The effects of morphology and fitness on catastrophic interference

Joshua Powers, Sam Kriegman, and Josh Bongard

Morphology, Evolution & Cognition Lab
University of Vermont, Burlington, VT, USA
jpowers4@uvm.edu

Abstract

Catastrophic interference occurs when an agent improves in one training instance but becomes worse in other instances. Many methods intended to combat interference have been reported in the literature that modulate properties of a neural controller, such as synaptic plasticity or modularity. Here, we demonstrate that adjustments to the body of the agent, or the way its performance is measured, can also reduce catastrophic interference without requiring changes to the controller. Additionally, we introduce new metrics to quantify catastrophic interference. We do not show that our approach outperforms others on benchmark tests. Instead, by more precisely measuring interactions between morphology, fitness, and interference, we demonstrate that embodiment is an important aspect of this problem. Furthermore, considerations into morphology and fitness can combine with, rather than compete with, existing methods for combating catastrophic interference.

Introduction

Currently, a popular method for realizing intelligent machines is to optimize the parameters of fixed-architecture deep neural networks (LeCun et al., 2015). However, increasing interest is coming to bear on optimizing the cognitive architecture of such networks as well (Miikkulainen et al., 2017). Indeed, investigations into the evolution of cognitive architectures has long been a target of study (Gruau and Whitley, 1993; Bongard and Pfeifer, 2001; Stanley and Miikkulainen, 2002) in the evolutionary computation community.

It follows from this, if dealing with robots, that optimizing body plan influences sensory repercussions of action, which in turn will influence which cognitive architecture will facilitate learning for a given task. To begin investigations into this last observation, here we investigate how the choice of robot morphology and fitness affect one specific aspect of neural networks: their ability to resist catastrophic interference. We employ an evolutionary robotics approach to investigate this question.

Evolutionary robotics.

Since its beginnings, many investigators in the field of evolutionary robotics (Floreano and Mondada, 1994; Harvey et al., 1997; Bongard, 2013) have used evolutionary algorithms to optimize both the body plan and neural controllers of robots (Sims, 1994; Lipson and Pollack, 2000; Cheney et al., 2013). Here we show that indeed the choice of body plan can influence the efficacy of training neural controllers: some body plans enable greater resistance to catastrophic interference.

Catastrophic interference.

It has been acknowledged since the early days of neural network research that catastrophic interference (McCloskey and Cohen, 1989), also known as catastrophic forgetting (French, 1999; Goodfellow et al., 2013), is a major challenge to training them effectively. Even in the most common forms of network training such as the backpropagation of error, there is no guarantee that reducing the network's error on the current training sample does not increase error on the other training samples.

For these reasons, much effort has been expended to combat this challenge. One family of solutions involves constructing modular networks (Lipson et al., 2002; Ellefsen et al., 2015; Clune et al., 2013; Espinosa-Soto and Wagner, 2010; Kashtan and Alon, 2005; Sabour et al., 2017) in which different modules deal with different subsets of the training set. In such networks, changes to one module may result in improved performance for the training subset associated with that module without disrupting performance on other subsets. Such modularity has indeed been demonstrated to minimize catastrophic interference (Ellefsen et al., 2014; Rusu et al., 2016; Lee et al., 2016; Fernando et al., 2017).

Related to this concept of modularity are networks in which some subsets of the network that have a large impact on the current training set are made less resistant to change during subsequent training (Kirkpatrick et al., 2017; Velez and Clune, 2017). The remaining parts of the network remain adaptive, and thus able to deal with new training in-

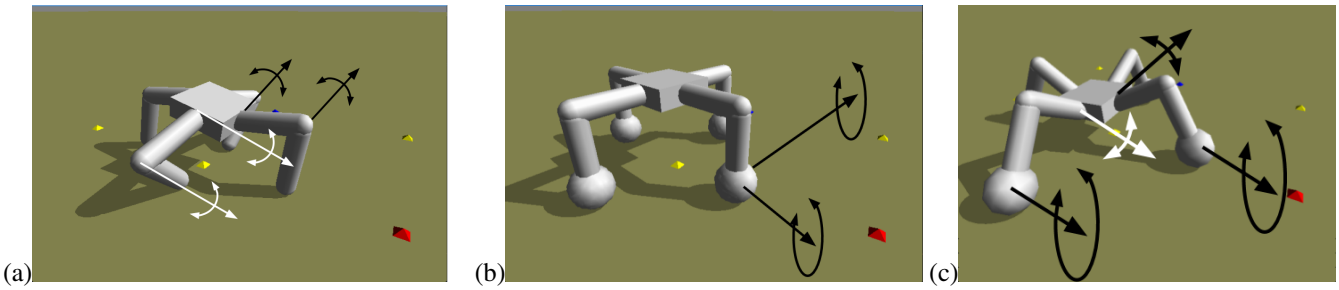

Figure 1: Three classes of phototaxic robots—legged (a), wheeled (b), and whegged (c)—and their environments were simulated using Pyrosim (ccappelle.github.io/pyrosim). Each robot has eight degrees of freedom, as depicted by the black and white arrows which indicate the axis (straight) and direction (curved) of rotation for a particular hinge-joint (a, c) or wheel (b, c). Video of all three robot types can be seen at youtu.be/yY7Vi7fw7Ik.

stances without disrupting behavior on previous instances.

The drawback of these approaches however is that network size tends to increase with the amount of training data, because new modules must be implicitly or explicitly added for new training data.

Another guard against catastrophic interference is to reduce the magnitude of behavioral impact after some change is made during training. The intuition here is that small changes to network behavior may increase the likelihood of local improvements for new training instances while minimizing or nullifying performance decreases on the previous training set.

In evolutionary methods, one way of reducing behavioral impacts is to dynamically tune mutation rates (Dang and Lehre, 2016) and/or crossover events (Teo et al., 2016). A recent approach demonstrated for neuroevolution is to dynamically tune individual synaptic weights proportionally to their impact on the network's behavior (Lehman et al., 2017). In the genetic programming community, semantic variation operators have been reported (Vanneschi et al., 2014; Szubert et al., 2016). These operators take into account the semantics of subtrees or individual tree nodes, and attempt to replace them with new genetic material that exhibits similar semantics.

In this work we show that the body plan itself as well as the manner in which fitness is defined can buffer the behavioral impact of mutations such that the embodied agent's behavior can improve in one environment without adversely impacting its performance in another environment in which it is already proficient without using any methods for specifically preventing interference in the controller.

Embodied impacts on neural properties.

Besides catastrophic interference, it has been shown elsewhere that embodiment can influence the positive or negative aspects of neural networks. For instance, work in morphological computation has shown that a good choice of morphology can allow for simplified neural networks (e.g. Hauser et al. (2011)).

Morphology may also render a robot more robust to external environmental perturbation (Bongard, 2011) or internal changes to the neural controller (Kriegman et al., 2017). Some body plans also facilitate or obstruct the discovery of modular neural networks (Bongard et al., 2015; Bernatskiy and Bongard, 2017). In this work we introduce a heretofore unexplored aspect of the interaction between body plan and neural control of embodied agents: how the choice of body plan may render the neural controller more or less resistant to catastrophic interference.

The next section describes our methodology for demonstrating this phenomenon; the following sections provide results from evolving neural controllers in different body plans; the final sections provide some discussion as to how and why this phenomenon arises and concludes with avenues for future study.

Methods

The robots.

Three types of robots were used in this study (Fig. 1). All are variations on a standard radially symmetric quadrupedal form used in other evolutionary robotics studies (Bongard et al., 2006; Lohmann et al., 2012; Belter et al., 2015). The robots differ in their differential use of legs and wheels.[1] Combining wheels and legs in different ways is itself an active area of study in robotics (Schroer et al., 2004; Jehanno et al., 2014; Kim et al., 2014).

The legged robot This robot consists of a body and four legs attached to the body by a joint. Each joint rotates 180° through the plane defined by the two cylinders comprising that leg. Each leg consists of two limbs bent at 90° attached by a joint which also rotates 180° through the plane defined by the upper and lower legs (Fig. 1a). Each of the resulting eight joints are actuated using position control: a value arriving from the motor neuron attached to the joint is treated

[1] github.com/jpp46/ALIFE2018 contains the source code necessary for reproducing the results reported in this paper.

as a desired angle. The length of each upper leg, lower leg, and the two sides of the main body are 0.3 units of length long (the physics engine is agnostic to the length scale).

The wheeled robot This robot freezes the previously mentioned joints at their initial angles and cannot change these positions. In addition, each leg has a wheel attached on the end that rotates through the plane defined by its upper and lower leg components. These wheels are attached to the legs by axles which rotate through the sagittal plane (Fig. 1b). This allows the wheels to change two directions of their rotation, acting as caster wheels. Values received at each of the two motors controlling the wheel treat the incoming values as desired angular velocity. In the spirit of keeping the robots as similar as possible, the motor neurons innervate the wheels output as desired angular velocities in $[-90°/s, +90°/s]$ to match the legged robot where motor neurons dictate desired angles in $[-90°, +90°]$.

The whegged robot This robot combines features of the previous two robots. In this robot, only the joints connecting the two leg limbs are frozen and can not change from their initial $90°$ angle. The joints joining the legs to the body are the same as the first robot: position controlled, actuated one-degree-of-freedom rotational joints. The robot also has wheels on the legs, but they do not act as caster wheels (Fig. 1c). This robot combines position control at the four shoulder joints and velocity control at the four wheels.

The sensors. Each robot contains a light sensor ℓ that responds to light as a float value according to the inverse square law for light propagation: $\ell = 1/d^2$, where d is the distance from a light source. Occlusion is not simulated in the light sensor: if an object is between the sensor and the light, there is no change in sensor value. Each robot contains a single binary touch sensor in each leg. These four sensors read $+1$ when in contact with the ground, and -1 otherwise.

The controllers.

The controller for each robot is a neural net with 5 input neurons, one for each sensor, and 8 output neurons, one for each motor. The neural net has no hidden layers and is fully connected. The connection weights are captured by a 5×8 matrix, which also represents the genome of a robot. The weights of the matrix are constrained to values in the interval $[-1, 1]$. The update function for a neuron during simulation follows this function:

$$m_i^{(t)} = \tanh\left[m_i^{(t-1)} + \tau_i \sum_{s=1}^{5} w_{s_i} m_s^{(t)}\right] \tag{1}$$

where $m_i^{(t)}$ denotes the value of the i^{th} motor neuron at the current time step, $m_i^{(t-1)}$ is a momentum term that guards against 'jitter' (high-speed and continuous reversals in the angular velocity of a joint), τ_i is a time constant that can strengthen or weaken the influence of sensation on the i^{th} motor neuron relative to its momentum, and w_{s_i} is the weight of the synapse connecting the s^{th} sensor neuron to the i^{th} motor neuron. In order to ensure that random controllers produce diverse yet not overly-energetic motion, all τ_i were set to 0.3 via empirical investigation.

The task environment.

Robots are evolved to perform phototaxis: minimizing the distance between themselves and a light source in their environment. Each robot is exposed to all training environments. Each environment consists of a light source that is 30 body lengths away from the robot (equivalent to nine units of distance). The robot starts at coordinate space $(0, 0)$ and the light sources are placed at coordinate space $(0, 9)$, $(0, -9)$, $(9, 0)$, $(-9, 0)$ for environments 1, 2, 3, and 4 respectively. The robots are evaluated for 1000 time steps in each environment using a fixed time step of 0.05 seconds.

The fitness functions.

The fitness function applied to a single environment is the value of the robot's light sensor at the end of an evaluation period ($t = 1000$). We combine fitness values drawn from multiple environments in three different ways:

$$F_\Sigma \doteq \sum_{i=1}^{n} f_i, \qquad F_\Pi \doteq \prod_{i=1}^{n} f_i, \qquad F_{\min} \doteq \min_{i=1}^{n} f_i. \tag{2}$$

where f_i is the individual's fitness in environment $i \in (1, n)$.

The evolutionary algorithm.

We employed a simple parallel hill climber to nine different experimental treatments as shown:

	Legged	**Wheeled**	**Whegged**
$\sum$	Treatment 1	Treatment 2	Treatment 3
$\prod$	Treatment 4	Treatment 5	Treatment 6
min	Treatment 7	Treatment 8	Treatment 9

For each treatment, 30 independent evolutionary runs were conducted each with a population size of $S = 100$. Each individual p was encoded as a 5×8 matrix of synaptic weights. Thus population P is represented by a $100 \times 5 \times 8$ tensor. We used a mutation strength of $m = 0.05$, and conducted each run for $G = 3000$ generations. At each generation g we took the current population P_g and generated a new population $P_{(g+1)}$ by mutation such that:

$$P_{(g+1)} = \mathcal{N}(P_g, m) \tag{3}$$

Where $\mathcal{N}(\mu, \sigma)$ is the standard normal distribution. We then updated the new population using:

$$p_{g+1} = \begin{cases} p_{g+1}, \ F(p_{g+1}) > F(p_g) \\ p_g, \ \text{otherwise} \end{cases} \tag{4}$$

where p denotes an individual in the population at generation g and the fitness function F is determined by the treatment as defined in Eq. 2. As shown, each child in the evolutionary algorithm only competes with its direct parent, creating S individual climbers.

Measuring catastrophic interference.

Catastrophic interference, in its simplest formulation, occurs when an improvement in one environment incurs reduced performance in one or more other environments. In an evolutionary setting, catastrophic interference can be measured at the highest temporal resolution by considering mutations: the change in performance between a parent and child for each environment experienced by both agents (Fig. 2).

For the purposes of analysis, we only track performance changes between mutations where overall fitness increased (we do not record deleterious or stagnant mutations). Also we only record and perform analysis on the change in Euclidean distance of the robot from the light source (blue and red bars in Fig. 2). This provides a more intuitive understanding of changes in performance and allows comparison between any 2 treatments regardless of the fitness function used in that treatment.

We define a function D on an individual such that it returns a vector $[x_1, x_2, ..., x_n]$, where x_i is the distance from the light source at the end of simulation in environment i. As shown in Fig. 2, fitness and distance are inversely correlated therefore we record change in distance after every successful mutation as:

$$\Delta D = -\left[D\left(p_{(g+1)}\right) - D\left(p_g\right)\right]. \tag{5}$$

We negate this difference, so that a positive increase in fitness in an environment causes an increase in the corresponding component of ΔD. We apply four metrics (M_1, M_2, M_3, M_4) to this data (ΔD) to measure the overall effect of catastrophic interference in each each of the treatments.

M_1: Average Worst Absolute Distance For this metric, we take the distances from the light sources for the champion from each run, for each treatment. We here define a run champion such that:

$$C = \min_{P_g}\left[\max D(p_g)\right], \text{ where } g = 3000. \tag{6}$$

Recall that distance is measured from the light source so higher is worse. Therefore a champion C is the one with the lowest distance in it's worst environment. We compute C for each run and take the average across all runs, for each treatment:

$$M_1 = \frac{1}{N}\sum_{i=1}^{N=30} C_i \tag{7}$$

If this number is particularly high for a given treatment then it either failed to evolve robots that behaved well in at least one environment, a sign of catastrophic interference. If this number is sufficiently low for a given treatment, it was able to evolve agents that solved all environments, and at worst suffered mild amounts of interference.

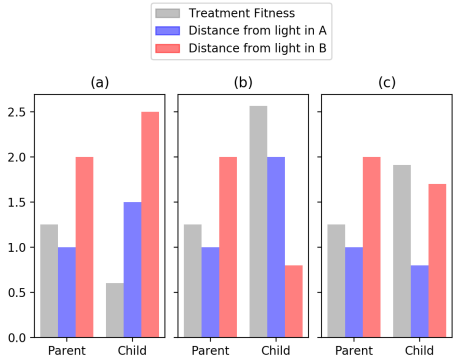

Figure 2: Example mutations that (a) are deleterious, (b) result in catastrophic forgetting, and (c) avoid catastrophic forgetting. The smaller the distance from the light source (blue, red), the higher the fitness (inverse square law).

M_2: Average θ from V If a treatment is able to avoid evolving specialists by producing children that enjoy increases in fitness in all environments (Fig. 2c), the most beneficial mutation possible would be represented by the vector $V = [x_1, x_2, ..., x_n]$ where all elements are equal, $x_1 = x_2 = ... = x_n$. To compute M_2 we take all of the distance vectors from the mutations of a champion C and record the cosine angle difference between ΔD and V. If the angle is less than or equal to $45°$, then that vector, although perhaps biased slightly toward one or the other environment, nevertheless represents a mutation that avoided catastrophic interference. For this metric, we again record all such beneficial mutations within the lineage of a run that produced that run's champion. This can be represented as:

$$\theta' = \frac{1}{A_i}\sum_{a=1}^{A_i} \arccos\frac{\Delta D_a \cdot (V)}{|\Delta D_a||V|} \tag{8}$$

$$\theta_i = \begin{cases} \theta' : \theta'_i \leq 180° \\ |\theta' - 360°| : \text{otherwise} \end{cases} \tag{9}$$

$$M_2 = \frac{1}{N}\sum_{i=1}^{N=30} \theta_i \tag{10}$$

where A_i is the number of recorded ΔD's, θ' denotes the angle between the ΔD and V, and θ_i represents the average amount of catastrophic interference that individuals in the lineage of the champion from run i experienced (lower θ_i represents less catastrophic interference). M_2 thus represents the average amount of catastrophic interference experienced by the run champion lineages in a treatment. If M_2 is low for a given treatment, then that treatment can be considered to be resistant to interference.

M_3: Average ΔD Length Again we take the vector ΔD of a beneficial mutation that occurred within the ancestral lines of run champions. We then perform a similar method as in Eq. 8 on the length $|\Delta D|$ as shown:

$$L_i = \frac{1}{A_i} \sum_{a=1}^{A_i} |\Delta D_a| \quad (11)$$

$$M_3 = \frac{1}{N} \sum_{i=1}^{N=30} \Delta L_i \quad (12)$$

M_3 represents the average magnitude of improvement made during mutations by the run champion lineages in a treatment. This metric only makes sense in conjunction with M_2. Consider a treatment with small M_2 yet also small M_3. The treatment will suffer less from catastrophic interference, but is insignificant since the mutations yield insignificant improvements. However, if M_2 is small and M_3 is large for a given treatment, that treatment not only yielded mutations that avoided catastrophic interference but also exhibited high evolvability.

M_4: Average Percentage of Points in Quadrant I Even though we attempt to measure catastrophic interference among just beneficial mutations, it is possible for a beneficial mutation to make a sufficiently large improvement in fitness in one environment even though there is some degradation in performance in the other environment (e.g. Fig. 2b). This equates to a ΔD where one of its elements is negative. When plotted as a point for the case of 2 environments the point falls in the upper left or lower right quadrant of a scatter plot. Thus, to buttress our measurement of catastrophic interference we devised a fourth metric, which is simply the fraction of beneficial mutations within run champions' ancestral lines where all elements of ΔD are positive. In the case of 2 environments they fall within the upper right quadrant (Quadrant I). This can be seen visually in Fig. 4.

Results

We analyze the performance and relative amounts of catastrophic interference for all nine treatments across M_{1-4}. We use the Mann-Whitney U test (with Bonferroni correction for eighteen comparisons) to indicate statistical significance at the $p = 0.05$ level. The metrics generally show that there is an interaction between fitness function and morphology. If there was no interaction, an entire row (morphology does not matter) or an entire column (fitness function does not matter) would not be significantly different.

M_1: Average Worst Absolute Distance

As can be seen in Table 1, evolutionary performance generally improves moving from top to bottom row-wise and left to right column-wise. This metric shows that the whegged robot significantly outperformed the other two robots and that the min function is generally a better fitness function for this task. Interestingly, however, the treatments under the product and min fitness functions aren't significantly different. This suggests morphology might have a greater impact on this metric.

M_2: Average θ from V

We now wish to investigate whether the greater evolvability seen for the whegged robot under the product and min fitness functions is a result of those treatments being better able to resist catastrophic interference. Table 2 shows that the whegged robot with the min fitness function achieves beneficial mutations that yield improvements in both environments, or least only slight decreases in fitness in one of them evidenced by their higher relative proximity to V.

M_3: Average ΔD Length

Even though the whegged robot with the min fitness function may yield evolutionary improvements in both environments after a single mutation, those increases in fitness may be very small and thus not contribute to the observed evolvability in that treatment. If so, one would expect M_3 to be very low for this treatment. However, as Table 3 reports, this is not the case: the legged robot with the min fitness function has the lowest M_3 value, and the value for the whegged robot with the min fitness function is significantly higher.

M_4: Average Percentage of Points in Quadrant I

Using this metric we can conclude that the min function, regardless of morphology, seems to better force mutations to results in less catastrophic interference, at least according to this particular metric (Table 4). We can conclude this because each M_4 value in the min row is significantly higher than the two M_4 values in the two entries above it.

Performance in four environments.

The same general pattern held when we scaled our approach from two to four environments. Namely, the pairwise comparisons that were significant (at the 0.05 level) in two environments remained so in four environments. However, while the whegged robot with the product and min fitness functions similarly outperformed the other treatments, they could not solve all four environments.

Discussion

As shown by Figures 4 and 3 as we change the morphology from legged to whegged the robots demonstrate increased evolvability. Thus the fitness landscape allows for larger jumps towards the optima. This includes those jumps that avoid catastrophic interference altogether: mutations visualized by points in Figs. 4 and 3 that lie in the upper right quadrant.

In conjunction, as we change the fitness function from sum to min, we see the spread of points in Figures 4 and

Table 1: Mean of M_1 across treatments. Arrows indicate statistical significance between adjacent cells.

D	**Legged**		**Wheeled**		**Whegged**	
$\sum$	8.181		8.386	$\leftrightarrow$	5.915	$\leftrightarrow$
	$\updownarrow$		$\updownarrow$		$\updownarrow$	
$\prod$	6.530	$\leftrightarrow$	3.593	$\leftrightarrow$	1.234	$\leftrightarrow$
min	5.827	$\leftrightarrow$	3.575	$\leftrightarrow$	1.296	$\leftrightarrow$

Table 2: Mean of M_2 across treatments. Arrows indicate statistical significance between adjacent cells.

θ	**Legged**	**Wheeled**		**Whegged**	
$\sum$	71.035	73.362		72.640	
	$\updownarrow$	$\updownarrow$		$\updownarrow$	
$\prod$	66.292	71.090		66.895	
	$\updownarrow$	$\updownarrow$		$\updownarrow$	
min	57.843	58.226	$\leftrightarrow$	48.844	$\leftrightarrow$
	$\updownarrow$	$\updownarrow$		$\updownarrow$	

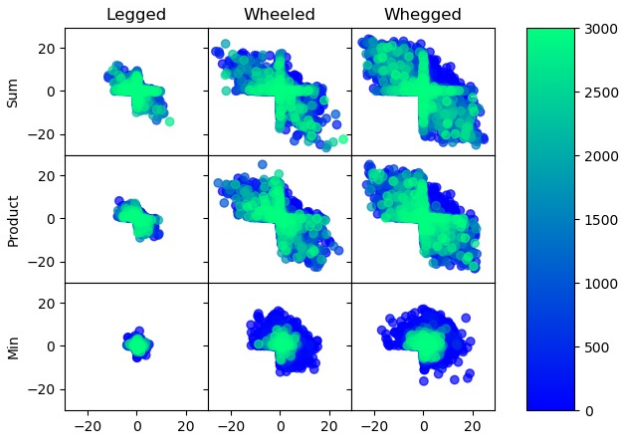

Figure 3: Change in fitness (ΔD, as defined in Eq.5), in two environments, for the three robots (Fig. 1) and three fitness metrics (Eq. 2), colored by the generation of the mutation. Dots in the upper-right quadrants of each robot-fitness cell represent beneficial changes in both environments; these mutations avoided catastrophic interference. Dots in the upper-left and lower-right quadrants of each cell are mutations that were beneficial in one environment but deleterious in the other; these changes caused catastrophic interference. We did not record mutations that were deleterious in both environments (lower-left quadrants).

3 condense toward the origin. When combined with the whegged robot, we see a significant improvement in the metrics we used to measure catastrophic interference. It

Table 3: Mean of M_3 across treatments. Arrows indicate statistical significance between adjacent cells.

$\lvert\Delta D\rvert$	**Legged**		**Wheeled**		**Whegged**	
$\sum$	1.734	$\leftrightarrow$	3.209		4.971	$\leftrightarrow$
			$\updownarrow$		$\updownarrow$	
$\prod$	1.889	$\leftrightarrow$	3.372	$\leftrightarrow$	5.341	$\leftrightarrow$
	$\updownarrow$		$\updownarrow$		$\updownarrow$	
min	1.164	$\leftrightarrow$	2.411		2.001	$\leftrightarrow$
	$\updownarrow$				$\updownarrow$	

Table 4: Mean of M_4 across treatments. Arrows indicate statistical significance between adjacent cells.

% in I	**Legged**	**Wheeled**		**Whegged**	
$\sum$	25.655	22.903		30.824	
	$\updownarrow$	$\updownarrow$		$\updownarrow$	
$\prod$	22.973	22.167		26.981	
	$\updownarrow$	$\updownarrow$		$\updownarrow$	
min	49.519	48.674	$\leftrightarrow$	54.332	$\leftrightarrow$
	$\updownarrow$	$\updownarrow$		$\updownarrow$	

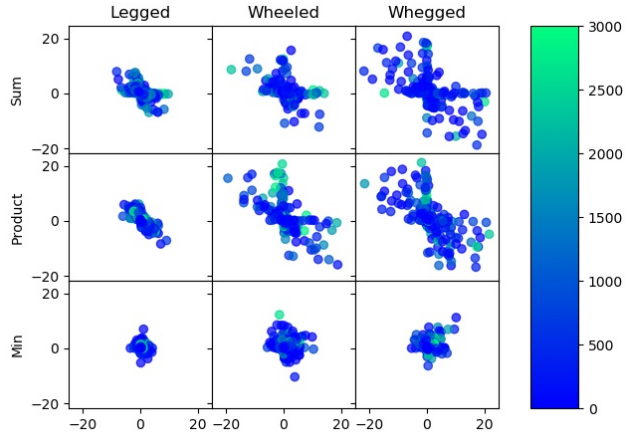

Figure 4: The same as Fig 3 , but only for the run champs.

appears that it is the combination of correct fitness function (min) with the correct morphology (whegged) that resists catastrophic interference: changes in morphology and fitness alone are not sufficient. We hypothesize that this greater resistance to catastrophic interference is what enables the whegged robot, under the min fitness function, to achieve higher fitness within environments and consistent fitness across environments.

One objection to this hypothesis could attribute the performance of the whegged robot to the increased speed allowed for by wheels. We do not feel that this is valid for two reasons: the wheeled robot also has wheels and does not achieve the same level of performance, and the evalua-

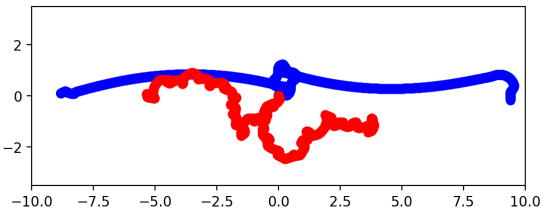

Figure 5: A tracing of a typical whegged robot (blue) and legged robot (red) trained in two environments, under the min fitness function. The light source is first placed at (9, 0), and then at (-9, 0). Video is available at youtu.be/uWy33A5HZGM .

tion time of a simulation was set such that all morphologies are able to reach the light source before the end of simulation. Indeed we observed that all morphologies reached and waited at the light source when trained against a single environment.

In observing the behavior of the robots we noticed a pattern among whegged robots that could account for their resistance to interference. Whegged robots move very rapidly in a circular pattern during the initial time steps of a simulation which may allow them to 'sidestep' catastrophic interference by rapidly turning unfamiliar environments into familiar ones. An example is shown in Fig. 5: the rotationally symmetric trajectories of the blue whegged robot indicates it has recognized two versions of the same environment. The red legged robot does not: its two trajectories are different, and take longer to diverge. The wheeled and legged robot both seem to have much more difficulty in turning.

Conclusions

This work suggests not only that the particular choice of morphology and fitness function for embodied agents can affect their resistance to catastrophic interference, but the very fact that the agent has a body can help. In effect, an agent can use its body to move in such a way that a seemingly different training instance converges sensorially to a familiar instance. The implication of this is that the very phenomenon of catastrophic interference itself may be to some degree a false problem arising from investigations using non-embodied systems: Since such systems do not have control over their input, they cannot align objects of interest in different training instances and thus reduce catastrophic interference.

A simple example may suffice here: a human face that appears in two different locations in an image may be difficult for a non-embodied learner to recognize, unless there is a large amount of training data that contains diversity along that feature (face position). In contrast, an embodied agent equipped with a camera that experiences the same two stimuli may learn to move such that the face is centered in its field of view. Furthermore, it may be that different types of embodied agents may more easily discover and perform this centering. Finally, such appropriately embodied agents may thus be able to generalize about faces regardless of position using less training instances than the non-embodied agent because of this ability. However, whether this latter system is indeed more scalable in this way compared to an equivalent non-embodied system remains as future work.

Acknowledgments

This work was supported by DARPA contract HR0011-18-2-0022. The computational resources provided by the Vermont Advanced Computing Core (VACC) are gratefully acknowledged.

References

Belter, D., Skrzypczyński, P., Walas, K., and Wlodkowic, D. (2015). Affordable multi-legged robots for research and stem education: a case study of design and technological aspects. In *Progress in Automation, Robotics and Measuring Techniques*, pages 23–34. Springer.

Bernatskiy, A. and Bongard, J. (2017). Choice of robot morphology can prohibit modular control and disrupt evolution. In *Proceedings of the 14th European Conference on Artificial Life*, pages 60–67. MIT Press.

Bongard, J. (2011). Morphological change in machines accelerates the evolution of robust behavior. *Proceedings of the National Academy of Sciences*, 108(4):1234–1239.

Bongard, J., Zykov, V., and Lipson, H. (2006). Resilient machines through continuous self-modeling. *Science*, 314(5802):1118–1121.

Bongard, J. C. (2013). Evolutionary robotics. *Communications of the ACM*, 56(8):74–83.

Bongard, J. C., Bernatskiy, A., Livingston, K., Livingston, N., Long, J., and Smith, M. (2015). Evolving robot morphology facilitates the evolution of neural modularity and evolvability. In *Proceedings of the 2015 on Genetic and Evolutionary Computation Conference*, pages 129–136. ACM.

Bongard, J. C. and Pfeifer, R. (2001). Repeated structure and dissociation of genotypic and phenotypic complexity in artificial ontogeny. In *Proceedings of the 3rd Annual Conference on Genetic and Evolutionary Computation*, pages 829–836. Morgan Kaufmann Publishers Inc.

Cheney, N., MacCurdy, R., Clune, J., and Lipson, H. (2013). Unshackling evolution: evolving soft robots with multiple materials and a powerful generative encoding. In *Proceedings of the 15th annual conference on Genetic and evolutionary computation*, pages 167–174. ACM.

Clune, J., Mouret, J.-B., and Lipson, H. (2013). The evolutionary origins of modularity. In *Proc. R. Soc. B*, volume 280, page 20122863. The Royal Society.

Dang, D.-C. and Lehre, P. K. (2016). Self-adaptation of mutation rates in non-elitist populations. In *International Conference on Parallel Problem Solving from Nature*, pages 803–813. Springer.

Ellefsen, K., Mouret, J., Clune, J., and Bongard, J. C. (2015). Neural modularity helps organisms evolve to learn new skills without forgetting old skills. *PLoS Comput Biol*, 11(4):e1004128.

Ellefsen, K. O., Mouret, J.-B., and Clune, J. (2014). Neural modularity reduces catastrophic forgetting. *The Evolution of Learning: Balancing Adaptivity and Stability in Artificial Agents*, page 111.

Espinosa-Soto, C. and Wagner, A. (2010). Specialization can drive the evolution of modularity. *PLoS Comput Biol*, 6(3):e1000719.

Fernando, C., Banarse, D., Blundell, C., Zwols, Y., Ha, D., Rusu, A. A., Pritzel, A., and Wierstra, D. (2017). Pathnet: Evolution channels gradient descent in super neural networks. *arXiv preprint arXiv:1701.08734*.

Floreano, D. and Mondada, F. (1994). Automatic creation of an autonomous agent: Genetic evolution of a neural network driven robot. In *Proceedings of the third international conference on Simulation of adaptive behavior: From Animals to Animats 3*, pages 421–430. MIT Press.

French, R. M. (1999). Catastrophic forgetting in connectionist networks. *Trends in cognitive sciences*, 3(4):128–135.

Goodfellow, I. J., Mirza, M., Xiao, D., Courville, A., and Bengio, Y. (2013). An empirical investigation of catastrophic forgetting in gradient-based neural networks. *arXiv preprint arXiv:1312.6211*.

Gruau, F. and Whitley, D. (1993). Adding learning to the cellular development of neural networks: Evolution and the baldwin effect. *Evolutionary computation*, 1(3):213–233.

Harvey, I., Husbands, P., Cliff, D., Thompson, A., and Jakobi, N. (1997). Evolutionary robotics: the Sussex approach. *Robotics and autonomous systems*, 20(2):205–224.

Hauser, H., Ijspeert, A. J., Füchslin, R. M., Pfeifer, R., and Maass, W. (2011). Towards a theoretical foundation for morphological computation with compliant bodies. *Biological cybernetics*, 105(5-6):355–370.

Jehanno, J.-M., Cully, A., Grand, C., and Mouret, J.-B. (2014). Design of a wheel-legged hexapod robot for creative adaptation. In *Mobile Service Robotics*, pages 267–276. World Scientific.

Kashtan, N. and Alon, U. (2005). Spontaneous evolution of modularity and network motifs. *Proceedings of the National Academy of Sciences of the United States of America*, 102(39):13773–13778.

Kim, Y.-S., Jung, G.-P., Kim, H., Cho, K.-J., and Chu, C.-N. (2014). Wheel transformer: A wheel-leg hybrid robot with passive transformable wheels. *IEEE Transactions on Robotics*, 30(6):1487–1498.

Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., et al. (2017). Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, page 201611835.

Kriegman, S., Cheney, N., and Bongard, J. (2017). How morphological development can guide evolution. *arXiv preprint arXiv:1711.07387*.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

Lee, S.-W., Lee, C.-Y., Kwak, D.-H., Kim, J., Kim, J., and Zhang, B.-T. (2016). Dual-memory deep learning architectures for lifelong learning of everyday human behaviors. In *IJCAI*, pages 1669–1675.

Lehman, J., Chen, J., Clune, J., and Stanley, K. O. (2017). Safe mutations for deep and recurrent neural networks through output gradients. *arXiv preprint arXiv:1712.06563*.

Lipson, H. and Pollack, J. B. (2000). Automatic design and manufacture of artificial lifeforms. *Nature*, 406:974–978.

Lipson, H., Pollack, J. B., Suh, N. P., and Wainwright, P. (2002). On the origin of modular variation. *Evolution*, 56(8):1549–1556.

Lohmann, S., Yosinski, J., Gold, E., Clune, J., Blum, J., and Lipson, H. (2012). Aracna: An open-source quadruped platform for evolutionary robotics. *Artificial Life*, 13:387–392.

McCloskey, M. and Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. *Psychology of learning and motivation*, 24:109–165.

Miikkulainen, R., Liang, J., Meyerson, E., Rawal, A., Fink, D., Francon, O., Raju, B., Navruzyan, A., Duffy, N., and Hodjat, B. (2017). Evolving deep neural networks. *arXiv preprint arXiv:1703.00548*.

Rusu, A. A., Rabinowitz, N. C., Desjardins, G., Soyer, H., Kirkpatrick, J., Kavukcuoglu, K., Pascanu, R., and Hadsell, R. (2016). Progressive neural networks. *arXiv preprint arXiv:1606.04671*.

Sabour, S., Frosst, N., and Hinton, G. E. (2017). Dynamic routing between capsules. In *Advances in Neural Information Processing Systems*, pages 3859–3869.

Schroer, R. T., Boggess, M. J., Bachmann, R. J., Quinn, R. D., and Ritzmann, R. E. (2004). Comparing cockroach and whegs robot body motions. In *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on*, volume 4, pages 3288–3293. IEEE.

Sims, K. (1994). Evolving 3D morphology and behaviour by competition. *Artificial Life IV*, pages 28–39.

Stanley, K. O. and Miikkulainen, R. (2002). Evolving neural networks through augmenting topologies. *Evolutionary computation*, 10(2):99–127.

Szubert, M., Kodali, A., Ganguly, S., Das, K., and Bongard, J. C. (2016). Semantic forward propagation for symbolic regression. In *International Conference on Parallel Problem Solving from Nature*, pages 364–374. Springer.

Teo, J., Tahir, A., Daut, N., Rusli, N. M., and Khamis, N. (2016). Fixed vs. self-adaptive crossover-first differential evolution. *Applied Mathematical Sciences*, 10(32):1603–1610.

Vanneschi, L., Castelli, M., and Silva, S. (2014). A Survey of Semantic Methods in Genetic Programming. *Genetic Programming and Evolvable Machines*, 15(2):195–214.

Velez, R. and Clune, J. (2017). Diffusion-based neuromodulation can eliminate catastrophic forgetting in simple neural networks. *arXiv preprint arXiv:1705.07241*.