000</td>
<td>571.0000</td>
<td>582.8000</td>
<td>581.5000</td>
</tr>
<tr>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mut 1%</td>
<td>745.2000</td>
<td>732.9000</td>
<td>741.6000</td>
<td>739.8000</td>
<td>726.8000</td>
</tr>
<tr>
<td>Mut 2%</td>
<td>696.3000</td>
<td>702.7000</td>
<td>694.8000</td>
<td>675.4000</td>
<td>668.3000</td>
</tr>
<tr>
<td>Mut 4%</td>
<td>546.7000</td>
<td>543.0000</td>
<td>547.0000</td>
<td>525.9000</td>
<td>527.7000</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mut 1%</td>
<td>712.5000</td>
<td>714.8000</td>
<td>718.9000</td>
<td>686.4000</td>
<td>682.4000</td>
</tr>
<tr>
<td>Mut 2%</td>
<td>636.0000</td>
<td>643.5000</td>
<td>624.4000</td>
<td>658.9000</td>
<td>650.8000</td>
</tr>
<tr>
<td>Mut 4%</td>
<td>520.1000</td>
<td>495.1000</td>
<td>495.1000</td>
<td>464.3000</td>
<td>458.1000</td>
</tr>
<tr>
<td>150</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mut 1%</td>
<td>716.4000</td>
<td>694.2000</td>
<td>687.6000</td>
<td>694.8000</td>
<td>691.5000</td>
</tr>
<tr>
<td>Mut 2%</td>
<td>627.0000</td>
<td>634.7000</td>
<td>602.6000</td>
<td>654.8000</td>
<td>652.5000</td>
</tr>
<tr>
<td>Mut 4%</td>
<td>480.5000</td>
<td>471.8000</td>
<td>435.8000</td>
<td>458.3000</td>
<td>455.5000</td>
</tr>
<tr>
<td>200</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mut 1%</td>
<td>667.4000</td>
<td>676.2000</td>
<td>678.6000</td>
<td>663.8000</td>
<td>652.5000</td>
</tr>
<tr>
<td>Mut 2%</td>
<td>601.0000</td>
<td>603.8000</td>
<td>598.4000</td>
<td>584.8000</td>
<td>572.4000</td>
</tr>
<tr>
<td>Mut 4%</td>
<td>460.5000</td>
<td>451.8000</td>
<td>425.8000</td>
<td>448.2000</td>
<td>435.6000</td>
</tr>
</tbody>
</table>