Evolving Diverse Strategies Through Combined Phenotypic Novelty and Objective Function Search

Davy Smith[1], Laurissa Tokarchuk[1], and Chrisantha Fernando[1]

Queen Mary University of London, Department of Computer Science and Electronic Engineering, Mile End Road, London, E1 4NS
`david.smith@qmul.ac.uk`

Abstract. Novelty search is an algorithm which proposes open-ended exploration of the search space by maximising behavioural novelty, removing the need for an objective fitness function. However, we show that when applied to complex tasks, training through novelty alone is not sufficient to produce *useful* controllers. Alongside this, the definition of phenotypic behaviour significantly effects the strategies of the evolved solutions. Controller networks for the spaceship in the arcade game *Asteroids* were evolved with five different phenotypic distance measures. Each of these phenotypic measures are shown to produce controllers which adopt different strategies of play than controllers trained through standard objective fitness. Combined phenotypic novelty and objective fitness is also shown to produce differing strategies within the same evolutionary run. Our results demonstrate that for domains such as video games, where a diverse range of interesting behaviours are required, training agents through a combination of phenotypic novelty and objective fitness is a viable method.

1 Introduction

The training of agents through the promotion of diverse behavioural characteristics is a recent area within Evolutionary Robotics (ER) that suggests promising directions towards open-ended evolution [5,17,3,13]. In complex tasks, and when faced with uncertainty, a range of differing behavioural strategies emerge. The behavioural sciences observe in nature, not only a vast array of different species, but also a scale of behavioural differences within the same species [14,24].

Varying both the phenotypic definition and the distance metric used in the fitness assessment of an evolutionary task has been shown to produce widely different strategies amongst the population, suggesting the importance of domain specificity and the assessment metric in measuring phenotypic distance [17,3]. However, there has been little research which addresses a series of behavioural measures in order to highlight the effect that particular phenotypic definitions play on both objective fitness and the resulting useful behavioural strategies of the agents in the same domain. A series of experiments were undertaken to assess the effect of varying phenotypic definitions on objective performance at an uncertain task. Agent controllers for the spaceship in the video game *Asteroids* were evolved on the same task, with fitness assessments based on a linear proportion of objective fitness and phenotypic novelty over five definitions of behaviour.

A linearly mixed fitness assessment is shown to produce controllers which adopt different strategies of play than controllers trained through standard objective fitness without significant effect on objective performance. A comparison of the complete state-action pairings of high scoring agents was undertaken to assess the diversity of the evolved solutions. Our results show that, dependant on an ideal mixing ratio, the linear combination of objective fitness and phenotypic novelty produces highly diverse solution populations.

2 Related Work

2.1 Novelty Search

Novelty search, as proposed by Lehman and Stanley [11,12], is an algorithm which removes the need for an objective fitness function through the assignment of high fitness values to novel behaviours in a population.

$$n(x) = \frac{1}{k}\sum_{i=0}^{k} dist(x, \mu_i) \tag{1}$$

The behavioural novelty $n(x)$ of an individual x is defined as the mean behavioural distance between x and its k nearest neighbours, where k is a user defined parameter and μ_i is the ith nearest neighbor of x with respect to the distance $dist$. The value of k includes both the behaviours of the current population and an archive of previous novel behaviours. Individuals with a value of $n(x)$ above a predefined novelty threshold are added to the archive, Equation (1).

2.2 Phenotypic Diversity

It has been suggested that the success of ER to extend beyond simple tasks will be dependant upon adaptable and open-ended evolutionary procedures [20]. Current definitions exist within the literature which draw separations between *genotypic space* as the binary representation of genes in the population, *phenotypic space* as the topology of the networks produced by the genotype and *behaviour space* as the actions produced by the agent [17]. Here, however, we adopt a more biologically informed definition, in which the phenotype refers to any observable characteristics of an organism, which may include both network topology and the related behaviour of an agent [10].

The design of effective fitness functions in ER is a subjective process, therefore susceptible to human error and deception. If generalisable across domains, the promotion of phenotypic diversity may alleviate this limitation [11]. However, assessment methods based upon phenotypic diversity, which require a definition of the particular behaviour to encourage or diversify in a given domain, strictly translate rather than remove the human design process. Experiments which have shown novelty search to outperform objective fitness have concentrated on domain specific behavioural metrics, such as maximising the distance of the end navigation points of a robot within a maze [11]. Metrics have been proposed to generalise phenotypic novelty, for example through measuring the distance of output values from randomised input vectors given to the controller [6]. In a comparison of generic behaviour based on motor actions, Gomez suggested measuring a Normailized Compression Distance (NCD) of binary action vectors as yielding the most promising method for translation to different domains [6].

Although not strictly a diversity maintenance algorithm, novelty search receives increasing interest within ER research for its unique way in expanding the search space to multiple solutions in any given domain [11]. The introduction of novelty search uncovered many of the stepping stones towards open-ended evolution. Although novelty search may outperform objective fitness search in specific tasks, especially when the design of an objective fitness function may be difficult, it has been shown that the assessment of behavioural novelty alone is insufficient as a generalisable evolutionary technique in many domains [2,16]. Due to its divergent nature, novelty search continues to produce new solutions throughout the evolution, however these solutions may not be useful for the task at hand. The combination of novelty search and objective fitness, in which the diverse and expanding search space explored by novelty search is limited to *useful* solutions is a promising direction for the application of the algorithm.

Alongside this, multiobjective optimisation of both novelty and objective fitness has shown to outperform purely objective fitness in biped locomotion tasks and maze navigation [13].

2.3 Neuro-Evolution Through Augmenting Topologies (NEAT)

Neuro-Evolution Through Augmenting Topologies (NEAT), developed by Stanley and Miikkulainen in 2002 [22], is an Evolutionary Algorithm (EA) for the evolution of Artifical Neural Networks (ANN). In addition to the mutation of weights between neurons, NEAT evolves the networks' topologies, creating phenotypically diverse populations. Increasing complexity is achieved by initialising a population of networks with minimal topologies and adding genes as the evolution progesses, leading to more diverse behaviour patterns. Neurons added to the network may be either feed-foward or recurrent, allowing for the emergence of a short term memory within networks. Additionally, in order to protect new innovations, historical markings are used to assign species to the population. Alongside evolutionary robotics tasks, NEAT has been widely applied to the evolution of video game agents in multiple domains [21,7,9,1].

2.4 Video Games for Evolutionary Research

Classic 2D video games are generally played within basic grid type worlds, with a limited number of agents and a small set of available actions. Although deceptively simple, the dynamics within these worlds are often complex and uncertain, requiring multiple diverse strategies and planning procedures for play. Unlike traditional maze navigation or pole balancing tasks, classic video games provide simulated worlds which require the acquisition of multiple skills, such as path finding, fights for survival, evasion, goals and sub-goals. Due to these rich worlds, there is a growing body of research in the application of EAs to classic video games, both for agents evolved for particular games, for example *Ms-Pacman* [23], and general games playing agents [9,8,15]. Typically within the literature, fitness assessment of the population is achieved through objective points scoring. However there are a number of games, including *Montezuma's Revenge* and *Pitfall* where points are not readily available. It has been recently suggested that an intrinsically motivated assessment of the diversity of agents' behaviours is a promising approach to allow evolution to continue in the absence of points [9].

3 Experimental Domain

A bespoke version of the classic arcade game *Asteroids* was used as an experimental domain (Figure 1). The aim of *Asteroids* is to score as many points as possible by shooting asteroids and avoiding collisions. The player controls a spaceship that has a left and right rotation, a forward thrust and the ability to shoot. Due to *Asteroids* being set in deep space, the spaceship is not effected by friction or gravity, therefore takes a long time to slow down after thrust is applied. As the name suggests, the enemies in the game are asteroids, which appear on the screen in waves, with random velocity and direction. Asteroids appear in three different sizes: starting off as large, and when hit splitting into two medium size asteroids, which in turn each split into a further two small asteroids. The first wave consists of three large asteroids. After the player has cleared all of the asteroids, the next wave begins, with one more large asteroid than the last. The playing field in *Asteroids* is constructed as a toroidal space, i.e. if asteroids or the ship move off the edge of the screen, they reappear on the opposite side.

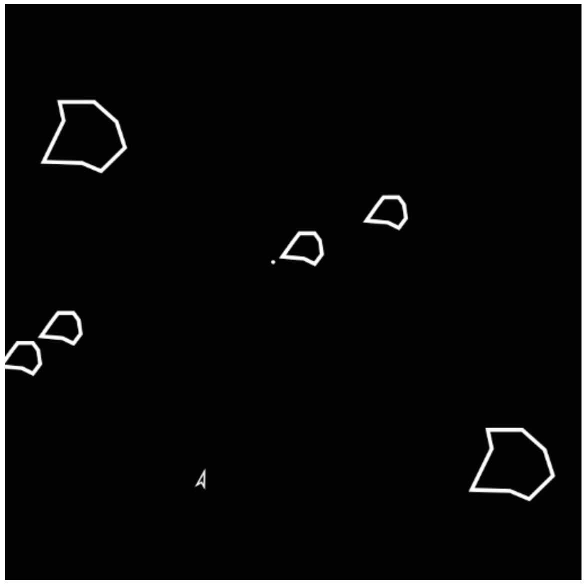

Fig. 1: Screenshot of the *Asteroids* video game

4 Agent Model

4.1 Perceptual Field

The perceptual field for the agent was constructed as a dartboard-style map with binary inputs centred on its position and rotation, providing a discrete representation of polar coordinates relative to the agent (Figure 2). Inputs to the map were assigned a value of zero if no asteroids were present within the related coordinate, and one if any number of asteroids appeared within the bounds. A series of trial experiments were conducted using a range of input maps with differing resolutions and sizes. The final perceptual field used consisted of 4 segments and 3 slices and a diameter of 0.8 of the world's length. In order to allow the agent the capacity to adapt to the toroidal nature of the playing space, it was also decided to allow the agents' perception to extend beyond the edges of the screen, overlapping to the opposite side (Figure 2).

4.2 Controller Network

The NEAT algorithm was used to train the agents' controllers throughout this experiment. The dartboard state map was passed to the inputs of the network as a 12 dimensional binary array (with a value of one if at least one asteroid was detected in the position and zero otherwise) at each time step in the game. The networks were assigned three floating point outputs used to control left and right rotation and thrust, and one binary output for shooting. The NEAT

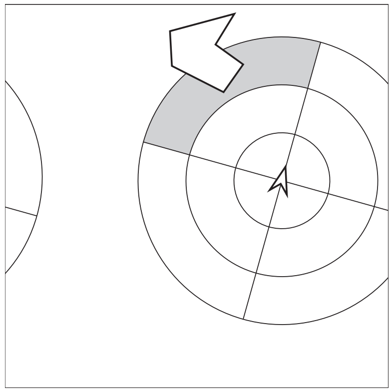

Fig. 2: Agent perceptual state map

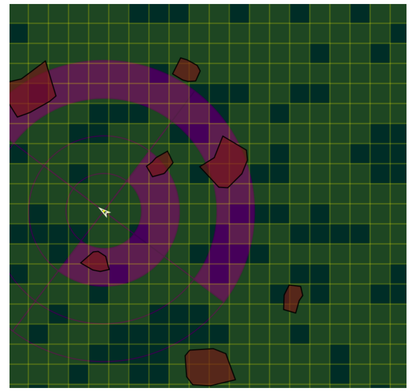

Fig. 3: **GC** phenotypic measurement

algorithm's parameters were set to enable the evolution of recurrent nodes within the networks, allowing for the possibility of a short term memory to develop within the controllers.

5 Phenotypic Definitions

In the presented experiment, in order to determine the possibilities for a generalised method of evolution based on behavioural novelty, we apply a combination of both generic and domain specific behaviour measures. Although [6] suggests a metric using NCD as being optimal, due to the computational complexity of both NCD, which requires a compression to be calculated with each distance measurement, and novelty search, which introduces its own computational load due to the maintenance of an expanding archive, we forgo the NCD measurement, comparing instead the Hamming distance of action vectors, which has previously produced comparably similar results [6,17].

The following five differing phenotypic definitions were evaluated to establish the significance of a fitness based evolution based on behavioural distance.

1. **(AC) Action Count** Due to the complexity of evaluating the behaviours of the agents during online gameplay, an evaluation procedure was devised to present each of the population with a randomly generated set of hypothetical game states. A set of one hundred states was randomly produced at the beginning of the evolutionary run and the set presented to the controllers before gameplay began in each generation. The euclidean distances between each of the resulting arrays of output vectors were then measured

to establish the novelty value of the agents[1].

2. **(GC) Ground Covered** The play area was divided into a 20x20 matrix and a counter was incremented the first time each time the agent was positioned within a square. The 1 dimensional euclidean distance was taken as the measurement between agents in the population (Figure 3).

3. **(MAD) Mean Asteroid Distance** The average distance from the agent to the asteroids in the playing field was taken at each time step during the game. The mean of these distances was then measured between agents.

4. **(MTR) Mean Thrust and Rotation** The average thrust and rotation of the agent was taken throughout game play. The euclidean distance of the resulting two dimensional vector was measured.

5. **(NA) N Actions** Each game was initiated with the game state, i.e. identical asteroid positions and velocities. The first N actions was stored as a string (L for left, R for Right etc) and the hamming distances between the strings taken (in this experiment $N = 100$).

6 Evolutionary Criteria

6.1 Experimental Parameters

Although relatively simple for a human player to grasp, *Asteroids* is a particularly difficult domain for evolutionary techniques. The agents' actions in the game directly effect the trajectory of the task, therefore introducing a level of uncertainty. The same controller can receive a wide range of scores in different games, subsequently effecting fitness measures. In order to reduce this level of noise, 10 games were played by each controller in each generation of the evolutionary phase, and the objective fitness averaged. Due to the computationally exhaustive nature of novelty search over multiple assessments, the behavioural values were evaluated on one random game per generation.

Agents were allocated three lives in each round of the game. To remove the possibility of an agent discovering a linear trajectory which avoids collisions with all asteroids, therefore making the game last infinitely long, a timer was added to the task, requiring agents to hit an asteroid every 1000 update loops, or approximately 15 seconds. Evolutionary runs were restricted to 1000 generations throughout, with the winning average scores over 40 evolutionary runs presented

[1] It may be noted that with this particular phenotypic definition, the behaviour is measured hypothetically, before the task has been performed. Therefore when evaluating through novelty search alone ($\lambda = 1.0$), the game does not need to even be played.

in Section 7.

6.2 Fitness Assessment

In each of the experiments the fitness of an individual x, where $x \in P$ was determined as a linear combination of behavioural novelty and points scored, with the ratio of each dependant upon a multiplier, λ $[0, 1]$. The values assigned for behavioural novelty x_b and points scored x_s were normalised $n(x_i)$ against the maximum and minimum scoring individuals in P, to assign a final fitness value $f(x)[0, 1]$ (Equations 2, 3).

$$n(x_i) = \frac{x_i \ - \ min(P_i)}{max(P_i) \ - \ min(P_i)} \tag{2}$$

$$f(x) = \lambda(n(x_b)) \ + \ (1 - \lambda)(n(x_s)) \tag{3}$$

7 Findings

7.1 Objective Performance

Figure 4 outlines the maximum points achieved by an agent for the average scores over 40 evolutionary runs of 1000 generations for varying mixing ratios of objective fitness and novelty search (λ $[0 : 1]$). In each generation of the experiments, the average score over 10 games was assessed. Our results show that phenotypic definition directly effects the ability of the novelty search algorithm. Although outperforming random search, pure novelty search ($\lambda = 1$) performs significantly sub optimally over all tested behaviours compared to objective fitness search.

The linear combination of novelty search and objective fitness increases the performance of the agents compared with novelty search ($\lambda < 1$). The results show no significant improvement in objective fitness with the addition of novelty, however, of the five tested behaviours, AC, MAD, GC and MTR produced median fitnesses which outperformed the median objective fitness for $\lambda = \frac{1}{6}$, with only NA under performing. Of all tested behaviours, MTR with a mixing ratio of $\lambda = \frac{1}{6}$ produced the most successful results. The ideal mixing ratio of novelty search and objective fitness remained relatively consistent throughout the experiments. A small ratio of novelty search to objective fitness ($\lambda = \frac{1}{6}$) produced the highest results for GC, MTR, MAD and NA. AC, however, produced comparable results for both $\lambda = \frac{1}{6}$ and $\lambda = \frac{2}{6}$.

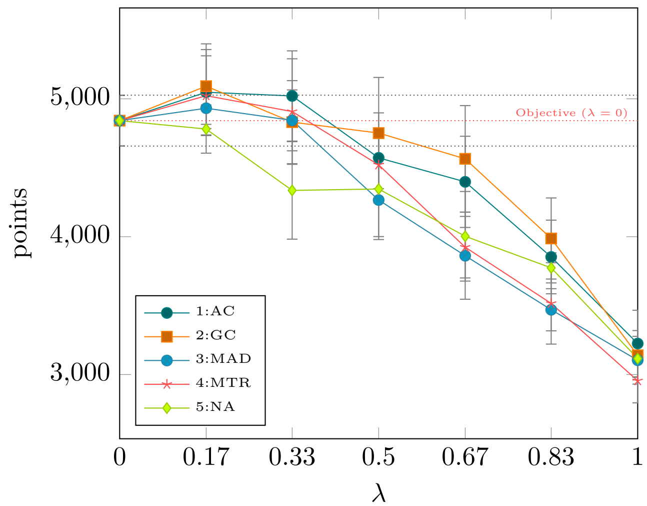

Fig. 4: Mean points scored using a linear combination of objective fitness and novelty search. (Error bars represent standard deviation)

7.2 State-Action Distance between Agents

Due to the tested domain being an interactive video game, good progress requires the agent to constantly alter the trajectory of play (i.e. by shooting asteroids). This makes the assessment of play strategies a difficult task. It was therefore decided to compare the full state-action parings for agent controllers to indicate the distance of actions between agents.

In order to assess the diversity of high performing strategies produced within a single evolutionary run, the four highest scoring individuals were stored over the course of one training cycle of 1000 generations for each phenotypic type, using the optimally combined phenotypic-objective ratio ($\lambda = \frac{1}{6}$) alongside a separate, purely objective fitness run ($\lambda = 0$).

The actions performed for all combinations of the input state space were compared for these high scoring agents in order to establish the diversity of controller networks produced by the addition of phenotypic novelty. The state input map used in the trials consisted of a 5×4 two dimensional binary input matrix, giving $2^{20} = 1048576$ possible combinations. The resulting actions were converted into action strings (e.g. "ULS"$= \{up, left, shoot\}$) and the Hamming distances for each state-action pair between agents within the same phenotypic definitions were compared as a percentage (Table 1).

All Agents across all of the phenotypic definitions (Tables 1a to 1e), with the exception of agents one and three in the MTR phenotypic type (Table 1a), produce state-action parings with equal or higher distances than the objectively trained agents (Table 1f). This strongly indicates a more diverse set of actions for input states within singular evolutionary runs.

Table 1: Hamming Distance of State-Action Mappings

(a) $\frac{1}{6}\lambda$ GC

	1	2	3	4
1	0	80	73	55
2	80	0	68	80
3	73	68	0	77
4	55	80	77	0

(b) $\frac{1}{6}\lambda$ AC

	1	2	3	4
1	0	75	66	76
2	75	0	75	59
3	66	75	0	76
4	76	59	76	0

(c) $\frac{1}{6}\lambda$ MAD

	1	2	3	4
1	0	71	76	66
2	71	0	94	62
3	76	94	0	75
4	66	62	75	0

(d) $\frac{1}{6}\lambda$ MTR

	1	2	3	4
1	0	100	9	61
2	100	0	99	84
3	9	99	0	61
4	61	84	61	0

(e) $\frac{1}{6}\lambda$ NA

	1	2	3	4
1	0	84	78	95
2	84	0	79	94
3	78	79	0	69
4	95	94	69	0

(f) $\lambda = 0$

	1	2	3	4
1	0	38	52	36
2	38	0	48	55
3	52	48	0	49
4	36	55	49	0

8 Limitations

Due to the nature of the tested domain, in which agents have the ability to alter the environment and thus the trajectory of play, assessing the diversity of strategies produced by agents is a difficult task. In this experiment, we chose to highlight the distance between full state space to action mappings. Although the Hamming distance of possible action strings may be applicable for domains with a small, or unalterable state-space, e.g. a board game, it may not be sufficient here to fully capture differing strategies of behaviour. Because of the highly uncertain and changeable nature of the domain, methods developed in the behavioural sciences may be more applicable [14]. Assessment criteria could be established, for example, through the qualitative human assessment of videos of the agents playing the game.

9 Conclusion

As shown in Section 7, training through a combination of objective function and novelty search is a viable method to produce controllers which are not only high scoring, but also adopt multiple strategies. Alongside this, varying the particular phenotypic definition or metric used may further increase the diversity of strategies adopted. Our results indicate that training through a linear combination of objective fitness and novelty search with multiple phenotypic definitions is a viable methods to produce a range of useful controllers which adopt a diverse range of strategies.

Sandbox [19] and open-world games [18] are recent video game genres which promote exploratory and intrinsically motivated forms of play rather than fixed objectives. Investigations could be undertaken to assess the applicability of training non-player character (NPC) behaviours in such games, using combined objective and phenotypic novelty, where a diverse range of unpredictable and unique behavioural characteristics are required.

Although some of the possible sets of strategies which emerge through novelty search are not directly useful to the domain at hand, it does not follow that these strategies are without use for all domains. An interesting direction to further extend studies analysing the diversity of strategies produced through combined objective and phenotypic search, could assess the transferability of trained agents or agent populations into either different domains, or domains which alter over time.

Our results indicate that evolution through the combination of objective and novelty search produces an extensive set of useful diverse strategies that have potential application in two main areas; firstly, to both provide new and interesting non-player characters (NPCs) and secondly, to further progress towards creating transferable agents for general game playing.

10 Acknowledgements

We would like to thank the reviewers for their insightful comments and suggestions.

This work was funded by EPSRC through the Media and Arts Technology Programme, an RCUK Doctoral Training Centre EP/G03723X/1. Computational facilities were provided by the MidPlus Regional Centre of Excellence for Computational Science, Engineering and Mathematics, under EPSRC grant EP/K000128/1

References

1. L. Cardamone, D. Loiacono, and P. L. Lanzi. Evolving competitive car controllers for racing games with neuroevolution. In *Proceedings of the 11th Annual conference on Genetic and evolutionary computation*, pages 1179–1186. ACM, 2009.

2. G. Cuccu and F. Gomez. When novelty is not enough. In *Applications of Evolutionary Computation*, pages 234–243. Springer, 2011.
3. A. Cully and J.-B. Mouret. Behavioral repertoire learning in robotics. In *Proceeding of the fifteenth annual conference on Genetic and evolutionary computation conference*, pages 175–182. ACM, 2013.
4. K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, and A. Fast. Nsga-ii. *IEEE transactions on Evolutionary Computation*, 6(2):182–197, 2002.
5. S. Doncieux, J.-B. Mouret, N. Bredeche, and V. Padois. Evolutionary robotics: Exploring new horizons. In S. Doncieux, N. Bredèche, and J.-B. Mouret, editors, *New Horizons in Evolutionary Robotics*, volume 341 of *Studies in Computational Intelligence*, pages 3–25. Springer Berlin Heidelberg, 2011.
6. F. J. Gomez. Sustaining diversity using behavioral information distance. In *Proceedings of the 11th Annual conference on Genetic and evolutionary computation*, pages 113–120. ACM, 2009.
7. E. J. Hastings, R. K. Guha, and K. O. Stanley. Evolving content in the galactic arms race video game. In *Computational Intelligence and Games, 2009. CIG 2009. IEEE Symposium on*, pages 241–248. IEEE, 2009.
8. M. Hausknecht, P. Khandelwal, R. Miikkulainen, and P. Stone. Hyperneat-ggp: A hyperneat-based atari general game player. In *Proceedings of the fourteenth international conference on Genetic and evolutionary computation conference*, pages 217–224. ACM, 2012.
9. M. Hausknecht, J. Lehman, R. Miikkulainen, and P. Stone. A neuroevolution approach to general atari game playing. 2013.
10. W. Johannsen. The genotype conception of heredity. *The American Naturalist*, 45(531):129–159, 1911.
11. J. Lehman and K. O. Stanley. Exploiting open-endedness to solve problems through the search for novelty. In *ALIFE*, pages 329–336, 2008.
12. J. Lehman and K. O. Stanley. Abandoning objectives: Evolution through the search for novelty alone. *Evolutionary computation*, 19(2):189–223, 2011.
13. J. Lehman, K. O. Stanley, and R. Miikkulainen. Effective diversity maintenance in deceptive domains. 2013.
14. P. R. Martin. *Measuring behaviour: an introductory guide*. Cambridge University Press, 1993.
15. V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*, 2013.
16. J.-B. Mouret. Novelty-based multiobjectivization. In *New Horizons in Evolutionary Robotics*, pages 139–154. Springer, 2011.
17. J.-B. Mouret and S. Doncieux. Encouraging behavioral diversity in evolutionary robotics: An empirical study. *Evolutionary computation*, 20(1):91–133, 2012.
18. S. Murray, D. Ream, and R. Doyle. No man's sky, 2015.
19. M. Persson. Minecraft, 2009.
20. K. Ruiz-Mirazo, J. Peretó, and A. Moreno. A universal definition of life: autonomy and open-ended evolution. *Origins of Life and Evolution of the Biosphere*, 34(3):323–346, 2004.
21. K. O. Stanley, B. D. Bryant, and R. Miikkulainen. Real-time neuroevolution in the nero video game. *Evolutionary Computation, IEEE Transactions on*, 9(6):653–668, 2005.
22. K. O. Stanley and R. Miikkulainen. Evolving neural networks through augmenting topologies. *Evolutionary Computation*, 10(2):99–127, 2002.

23. R. Thawonmas and T. Ashida. Evolution strategy for optimizing parameters in ms pac-man controller ice pambush 3. In *Computational Intelligence and Games (CIG), 2010 IEEE Symposium on*, pages 235–240. IEEE, 2010.
24. N. Tinbergen. On aims and methods of ethology. *Zeitschrift für Tierpsychologie*, 20(4):410–433, 1963.