UNIVERSITY OF CAPE TOWN
DEPARTMENT OF COMPUTER SCIENCE

CS/IT Honours
Final Paper 2020

Title: The Impact of a Complexity Cost on the Complexity of Robots Evolved Using Novelty Search

Author: Brooke Stewart

Project Abbreviation: EVOBAB

Supervisor(s): Geoff Nitschke

Category	Min	Max	Chosen
Requirement Analysis and Design	0	20	0
Theoretical Analysis	0	25	0
Experiment Design and Execution	0	20	20
System Development and Implementation	0	20	5
Results, Findings and Conclusions	10	20	20
Aim Formulation and Background Work	10	15	15
Quality of Paper Writing and Presentation	10		10
Quality of Deliverables	10		10
Overall General Project Evaluation (*this section allowed only with motivation letter from supervisor*)	0	10	0
Total marks		**80**	**80**

The Impact of a Cost on Complexity on the Complexity and Task Performance of Robots Evolved Using Novelty Search
Draft Paper

Brooke Stewart
University of Cape Town
STWBRO003@myuct.ac.za

ABSTRACT
An open question for both natural and artificial evolutionary systems is how, and under what environmental and evolutionary conditions complexity evolves. The *arrow of complexity* hypothesis posits that the most complex products of open-ended evolutionary systems tend toward complexity over time. This study investigates the impact a cost on overall robot complexity has on the complexity and fitness of morphology-controller couplings co-evolved for locomotion using a novelty search approach over a range of increasingly complex environments. The main finding of the study was that an energy cost imposed on complex robots during fitness evaluation had no statistical effect on the populations evolved in almost all simulation environments when compared to populations evolved without such a cost on complexity. In addition the arrow of complexity hypothesis was shown to not hold for this evolutionary system. It was also found that obstacles in the environment predictably necessitated a greater overall complexity for optimal task performance, whilst terrain friction was shown to not have a predictable impact on optimal task performance or the complexity of optimal robots.

1 INTRODUCTION

Darwin himself, in his groundbreaking work proposing the theory of descent with modification and natural selection, recognised the difficulty in conceiving that such a process could produce "organs of extreme perfection and complication", such as an eye [10]. There has since been evidence in support of Darwin's model [11, 30], however it remains difficult to track the incremental stages leading to the development of such complex features; in part due to the incompleteness of the paleobiological record and the large time-scale over which natural evolution occurs. Indeed, how and under what evolutionary and environmental conditions complexity evolves remains a largely open question for both natural [14] and artificial evolutionary systems [12, 52].

Evolutionary Robotics (ER) provides an ideal, controlled environment for simulated evolution through the application of evolutionary algorithms to the evolution of the morphology, controller, or both, for autonomous and simulated robots [22]. Two of the main approaches to research within the field of ER are Automated Engineering and Synthetic Biology [12, 16]. The Automated Engineering approach aims to successfully evolve robotic system designs fit for the given task [16], whilst the Synthetic Biology approach uses ER as a platform for hypothesis testing and investigation in the field of evolutionary biology [12]. The evolution of complexity has important implications for both of these approaches. Additionally, ER has notable overlap with the field of Artificial Life which often

makes use of tools from computer modelling and robotics to investigate emergent phenomena in life-like systems [16]. Hence, any particularly useful findings of this study may be applicable to the field of Artificial Life.

The impact of evolved complexity on engineering applications of ER is clear: the degree of controller and morphological complexity directly impacts the applicability of the evolved solutions to real-world contexts. Excessively complex solutions may be infeasible due to the expense involved in manufacturing and maintenance. Complexity management strategies, such as imposing a cost on complexity during evolution, if shown to not greatly sacrifice task performance would be valuable in ensuring the evolution of maximally feasible and effective solutions. Additionally, establishing the relationship between task or environment complexity and the relative effectiveness and complexity of the robots evolved is especially beneficial to automated engineering applications aiming to evolve solutions for remote or hazardous tasks that permit little practical testing.

The 'arrow of complexity' hypothesis from evolutionary biology asserts that the most complex products of open-ended evolutionary systems tend towards increasing complexity over time[6]. Studying the impact of environmental complexity on the complexity of evolved morphology-controller (body-brain) couplings in an open-ended ER system [26] would show whether this hypothesis holds for synthetic evolutionary biology. Additionally, the findings would contribute to the debate about whether natural selection drives the growth of biological complexity over time or, controversially, that the drive towards complexity in natural evolution is a passive force [28, 52].

In this study the concept of novelty search [28] is applied in the selection of *fit* individuals for reproduction in the evolutionary algorithm. It is often useful to refer to a *genotype* which is a representation of an individual or candidate solution [43]. Populations of robots are evolved for the common ER benchmark task of locomotion [44] across a range of increasingly complex task environments. In each task environment evolution is conducted once with, and once without, an imposed cost on overall robotic complexity. In this study the cost associated with complexity takes the form of an *energy cost* affecting the simulation time for which an individual is evaluated.

1.1 Research Objectives
The primary aim of this study will be to investigate the hypothesis that *associating a cost with complexity during evolution facilitates the evolution of lower complexity robots without sacrificing task performance.* The notion of imposing a cost on complexity is inspired

by the certain disadvantages that accompany increased complexity in nature. For instance, a random mutation is far more likely to disrupt a complex organism than to aid it therefore affecting the speed with which such organisms experience adaption [14, 38]. Hence, complex organisms do pay a cost on complexity. *The impact of a complexity cost on the complexity and fitness of individuals in the evolved population will be comparatively evaluated with respect to the population evolved without such a cost on complexity. Additionally, by comparatively evaluating the populations evolved in the different task environments, this study aims to further elucidate the environmental conditions under which complexity evolves.* Furthermore, this study contributes to the research by investigating the impact of a cost of complexity using novelty search, where previous studies have made use of objective and multi-objective search [5, 37, 39].

Novelty search imitates the open-ended dynamic seen in natural evolutionary systems by searching directly for the novel forms open-ended evolutionary systems characteristically produce [26, 28, 48]. Therefore, by investigating the trends in complexity seen over all generations of the different evolutions, this study aims to establish whether the arrow of complexity hypothesis holds for open-ended ER systems.

Previous ER studies have investigated the impact of an imposed cost on complexity [5, 37, 39]. However, the complexity metrics used in these works measure controller or morphological complexity in isolation, whilst limiting the potential complexity of the other. There have also been studies investigating the impact of environmental complexity (task difficulty) on evolved populations [4, 5] but again with the limitation of a metric only accounting for morphological complexity. Another aim of this study is to present and use a complexity metric which takes into account both morphological and controller complexity in order to address this gap in the literature. This choice is motivated by the principles of embodied cognition which dictate that intelligent behaviour arises from the coupled dynamics of an agent's body, brain and environment [2]. The use of such a metric addresses a limitation of previous studies: the strongly coupled nature of the body and mind may, where neural complexity is limited, result in the evolution of a more complex morphology in order to compensate and vice versa [2].

2 BACKGROUND & RELATED WORK

ER has grown hugely since the early 1990s [20] and as such numerous approaches and techniques have been proposed, applied and investigated. This section will go on to overview the research that has informed this study as well as the techniques applied in this study.

2.1 Behavioural Diversity Maintenance

The evolutionary algorithms used in ER rely on the Darwinian principle [10] of selection of the fittest. Darwinian evolution also relies on diversity within the population, however the operations of mutation and crossover are often insufficient to maintain enough diversity [13]. Traditional fitness function based approaches to evolutionary search implicitly assume that pursuing increased performance with respect to the objective will lead the search in the direction of desired behaviours [12]. However, such an approach can encounter the issue of *deception* [51] where the fitness-function

actively misguides the search, resulting in premature convergence around local optima. Objective-based search may also suffer from the *bootstrap problem* in ER where all individuals of a randomly generated initially population are of equal fitness resulting in no viable solution being generated by the evolutionary process [35]. The concept of diversity maintenance is that by maintaining a greater spread of individuals within the evolutionary search space the problem of early convergence the lack of fitness gradient present in the bootstrap problem can be mitigated. Diversity maintenance techniques such as fitness sharing [17] and the fitness uniform selection scheme [21] emerged as alternative approaches to objective based search.

2.1.1 Novelty Search. The novelty search algorithm, first described by Lehman and Stanley [26], takes the concept of behavioural diversity maintenance to its extreme and searches explicitly for behavioural diversity. The novelty search algorithm uses a domain-specific behavioural novelty metric in place of a traditional fitness function for the selection of fit individuals in the evolutionary algorithm. The novelty score of an individual is calculated with respect to the rest of the population and, optionally, an archive of past individuals. The archive serves to characterize the spread of solutions in behaviour space.

Novelty search does not pursue a static objective and so avoids the problem of deception [51] and premature convergence seen in objective search. A further benefit is that novelty search will discover diverse solutions whereas fitness-based evolution typically converges on a single area in the search space [19]. Novelty search has also been shown to significantly outperform objective-based search and often evolves superior behaviour [26, 28]. One of the most critical considerations for novelty search's evolutionary dynamic is the measure of behaviour similarity [25]. The chosen novelty metric should indicate the sparseness of the individual's position in novelty space.

Previous work has shown that novelty search may struggle to find optimal solutions where the behaviour space is extensive [27]. Another issue presented by novelty search is that significant resources may be squandered exploring unfruitful regions of the search space [9, 27]. Additionally, novelty search does not explicitly fine-tune solutions as the evolutionary pressure is kept towards finding new behaviours [34]. In an effort to address these shortcomings novelty search variations such as novelty search with local competition [29] have emerged. Such strategies attempt to optimise both the novelty and performance of solutions.

2.2 Body-Brain Representation and Co-evolution

How best to define a compact representation encompassing all parts of a robot, from its morphology to its controller, remains an open question in ER [12]. However, it has been shown that genotypes that allow for the simultaneous evolution of morphology and controller facilitate the evolution of more robust behaviours [7].

Directed graphs of nodes and connections have been established as a useful way to represent a robot's morphology and allow the robot to be synthesized through traversal from the root node down all connections. Additionally, recombination and mutation operations are well defined for such tree-based encodings. In Sims' seminal

work [42, 43] the nodes of the directed graph encoding a creature's morphology each describe a rigid part - the physical shape of which is determined by the associated dimensions and the relative motion between the part and its parent is constrained by the assigned joint-type. The connections between nodes describe the placement of the child part relative to the parent with separate fields describing the position, orientation, scale and reflection of the part. Similarly, the morphological genotype used in the paper by Jelisavcic et al. [22] is a directed graph where each node represents one of three components (active hinge, fixed brick, core component) available within the RoboGen framework and where edges between nodes describe the physical connections between components. In both of these approaches the nervous system of the robot is distributed throughout the body.

In Sims' work [42, 43] directed graphs representing the brain are nested within morphological nodes with which there is an association. As a result, mutation and crossover operate on the body and brain as a unit. In the paper by Jelisavcic et al. [22] the controller is two-part and consists of a morphology-independent connective Compositional Pattern Producing Network (CPPN) and a network of Central Pattern Generators (CPGs). The network of CPG nodes is implicitly represented by the morphological encoding as each active hinge defined in the robots morphological structure has an associated CPG node with the same relative position as the active hinge corresponding to the robot's core component. The connection weight between any two CPGs is found by querying the CPPN. The separation of the controller into two parts allowed the researchers to employ lifetime learning of the controller without affecting the morphology as well as enabling them to transfer the CPPN to offspring. An additional benefit of this encoding is that the HyperCube-based NeuroEvolution of Augmenting Topologies (HyperNEAT)[46] method for neuroevolution can be applied to the evolution of the connective CPPNs, which are able to represent connectivity patterns as functions in hyperspace. As a result connective connective CPPNs evolved with HyperNEAT are able to accomodate changes in the substrate network (for example the CPG network in [22]) without the need for further evolution. HyperNEAT is an extension of the NeuroEvolution of Augmenting Topologies (NEAT) method for neuroevolution which was shown to outperform the best fixed-topology neuroevolution method and resulted in significantly faster learning [47]. Furthermore, connective CPPNs evolved with HyperNEAT are able to indirectly encode large scale Artificial Neural Networks (ANN) which are the most widely used controller paradigm in ER[12].

2.3 Evolving Complexity in Evolutionary Robotics

Two previous studies by Auerbach and Bongard [4, 5] have examined the impact of environmental complexity on the evolved morphological complexity of virtual organisms with restricted nervous systems. The same morphological complexity metric - the Shannon diversity [40] of the external curvature of the trimesh robot morphologies - is applied in both studies and in both studies the virtual organisms are evolved in a simple, flat control environment and 49 other environments of increasing complexity. The more complex environments consist of infinite series of differently spaced

low-friction, rectangular solids of varying heights over which the virtual organisms must move. The fitness of the organism is determined by the directed displacement it achieves in the environment in a fixed amount of time. The results of the earlier study [4] lend some support to the hypothesis that the complexity of a robot's task environment could create an evolutionary pressure leading to the evolution of more complex morphologies, as some environments saw the evolution of more complex morphologies than seen in the simple control environment. However, the majority of environments evolved organisms with morphologic complexities not significantly different to those of the organisms evolved in the control environment. Overall, the results demonstrate that the increased environment complexity did not necessarily result in increased morphological complexity. These results contrast with the results of the later study [5] in which a multi-objective selection mechanism that selects for simplicity as well as locomotion ability is applied. In this experiment the 49 more complex environments all actively induce a selection pressure favouring greater morphological complexity than simpler environments when a cost is imposed on complexity.

Contrary to the findings of [5], the key finding of Furman, Nagar and Nitschke's work [37] was that the imposition of a complexity cost on morphological complexity enabled the evolution of simpler morphologies with negligible differences in task performance. In both studies, evolution was performed over a number of environments of different complexity. In this study a collective robotics gathering task was performed in environments with differing numbers and sizes of blocks and different mandated degrees of cooperation. In each environment robots were evolved using a single-objective evolution method for comparison with those evolved using a multi-objective evolution method. Both methods attempt to maximize task performance which in this case is the number of blocks successfully moved into the gathering zone in the given time. The multi-objective method concurrently tries to minimise morphological complexity - thus imposing a cost on complexity.

Previous work has also investigated the effect of environmental complexity on neural complexity. Revello and McCartney [39] found that the inclusion of a dynamically scaled cost term in the fitness function used in evolution not only reduced the complexity of the control program evolved for the robot, but also that these solutions had improved performance compared to those evolved without cost terms. The cylindrical, wheeled robots in this experiment have *bump* and *sonar* sensors placed on the front and 90° to either side. The configuration of used sensors and robot control program are co-evolved during simulation. The robots were evolved in a simple maze environment, the same maze environment but with noisy input to sensors and in a purposely deceptive maze environment. Fitness was determined by the amount of progress made toward the goal less the cost terms associated with program size (in bits) and time of execution. The cost terms aimed at limiting controller complexity were scaled by the progress made toward the goal so as not to restrict the early exploration of the solution space. In all scenarios the application of the cost terms allowed the effective management of the complexity of the evolved control programs.

3 METHODS

3.1 Evolutionary Body-Brain Representation

The body-brain representation used in this paper is similar to that used in a paper by Jelisavcic et al. [22] where simulated robots were successfully co-evolved for locomotion, thus demonstrating the suitability of this representation for this experiment. The similarity in representation arises from the fact that they based their chosen representation off the one provided by the Robogen framework which is the one we use. Additionally, the chosen body-brain representation allows for a progression in complexity over time which has been shown to have demonstrable benefits over encodings that do not allow for this [5].

The morphological genotypes of the robots in this work are represented by directed graphs where each node of the graph is representative of a morphological module (part) provided by Robogen and the root node represents the core component. Each node contains information pertaining to the attachment of child nodes (parts) and the part's own orientation. The possible parts comprising a robot's morphology are limited to: active hinges, passive hinges and fixed bricks in addition to the core component. This was done to simplify the complexity calculation and analysis of the evolved morphologies. Additionally, robot morphologies are limited to a maximum of 50 parts. However, this restriction should not impede morphological development as it provides quite a generous upper bound on the number of morphological parts.

The controller representation for robots is two part. First, there is a Central Pattern Producing Network (CPPN) associated with each robot. Secondly each robot has an Artificial Neural Network (ANN) 'brain'. The sensors present on the robot's morphology serve as inputs to the ANN and the output neurons of the ANN correspond with motors (for active hinges). Additionally, the ANN may include a number of internal or hidden units not connected to any inputs or outputs. The ANNs in Robogen differ from classical recurrent ANNs in that the possibility of oscillator neurons is allowed alongside the default sigmoid neurons [3]. The CPPN associated with the robot is used to set the weights of the substrate ANN. An example of a robot and its associated representation is presented in Figure 1.

3.2 Evolutionary Search

The approach to evolutionary search taken in this study is close to the original novelty search approach described by Lehman and Stanley [26] with some adjustments. The novelty score of an individual is calculated with respect to the rest of the population and an archive of maximum fifty past individuals upon it's origin. In this study the archive of past individuals is randomly composed and maintained. That is, on the origin of a new individual it will be probabalistically added to the record of past individuals in the novelty archive. Gomes et al. [18] showed that the use of such an archive outperforms an archive consisting of only highly novel past individuals as used by Lehman and Stanley [26]. Their results also demonstrated that novelty calculations with respect to the current population alongside the archive yielded better results than when novelty is only calculated using an archive. In this work the novelty of individuals is calculated with respect to the 15 nearest neighbours in the population and archive. Gomes et al. [18] showed

that this commonly chosen number of nearest neighbours for the novelty calculation performed well across all metrics.

3.2.1 Novelty Metric. Genotypic difference doesn't always correlate with behavioural differences [13]. Hence, the approach to measuring novelty taken in this paper attempts to directly measure behavioural difference. To this end, the final ending position of each robot is recorded after evaluation. After the newly created individuals have been evaluated, and their end positions recorded, the novelty score for each of them is calculated. The novelty score is calculated as the average euclidean distance between the end position achieved by the robot and the ending positions of the 15 nearest robots in the population and archive. Each of these newly evaluated individuals is then added to the novelty archive with a 30% probability. The use of ending positions to characterise behavioural novelty was successfully applied by Lehman and Stanley in evolving robots with novelty search for a maze navigation task [28].

This novelty metric captures how well explored the region of the environment which the robot reaches during its simulation is. In this case a low novelty score indicates that the robot has similar locomotive ability to others due to the similarity in ending positions achieved. A high novelty score indicates that the robot has achieved a unique ending position. Conceptually we can also see that this metric subliminally guides the search toward greater locomotion ability: robots always start their evaluation at the center of the environment and hence this area of the behaviour space will quickly be populated by robots with low locomotive ability. Subsequently, robots will have to achieve a greater distance in order to reach more sparsely populated regions and achieve greater novelty scores.

3.3 Body-Brain Co-Evolution

In this study full-bodied evolution was conducted meaning that the robot body and brain are simultaneously evolved. The Robogen Framework was extended for this project to enable neural evolution (of the CPPNs) with a simplified version of the HyperNEAT [46] approach when conducting full-bodied evolution. The morphologies are simultaneously evolved through application of mutation operations to the morphological representation. The possible mutation operations performed on the morphological representation include: node (part) removal and addition, along with subtree duplication, removal and swapping. The simplications to the HyperNEAT method were made due to time constraints as well as the complexity involved in extending Robogen.

Instead of initialising a population of minimal CPPNs with random weights, each member of the initial robot population is directly assigned a minimal CPPN initialised with random weights. The CPPN assigned to a robot is then permanently associated with that robot as part of its neural representation. This direct assignment means that assigning historical markers to keep track of new genes, and dividing the population into species is avoided and reproduction simply occurs between the two CPPNs of the parents selected by the evolutionary algorithm. The CPPN created is then assigned to the robot offspring. The approach to body-brain co-evolution is summarised by the following steps:
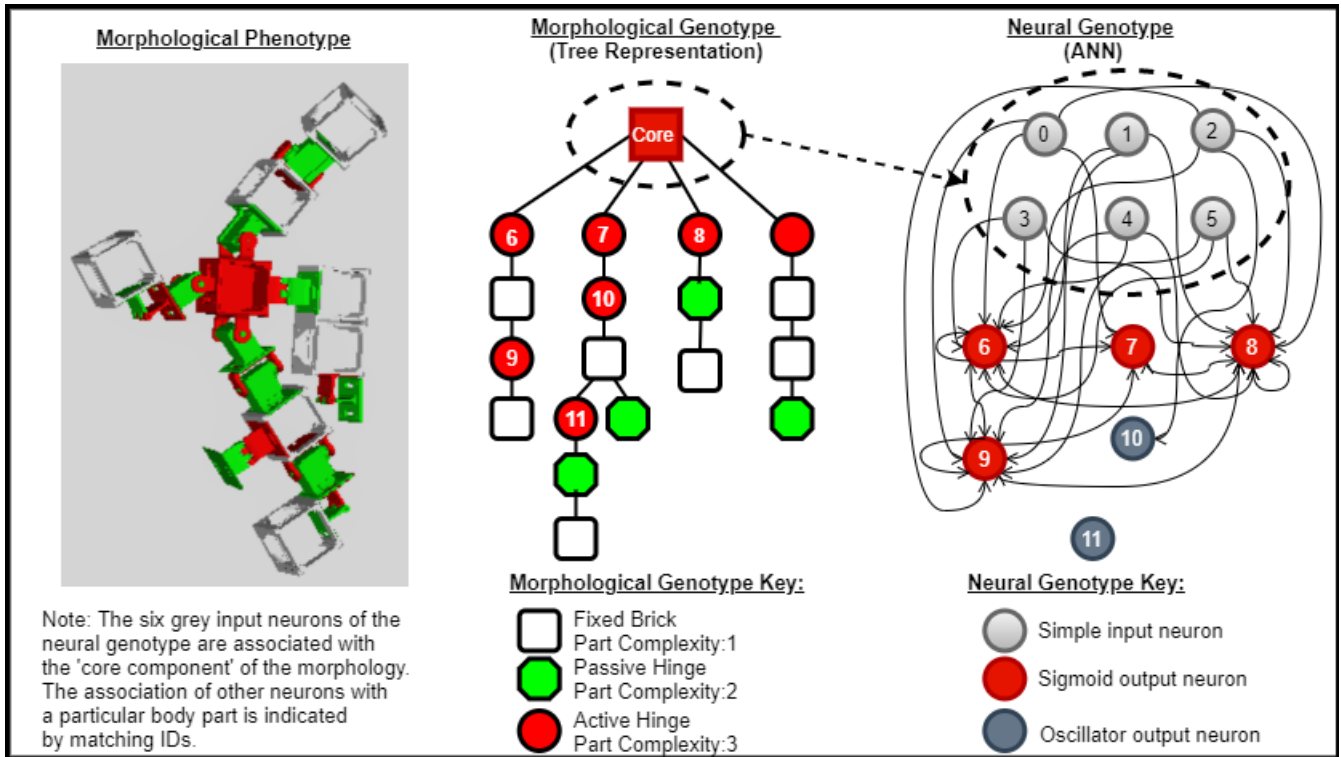
Figure 1: This figure presents an example robot (left) alongside its evolutionary representation. The association between parts of the robot's morphological representation (middle) and particular neurons in the ANN (right) is indicated. The second part of the neural representation of the robot, the CPPN, is not shown here.

(1) The initial population of robots is created by seeding the morphologies from a seed robot and assigning each a minimal CPPN with randomised weights.
(2) For each robot representation: weights in the substrate ANN are filled by querying the CPPN.
(3) The population is evaluated.
(4) Each individual in the population is assigned a novelty score and is probabalistically added to the novelty archive.
(5) Until required number of offspring are created:
  (a) Parents are selected from the population
  (b) The morphology of the first parent is copied into the child
  (c) The CPPNs of the parents are mated and the resulting CPPN is assigned to the child
  (d) The morphology of the child is then mutated through stochastic mutation operations performed in the morphological representation
(6) Steps 2 to 4 are performed for the newly created population of children
(7) The population of children is combined with the current population in accordance with the replacement strategy
(8) Steps 5 to 7 are repeated for the given number of generations.

3.4 Simulation Environments

The twelve simulation environments used in this study were designed in four sets of three, where each consecutive set increasing in environmental complexity from the previous one. The division of the environments into sets can be seen in Table 1. The first environment set is a perfectly flat and unobstructed and is intended to serve as the control environment set. The next set of environments consists of mirrored, evenly spaced, low obstacles over which the simulated robot must move. The challenge presented by this set of environments is then made greater in the next set of environments which contains mirrored, irregularly spaced and sized obstacles of differing heights. The final set of environments consists of tilted planes thus introducing the challenge of moving up a gradient. The environment design for each set can be seen in Figure 2.

Each set consists of three identically environments except that one has full surface friction (1.0), one has moderate surface friction (0.6) and one is a low friction environment (0.2). This means there are three grades of difficulty within each set. This extra division allows for an even deeper investigation into the impact of the environmental complexity on evolved complexity and task performance.

These environments were implemented using the available physics simulator configuration options available as part of the Robogen Framework [3]. All robots start their simulation in the center of the environment. The environments were designed to be mirrored so as not to restrict or advantage the robots' locomotion in any direction and thus allow the greatest potential for the evolution of genuine locomotive ability.
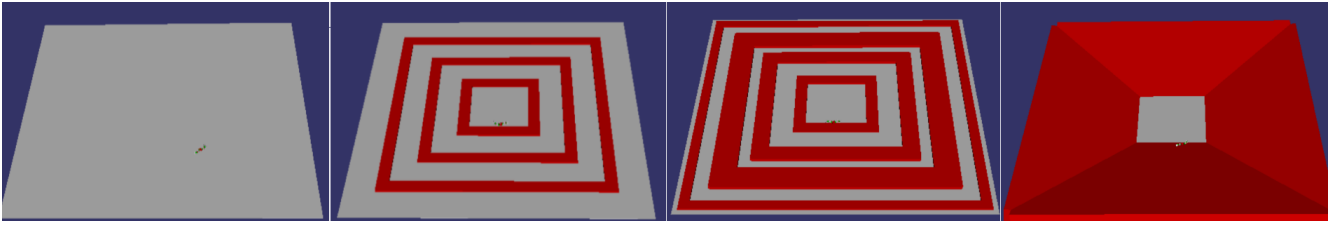
Figure 2: Simulation Environments from left to right: Environment Set 1 (flat), Environment Set 2 (regular obstacles), Environment Set 3 (irregular obstacles), Environment Set 4 (tilted)

Environment Set	Set1			Set2			Set3			Set4		
Environment	1	2	3	4	5	6	7	8	9	10	11	12
Terrain Friction	1.0	0.6	0.2	1.0	0.6	0.2	1.0	0.6	0.2	1.0	0.6	0.2

Table 1: Here the classification of the twelve simulation environments into sets is shown. The environment configuration for each environment set can be seen in Figure 2 and the terrain friction for each environment is indicated.

3.4.1 Fitness Evaluation. The objective fitness score for each individual in the evolving population is never used during evolution, but was still calculated and stored in order to track the locomotion ability (task performance) of the population. This fitness score was calculated as the minimum euclidean distance the robot achieved in the environment from it's initial start position over all simulated evaluations. This is a commonly chosen measure of objective fitness for locomotion [44]. Robots are simulated in the environment for 10 seconds, unless a cost is imposed on complexity as explained in the next section.

3.5 Imposing a Cost on Complexity

The approach to imposing a cost on complexity in this study was to impose an *energy cost* based on the genotypic complexity of the robot. This approach draws its inspiration from evolutionary biology [38]. This energy cost was realised in the form of a reduction in the simulation time a robot's fitness was evaluated for. The calculation for the adjusted simulation time for a robot given the robot's complexity score (which is scaled to a value between 0 and 1) is calculated using the following equation where the default simulation time is 10 seconds:

$$simulationTime = 10.0 - ((1 - complexityScore) * 10.0) \quad (1)$$

Previous studies have used multi-objective functions that explicitly select for lower complexity alongside performance as the means of imposing an evolutionary cost on complexity [5, 37]. The reasoning behind the different approach taken in this study is to allow for the evolution of complexity where it confers a significant boost in the efficiency of task performance. The results of the complexity cost experiment will therefore better represent the relationship between task difficulty, task performance and robot complexity. Additionally, explicitly selecting for individuals of lower complexity would disrupt the open-ended evolutionary dynamic created by novelty search.

Obviously, in order to impose a cost on complexity there must be some measure of an individual's complexity. The complexity metric used in this study is presented in the next section.

3.6 Complexity Metric

The aim of this metric is to adequately quantify the overall complexity of the robot. That is, the complexity score should be representative of both morphological (body) and controller (brain) complexity. Complexity scores are scaled to between zero and one, where a score of one would indicate a maximally complex robot. The overall complexity was calculated with the following weighted sum:

$$complexityScore = (0.7 * morphologicalComplexity) + (0.3 * neuralComplexity) \quad (2)$$

3.6.1 Morphological Complexity. The morphological complexity calculation relies on the inherent complexity of the parts that make up the robot's morphology. The fixed brick is the simplest possible part and was therefore assigned a complexity of 1. The passive hinge is able to facilitate movement, making it more complex than the fixed brick, and is therefore assigned a complexity of 2. Additionally, passive hinges present an opportunity for morphology facilitated control [36] whereby the morphology can contribute joint manipulation not explicitly performed by the controller. The active hinge is the most complex part available as it can be intentionally manipulated and it is always associated with an output neuron of the ANN. Hence it was assigned the highest part complexity of 3.

The morphological complexity is simply the sum of the part complexity for each part found in the morphological genotype. The morphological complexity is scaled to a number between 0 and 1 using the minimum complexity score possible (2) and the experimentally determined maximum morphological complexity score possible (90).

For example, application of this morphological complexity metric to the robot presented in Figure 1 would begin with the sum of the part complexity of all nodes in the morphological genotype resulting in a total value of 36. Scaling this value using the minimum and maximum complexity scores gives the final morphological complexity score of 0.3469. We see this score is relatively low as the number of parts comprising the robot is far less than the maximum number of parts permitted (50 parts).

The higher weight attributed to the morphological complexity component in the weighted sum is due to the fact that the morphological graph is representative of some aspects of neural complexity.

For instance, the number of active hinges corresponds with the number of output neurons found in the ANN. Additionally, the morphology dictates the number of possible sensor inputs to the ANN. Thus the morphological genotype is at least partly representative of the neural complexity as the network topology bounds the complexity of achievable behaviours [12].

3.6.2 Neural Complexity. The morphological complexity score in part represents the possible complexity of the brain, but just how complex the 'brain' actually is remains to be seen. The inspiration behind the metric presented here stems from two previous information-theoretic neural complexity measures presented in work by Tononi et al. [50] and Tononi and Sporns [49]. Both of the measures presented in these papers are maximised when two properties seen in the brains of higher vertebrates are show to coexist: namely functional segregation and local specialisation should coexist alongside global integration [50]. This measure avoids maximisation by randomness and in an intuitive sense captures the manner in which the system processes information at multiple scales. The measure was shown to be low in completely integrated or completely segregated systems and high where a balance between local specialisations and global integration was achieved. This logic informs the design of the neural complexity metric presented here.

Tononi and Sporns [49] identify complexes which are subsets of the system, not part of any larger subset, that are capable of integrating information. Functional specialisation is shown by the connection patterns developed between the elements of a complex while global integration is a measure of the ability to exhange information over any bipartition of the network. The requirement for interactions between components at multiple levels in the system can be seen as a unifying principle of complexity [24]. In this work, the identification of 'complexes' takes the form of the identification of strongly connected components (groups of neurons where a path exists between all pairs of neurons). These groupings of neurons represent a computational unit (local specialisation) capable of producing isolated functionality since the output neurons of the ANN control associated active hinges of the morphology. The specialisation of the strongly connected components is further exemplified by the number of distinct cycles which have developed. Additionally, the division of the ANN into strongly connected components in part captures the modularity of the network. The modularity observed in many biological structures and processes has been noted as a factor which greatly contributes to evolvability [8]. By definition, information can be exchanged between any bipartition of a strongly connected component, hence the ability of the network to exchange information over any bipartition can be determined by the existence of inter strongly connected component connections. The connectivity of a graphs has also previously been used as a measure of complexity [45].

The brain complexity calculation can be described by the following set of steps:

(1) An adjacency list representation of the ANN is constructed
(2) Neurons are grouped into strongly connected components of more than one neuron
(3) Connections between strongly connected components are tallied

(4) The degree of global integration present in the brain is estimated by the following equation:

$$globalIntegration = numberOfStronglyConnectedComponents/$$
$$numberOfInterStronglyConnectedComponentConnections$$
$$(3)$$

(5) The number of cycles present in each strongly connected component is determined using Johnson's algorithm [23]
(6) The proportion of the brain specialised is calculated as the ratio of the total number of neurons in a strongly connected components (specialisations) to the total number of neurons in the ANN.
(7) The average degree of specialisation is estimated by the average number of cycles present in the strongly connected components
(8) The brain is assigned a specialisation score calculated as follows:

$$specialisationScore = averageDegreeOfSpecialisation*$$
$$proportionOfBrainSpecialised \quad (4)$$

(9) Finally, in order to determine the balance of the brain in terms of local specialisation and global integration, the globalIntegration and specialisation score are ratioed to give an overall score of neural complexity.

4 EXPERIMENTS

In order to determine the impact of a cost on complexity on the complexity and task performance of evolved robots, two sets of experiments were conducted for comparison [1]. In the first, which will henceforth be referred to as the baseline experiment, evolution proceeded without a cost on complexity and in the second, which will be referred to as the complexity cost experiment, a cost on complexity was imposed during evaluation as described in Section 3.5. In both experiments populations of one hundred morphology-controller couplings (robots) were co-evolved (as described in section 3.3) for locomotion for one hundred generations in each of the twelve simulation environments described in Section 3.4 in order to further elucidate the effect of environmental complexity on task performance and robot complexity as well as the conditions under which complexity evolves. Ten runs were conducted for each experiment. The experimental parameters common to both experiments are summarised in Table 2.

This study made use of the open-source Robogen Framework's [3] evolution engine and physics simulator as the basis for the co-evolution of robot morphologies and controllers and the evaluation of robots in the simulation environments [2]. Both experiments were run on an Ubuntu Virtual Machine [3].

5 RESULTS AND DISCUSSION

For each environment, the average task performance, average complexity, average maximum task performance and associated average complexity of the populations evolved over the ten runs of the baseline and complexity cost experiments are statistically compared for

---

[1]https://github.com/BrookeSte/EVOBAB
[2]http://robogen.org/
[3]8 Intel(R) Xeon(R) Gold 6254 CPUs @ 3.10GHz. 32GB RAM.

Experiment Parameters	
Experiment	Complexity Cost or Baseline
Runs	10
Simulation Time	Default 10.0s (If complexity cost experiment - simulation time adjusted as described in section 3.5)
Environment	Environment Parameters specified in table 1
Evolutionary Algorithm Parameters	
Evolution Mode	Full-Bodied (body-brain co-evolution)
Use Seed Robot	True [4]
Population Size	100
Generations	100
Parent Selection	Deterministic Tournament Selection (based on novelty score)
Tournament Size	2
Number of Parents (μ)	100
Number of offspring (λ)	100
Mutation Probability	.3
Probability of addition to Novelty Archive	.3
Replacement Strategy	Plus replacement [15] (based on novelty score)
Novelty Archive Size	Maximum size of 50

Table 2: Table summarising the Experiment Parameters

each environment to determine the impact of a cost on complexity. Application of the Shapiro-Wilk test [41] showed that these datasets do not necessarily exhibit normal distributions for all environments. Hence, the non-parametric Mann-Whitney U-test [32] ($p<0.05$) was used to test for statistical difference. Levene's test [31] was performed on the datasets prior to comparison to ensure that the assumption of equal variances required by the Mann-Whitney U-test was met. The outcomes of these comparisons is recorded in table 3.

As seen in table 3, almost all comparisons between the baseline and complexity cost experiments show no statistical difference between the experiments. The lack of statistical difference between the two experiments in terms of average evolved complexity is to be expected as in both experiments the selective pressure is solely toward novel behaviours. However, we see that in environment 6 and 7 there is a statistical difference in the average complexity evolved over the ten runs in the different experiments. This is hypothesized to be a result of the relatively low number of repeats.

Due to the energy cost imposed on the most complex individuals of the population during evolution, we may have predicted a difference in the average task performance, best average task performance and perhaps, consequently the average complexity associated with best task performance between the baseline and complexity cost experiments. However, we observe in figures 4 and 3 as well as table 6 that the populations evolved by both experiments show no statistical difference for these criteria for almost all of the simulation environments in which evolution was conducted. This may seem surprising, but this result could just be a result of the explorative nature of novelty search enabling the search to find similarly fit individuals in spite of the imposed cost on complexity. Another possibility is that the energy cost was too lenient and had negligible effect on the task performance of individuals during simulation although conceptually this seems unlikely to be the case.

Since novelty is calculated with respect to the ending position the robot achieves in simulation, the task performance of the robot has

some bearing on the novelty score it achieves. In retrospect the task performance without an imposed cost on complexity should have been recorded in addition to task performance with an imposed energy cost for each individual in the complexity cost experiment. As it stands, the results do not provide enough evidence to rule out the possibility that more efficient task performance evolved in the complexity cost experiment allowing individuals of similar complexity to those evolved in the baseline experiment to achieve statistically similar task performance in spite of the energy cost.

In order to further elucidate the environmental conditions under which complexity evolves Mann-Whitney U-tests [32] ($p<=0.05$) were used to test for statistical difference between the complexity associated with best task performance between environments within the same set (i.e. same obstacle configuration) and between environments with the same terrain friction. The results of these comparisons can be seen in table 4 and 5. In both environment set 1 (flat) and environment set 3 (irregular obstacles) all the average highest complexity is statistically comparable indicating that terrain friction had no impact on the complexity of the fittest individual. In figure 3 we see that set 1 exhibits the highest task performance whilst set 3 exhibits the lowest suggesting that friction played little to no role in the first instance because the environment configuration made the task too easy and in the second instance for the opposite reason. In set 2 lower friction correlates with lower complexity whilst the opposite is true for set 4. From these results we may conclude that terrain friction cannot necessarily be considered as a marker of environmental complexity in isolation of the rest of the environmental configuration.

Where environments with the same friction were compared (i.e. the obstacle configuration is what differs) in general, the environments in set 3 were associated with greater evolved complexity. The environments in set 2 were comparable with those in set 3 in terms of evolved complexity, except at the lowest terrain friction where the environments in set 3 evolved resulted in the evolution of higher complexity. From this we can conclude that the irregularly space and sized obstacles of set 3 pose the greatest challenge resulting in relatively higher complexity robots just to achieve relatively low task performance. In general the second highest complexity was induced by environment set 2 while environment set 1 and 4 were statistically comparable in terms of the complexity evolved.

Looking at Figure 4 we clearly see that the average complexity associated with best task performance does not exceed even half of the possible complexity score. There are no other studies which measured overall robot complexity with which we can compare our results and complexity metric. It is very possible that the main reason for the relatively low complexity scores in general is that the upper bound of morphological complexity was too generous and as a result the morphological complexity scores are scaled down too harshly. On the other hand, previous work has noted the big influence the definition of the phenotype has over complexity values [1]. Since in this work, selection is behaviour-based it is also possible that the definition of behaviour did not warrant higher complexity values. Regardless, one of the shortcomings of this project is the lack of comparison and evaluation of this complexity

Environment Number	Average Task Performance	Average Complexity	Best Task Performance	Complexity Associated with Best Task Performance
1	B == CC	B == CC	B == CC	B == CC
2	B == CC	B == CC	B > CC	B == CC
3	B == CC	B == CC	B == CC	B == CC
4	B == CC	B == CC	B == CC	B == CC
5	B == CC	B == CC	B == CC	B == CC
6	B == CC	B < CC	B == CC	B < CC
7	B == CC	B > CC	B > CC	B > CC
8	B == CC	B == CC	B == CC	B == CC
9	B == CC	B == CC	B == CC	B == CC
10	B == CC	B == CC	B == CC	B == CC
11	B == CC	B == CC	B == CC	B == CC
12	B == CC	B == CC	B == CC	B == CC

Table 3: Statistical task performance and robotic complexity comparisons of the average and best evolved robots in the Baseline (B) and Complexity Cost (CC) experiments for all simulation environments. == : Statistically Comparable. <,>: Statistical Difference, less than or more than used to indicate the dataset with the higher median value. Robot complexity is defined in Section 3.6 and Task Performance is defined in Section 3.4.1.
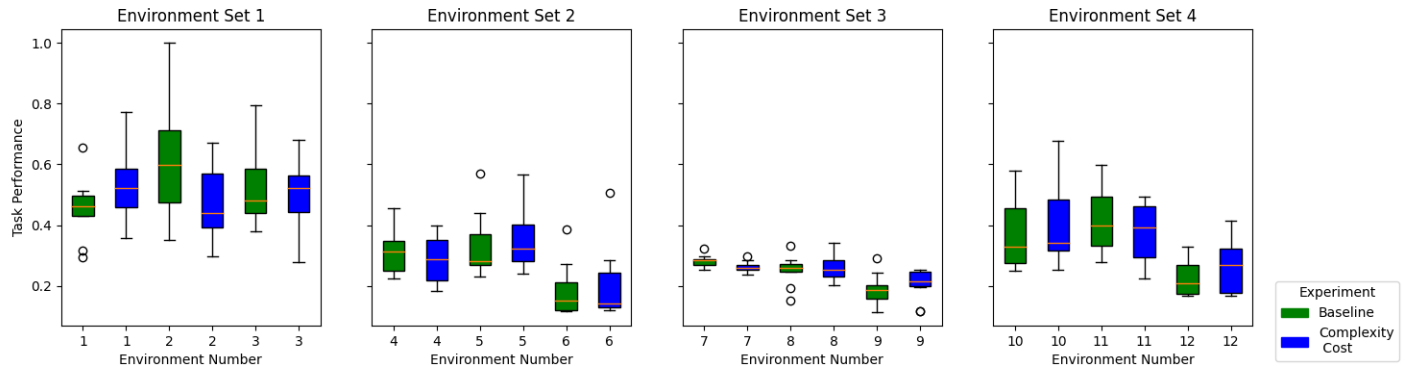
Figure 3: Average maximum task performance of the Baseline versus Complexity Cost experiments in all environments
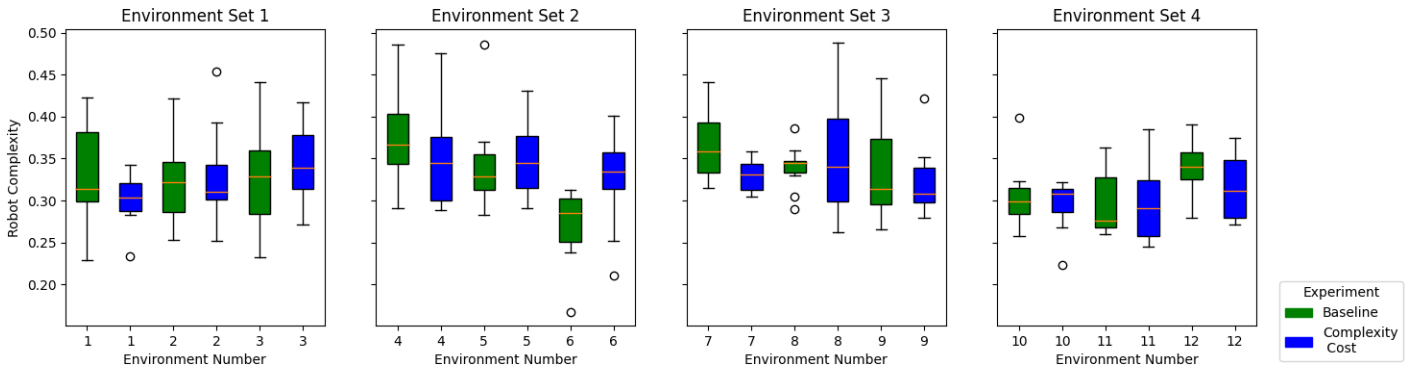

Figure 4: Average maximum complexity (complexity associated with maximum task performance) of the Baseline versus Complexity Cost experiments in all environments

metric as a measure of overall complexity. Unfortunately, designing and implementing another metric for comparison was beyond the scope of this project.

Statistical Comparison of Average Highest Complexity Evolved in Environments Within the Same Set															
Set 1				Set 2				Set 3				Set 4			
	1	2	3		4	5	6		7	8	9		10	11	12
1	-	1==2	1==3	4	-	4==5	4>6	7	-	7==8	7==9	10	-	10==11	10<12
2	2==1	-	2==3	5	5==4	-	5>6	8	8==7	-	8==9	11	11==10	-	11<12
3	3==1	3==2	-	6	6<4	6<5	-	9	9==7	9==8	-	12	12>10	12>11	-

Table 4: Statistical comparisons (between environments with the same obstacle configuration) of the average highest complexity evolved in the baseline experiment. == indicates the average highest complexity evolved in the two environments is statistically comparable. Where the results are not statistically comparable (p<=0.05), < and > are used to indicate the environment in which the statistically larger highest average complexity was evolved

Statistical Comparison of Average Highest Complexity Evolved in Environments With the Same Terrain Friction															
Terrain Friction : 1.0					Terrain Friction : 0.6					Terrain Friction : 0.2				
	1	4	7	10		2	5	8	11		3	6	9	12
1	-	1==4	1<7	1==10	2	-	2==5	2==8	2==11	3	-	3>6	3==9	3==12
4	4==1	-	4==7	4>10	5	5==2	-	5==8	5>11	6	6<3	-	6<9	6<12
7	7>1	7==4	-	7>10	8	8==2	8==5	-	8>11	9	9==3	9>6	-	9==12
10	10==1	10<4	10<7	-	11	11==2	11<5	11<8	-	12	12==3	12>6	12==9	-

Table 5: Statistical comparisons (between environments of the same friction) of the average highest complexity evolved in the baseline experiment. ==, < and > have the same meaning as in table 4
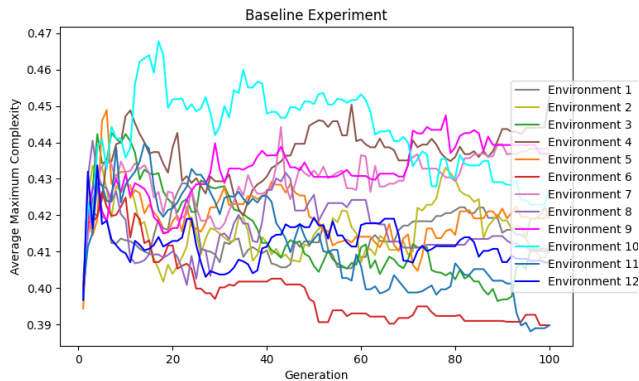

Figure 5: The average maximum complexity achieved per generation over all runs of the baseline experiment is plotted for all simulation environments

Figure 5 plots the average maximum complexity score achieved by a member of the robot population over all generations for the baseline experiment. We can clearly see that the complexity of the maximally complex individual does not necessarily tend toward increasing complexity over time, hence the arrow of complexity hypothesis does not hold for this evolutionary system. Miconi [33] presents a detailed and convincing argument that the arrow of complexity does occur in evolution and that it is most reasonably interpreted as the result of a passive trend than a driven one. Miconi, specifies the conditions under which Darwinian evolution will generate a sustained trend of increasing maximum complexity. One of the requirements is that at any given time, the probability of mutation toward successful designs of higher and equal complexity

should be higher than that of reaching designs that would make the starting points of these jumps unsuccessful. In this study, selection by novelty does not necessarily have a higher probability of selecting successful designs of equal or higher complexity. Therefore, we hypothesize that the contravention of this requirement is the reason why the arrow of complexity hypothesis doesn't hold for this system as the other noted requirements are met.

6 CONCLUSIONS AND FUTURE WORK

The results demonstrate that the imposition of an energy cost had practically no effect on the average complexity and complexity associated with best task performance of evolved robots as compared with populations evolved without such a cost on complexity. Additionally, the arrow of complexity hypothesis was shown not to hold for this evolutionary system and as a result no predictions on the maximal complexity evolved in such a system can easily be made. Perhaps surprisingly, the results also show that average task performance and best average task performance were almost entirely unaffected by the cost on complexity. However, we cannot entirely rule out the possibility that the imposition of a cost on complexity allowed the evolution of more efficient task performance. Future work would do well to extend this research by investigating the impact of a complexity cost imposed through multi-objective selection selecting for novelty and lower complexity thus introducing an explicit selective pressure toward lower complexity.

In terms of the environmental conditions under which complexity evolves it was found that obstacles in the environment predictably necessitated a greater overall complexity for optimal task performance, whilst terrain friction was shown to not have a predictable impact on optimal task performance or the complexity of optimal robots. Irregularly sized and spaced obstacles posed the

greatest task complexity, followed by the regularly spaced, low obstacles. It was also discovered that the average highest complexity evolved in the tilted environment was statistically comparable to that evolved in the flat environment where terrain friction was the same.

7 ACKNOWLEDGEMENTS

REFERENCES

[1] C. Adami and C. Adami. 1998. *Introduction to artificial life.* Springer Science & Business Media.
[2] M. Anderson. 2003. Embodied cognition: A field guide. *Artificial intelligence* 149, 1 (2003), 91–130.
[3] J. Auerbach, D. Aydin, A. Maesani, P. Kornatowski, T. Cieslewski, G. Heitz, P. Fernando, I. Loshchilov, L. Daler, and D. Floreano. 2014. Robogen: Robot generation through artificial evolution. In *Artificial Life Conference Proceedings 14.* MIT Press, 136–137.
[4] J. Auerbach and J. Bongard. 2012. On the relationship between environmental and morphological complexity in evolved robots. In *Proceedings of the 14th Annual Conference on Genetic and Evolutionary Computation (GECCO '12).* ACM, 521–528. https://doi.org/10.1145/2330163.2330238
[5] J. Auerbach and J. Bongard. 2014. Environmental influence on the evolution of morphological complexity in machines. *PLoS computational biology* 10, 1 (2014).
[6] M. Bedau. 1998. Four Puzzles About Life. *Artificial Life* 4, 2 (1998), 125–140. https://doi.org/10.1162/106454698568486
[7] J. Bongard. 2011. Morphological change in machines accelerates the evolution of robust behavior. *Proceedings of the National Academy of Sciences* 108, 4 (2011), 1234–1239.
[8] J. Clune, J. Mouret, and H. Lipson. 2013. The evolutionary origins of modularity. *Proceedings of the Royal Society b: Biological sciences* 280, 1755 (2013), 20122863.
[9] G. Cuccu and F. Gomez. 2011. When novelty is not enough. In *European Conference on the Applications of Evolutionary Computation.* Springer, 234–243.
[10] C. Darwin. 1968. On the origin of species by means of natural selection. *London: Murray Google Scholar* (1968).
[11] R. Dawkins. 1986. The blind watchmaker: Why the evidence of evolution reveals a universe without design, Nueva York.
[12] S. Doncieux, N. Bredeche, J. Mouret, and A. Eiben. 2015. Evolutionary robotics: what, why, and where to. *Frontiers in Robotics and AI* 2 (2015), 4.
[13] S. Doncieux and J. Mouret. 2010. Behavioral diversity measures for evolutionary robotics. In *IEEE congress on evolutionary computation.* IEEE, 1–8.
[14] A. Edwards. 2000. The Genetical Theory of Natural Selection. *Genetics* 154, 4 (2000), 1419–1426.
[15] A. Eiben and M. Schoenauer. 2002. Evolutionary computing. *Inform. Process. Lett.* 82, 1 (2002), 1–6.
[16] D. Floreano and J. Urzelai. 2000. *Evolutionary Robotics: The Next Generation.* AAI Books. http://infoscience.epfl.ch/record/63910/files/er2000.pdf
[17] D. Goldberg and J. Richardson. 1987. Genetic algorithms with sharing for multimodal function optimization. In *Genetic algorithms and their applications: Proceedings of the Second International Conference on Genetic Algorithms.* Hillsdale, NJ: Lawrence Erlbaum, 41–49.
[18] J. Gomes, P. Mariano, and A. Christensen. 2015. Devising effective novelty search algorithms: A comprehensive empirical study. In *Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation.* 943–950.
[19] J. Gomes, P. Urbano, and A. Christensen. 2013. Evolution of swarm robotics systems with novelty search. *Swarm Intelligence* 7, 2-3 (2013), 115–144.
[20] I. Harvey, E. Paolo, R. Wood, M. Quinn, and E. Tuci. 2005. Evolutionary robotics: A new scientific tool for studying cognition. *Artificial life* 11, 1-2 (2005), 79–98.
[21] M. Hutter. 2002. Fitness uniform selection to preserve genetic diversity. In *Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02 (Cat. No. 02TH8600)*, Vol. 1. IEEE, 783–788.

[22] M. Jelisavcic, K. Glette, E. Haasdijk, and A. Eiben. 2019. Lamarckian Evolution of Simulated Modular Robots. *Frontiers in Robotics and AI* 6 (2019), 9.
[23] D. Johnson. 1975. Finding all the elementary circuits of a directed graph. *SIAM J. Comput.* 4, 1 (1975), 77–84.
[24] W. Kinsner. 2008. Complexity and its measures in cognitive and other complex systems. In *2008 7th IEEE International Conference on Cognitive Informatics.* IEEE, 13–29.
[25] S. Kistemaker and S. Whiteson. 2011. Critical factors in the performance of novelty search. In *Proceedings of the 13th annual conference on Genetic and evolutionary computation.* 965–972.
[26] J. Lehman and K. Stanley. 2008. Exploiting open-endedness to solve problems through the search for novelty. In *ALIFE.* 329–336.
[27] J. Lehman and K. Stanley. 2010. Revising the evolutionary computation abstraction: minimal criteria novelty search. In *Proceedings of the 12th annual conference on Genetic and evolutionary computation.* 103–110.
[28] J. Lehman and K. Stanley. 2011. Abandoning objectives: Evolution through the search for novelty alone. *Evolutionary Computation* 19, 2 (2011), 189–223.
[29] J. Lehman and K. Stanley. 2011. Evolving a diversity of virtual creatures through novelty search and local competition. In *Proceedings of the 13th annual conference on Genetic and evolutionary computation.* 211–218.
[30] R. Lenski, C. Ofria, R. Pennock, and C. Adami. 2003. The evolutionary origin of complex features. *Nature* 423, 6936 (2003), 139–144.
[31] H. Levene. 1961. Robust tests for equality of variances. *Contributions to probability and statistics. Essays in honor of Harold Hotelling* (1961), 279–292.
[32] H. Mann and D. Whitney. 1947. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics* (1947), 50–60.
[33] T. Miconi. 2008. Evolution and complexity: The double-edged sword. *Artificial Life and Robotics* 14, 3 (2008), 325–344.
[34] J. Mouret. 2011. *Novelty-based multiobjectivization.* Springer, 139–154.
[35] J. Mouret and S. Doncieux. 2009. Overcoming the bootstrap problem in evolutionary robotics using behavioral diversity. In *2009 IEEE Congress on Evolutionary Computation.* IEEE, 1161–1168.
[36] V. Müller and M. Hoffmann. 2017. What is morphological computation? On how the body contributes to cognition and control. *Artificial life* 23, 1 (2017), 1–24.
[37] D. Nagar, A. Furman, and G. Nitschke. 2019. The cost of complexity in robot bodies. In *2019 IEEE Congress on Evolutionary Computation (CEC).* IEEE, 2713–2720.
[38] H. Orr. 2000. Adaptation and the cost of complexity. *Evolutionary computation* 54, 1 (2000), 13–20.
[39] T. Revello and R. McCartney. 2000. A cost term in an evolutionary robotics fitness function. In *Proceedings of the 2000 Congress on Evolutionary Computation. CEC00 (Cat. No. 00TH8512)*, Vol. 1. IEEE, 125–132 vol. 1.
[40] C. Shannon. 1948. A mathematical theory of communication. *Bell system technical journal* 27, 3 (1948), 379–423.
[41] S. Shapiro and M. Wilk. 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52, 3/4 (1965), 591–611.
[42] K. Sims. 1994. Evolving 3D morphology and behavior by competition. *Artificial Life and Robotics* 1, 4 (1994), 353–372.
[43] K. Sims. 1994. Evolving virtual creatures. In *Proceedings of the 21st annual conference on Computer graphics and interactive techniques.* 15–22.
[44] C. Sprong. 2011. Common tasks in Evolutionary Robotics, an overview. (2011).
[45] R. Standish. 2008. *Concept and definition of complexity.* IGI Global, 105–124.
[46] K. Stanley, D. D'Ambrosio, and J. Gauci. 2009. A hypercube-based encoding for evolving large-scale neural networks. *Artificial life* 15, 2 (2009), 185–212.
[47] K. Stanley and R. Miikkulainen. 2002. Evolving neural networks through augmenting topologies. *Evolutionary computation* 10, 2 (2002), 99–127.
[48] T. Taylor, M. Bedau, A. Channon, D. Ackley, W. Banzhaf, G. Beslon, E. Dolson, T. Froese, S. Hickinbotham, and T. Ikegami. 2016. Open-ended evolution: Perspectives from the OEE workshop in York. *Artificial life* 22, 3 (2016), 408–423.
[49] G. Tononi and O. Sporns. 2003. Measuring information integration. *BMC neuroscience* 4, 1 (2003), 31.
[50] G. Tononi, O. Sporns, and G. Edelman. 1994. A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proceedings of the National Academy of Sciences* 91, 11 (1994), 5033–5037. https://doi.org/10.1073/pnas.91.11.5033
[51] L. Whitley. 1991. *Fundamental principles of deception in genetic search.* Vol. 1. Elsevier, 221–241.
[52] L. Yaeger. 2009. How evolution guides complexity. *HFSP journal* 3, 5 (2009), 328–339.

8 SUPPLEMENTARY WORK

Table 6: Table showing the p-value produced by the Mann-Whitney U-test comparing the populations evolved in the baseline and complexity cost experiments for the ten repeated runs in each environment in terms of average fitness (task performance), average robot complexity, best fitness and the complexity of the most fit individuals. Cells are highlighted where the p-value<=0.05 indicating a statistical difference.

| Environment	Average Fitness	Average Robot Complexity	Best Fitness	Complexity Associated with Best Fitness
1	0.092938	0.260261	0.153745	0.192247
2	0.060612	0.311588	0.037831	0.454861
3	0.454861	0.395668	0.454861	0.236338
4	0.236338	0.172352	0.338792	0.213678
5	0.311588	0.338792	0.285375	0.236171
6	0.285375	0.032011	0.285375	0.006992
7	0.192337	0.012874	0.022577	0.020435
8	0.484925	0.285375	0.425053	0.425053
9	0.120661	0.236338	0.106147	0.425053
10	0.192337	0.285375	0.192337	0.454827
11	0.192337	0.080986	0.236338	0.425053
12	0.213678	0.260261	0.192337	0.051860
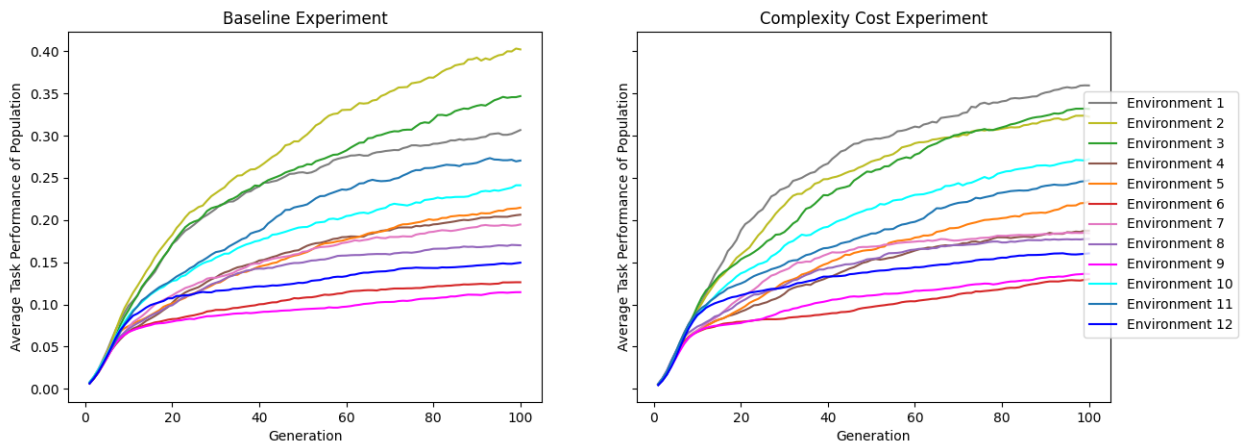
Figure 6: Graphs showing the average task performance of the population over all generations in all 12 environments for the baseline and complexity cost experiments
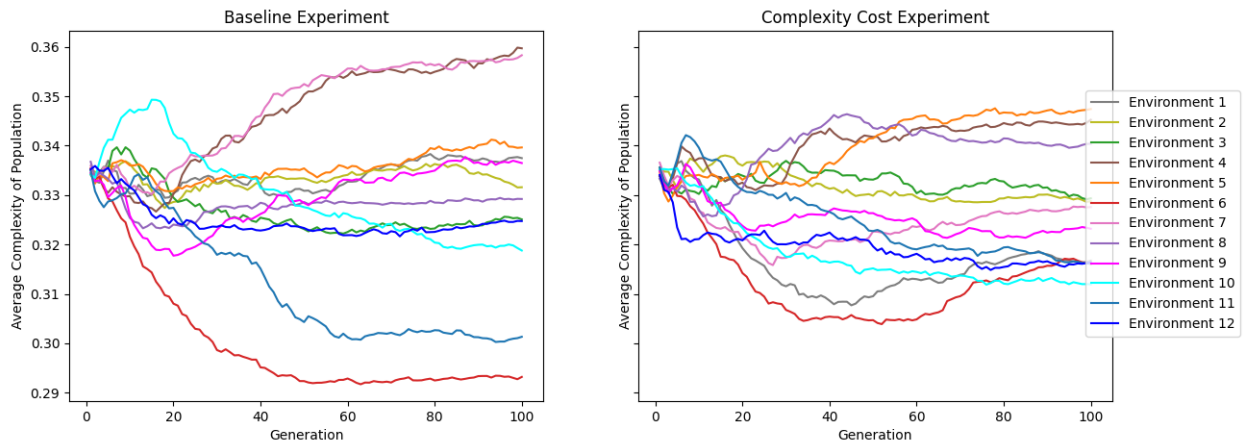
Figure 7: Graphs showing the average complexity of the population over all generations in all 12 environments for the baseline and complexity cost experiments
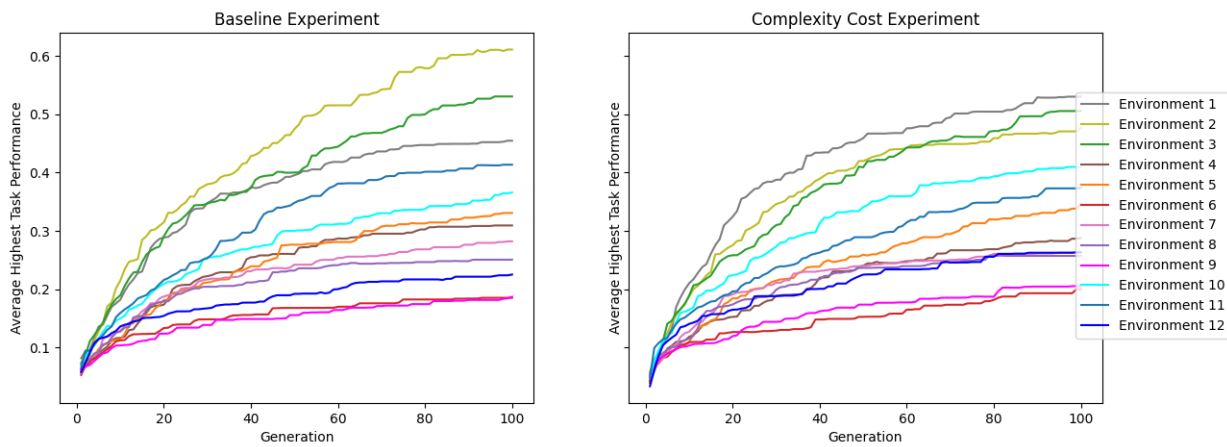

Figure 8: Graphs showing the average maximum task performance of the population over all generations in all 12 environments over all 10 runs for the baseline and complexity cost experiments
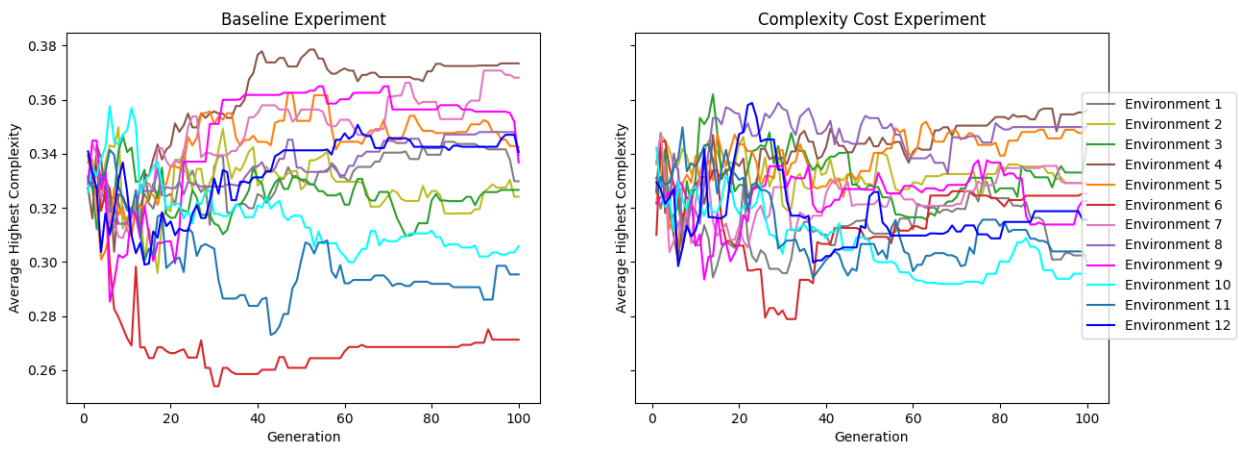
Figure 9: Graphs showing the average complexity associated with the fittest individual of the population over all generations in all 12 environments over all 10 runs for the baseline and complexity cost experiments