Reweighting Rewards in Embodied Evolution to Achieve a Balanced Distribution of Labour

Sebastian Bangel¹, Evert Haasdijk²

¹Rhine-Waal University of Applied Sciences, Cleves, GER 47533
²Vrije Universiteit Amsterdam, Amsterdam, NL 1081 HV
sebastian.bangel@hsrw.org, e.haasdijk@vu.nl

Abstract

Embodied evolution aims at self-sufficient adaptation in robot collectives in their task environment. An open question is how to achieve a good balance of effort over multiple tasks using embodied evolution. Most efforts to date rely on switching between predefined behaviours or on spatial or temporal separation of the tasks to achieve this. The research presented here is part of an effort to enable embodied evolutionary systems to achieve a balanced distribution of effort over multiple tasks without predefined behaviour and without any separation of the tasks. We propose and experimentally evaluate a selection mechanism that introduces a local reproductive advantage to individuals that specialise in under-represented tasks. The paper shows that an embodied evolution implementation with this mechanism leads to balanced populations of generalist individuals, even when the environment severely penalises generalist behaviour. An extension that combines task-based and purely environmental selection in some cases leads to a balanced population of specialists, but does so inconsistently.

Introduction

Embodied Evolution was introduced by Watson et al. (2002) as an attempt at a robust algorithm that generates autonomous, self-sufficient robot collectives able to handle scenarios unpredicted by their designers. An important requirement for such autonomy is that the robots must develop robust control strategies without human supervision but through adaptation by artificial evolution based on specification of requirements (Nelson et al., 2009). In contrast to standard evolutionary computation, the evolutionary process in embodied evolution is not guided by a global authority with a view of the entire population, but results from decentralised, local interactions between the individual robots. The autonomous, distributed evolutionary process implies the possibility of adapting to changing environments and tasks, and it can accelerate the evolutionary process by its inherent parallelism (Silva et al., 2016).

In this paper, we consider the use of embodied evolution in situations where the robot collective has multiple tasks that cannot be efficiently tackled together. Montanier et al. (2016) showed that this is a challenging proposition, particularly when the task behaviour (how to tackle each task) has to evolve, not only the choice of which task to perform.

Jones and Mataric (2003) showed that robot collectives can efficiently achieve a division of labour in a concurrent foraging task (i.e., a task with two types of resource, with robots able to forage for one type at a time) by selecting behaviour based on local estimates of the proportion of robots that forage for either resource. A division of labour or sequencing of subtasks can also be achieved by providing positive feedback to robots that repeatedly perform the same task (Brutschy et al., 2012) or penalising switching behaviour (Goldsby et al., 2010). Similarly, Trueba et al. (2013) showed that genetically encoding the choice of subtask allows a robot collective to achieve optimal proportions of subtask specialists. Ferrante et al. (2015) showed that a population can achieve division of labour in a foraging scenario where a steep slope separates two subtasks. In their experiments, the robot collective split into two subgroups: one group at the top of the slope slides pucks down and the other group picks up the pucks at the lower end.

Some of these approaches require spatial, temporal or sequential separation of tasks as a precondition. However, division of labour is also achievable with a uniform task distribution and non-sequential tasks in non-evolutionary (Jones and Mataric, 2003) and simplified evolutionary (Trueba et al., 2013) set-ups. All these systems consider a search space of predefined behaviours: any adaptation (evolutionary or not) is at the level of selecting or ordering (sub)tasks, not at the level of learning how to perform them.

Montanier et al. (2016) considered the problem of specialisation while learning the appropriate behaviour. They pointed out that the driving force for behavioural specialisation in such a setting is primarily reproductive isolation by spatially separated task types: their embodied evolution experiments showed no specialisation unless the tasks were spatially separated.

It is unclear how to achieve an equitable distribution of tasks in a population of specialist robots while they learn how to perform the tasks without reproductive isolation. This paper investigates if a selection mechanism that in-
Figure 1: Experiment screen shot. Robots are shown as small blue circles with sensors indicated as beams. Two types of pucks are shown as small green and red squares. The shaded orange rectangles indicate arena walls and obstacles.

Introduces a local reproductive advantage to individuals that specialise in under-represented tasks can achieve a balanced population of specialists, even if there is no reproductive isolation by spatial variation of task distribution. Such a mechanism should achieve a situation where the reproductive advantage of choosing a rare task outweighs the cost of behavioural specialisation (Brutschy et al., 2012).

To this end, we propose a weighting scheme that evaluates the local task distribution cf. Jones and Mataric (2003) in an embodied evolution setting where this information is used to select genetic material. This amounts to a more sophisticated variant of the market mechanism introduced by Haasdijk et al. (2014) that was shown to promote generalist behaviour and so prevent the population from focussing on one task exclusively when the penalty for generalist behaviour was limited. Specifically, we propose to use a sigmoid weighting function to determine the relative value of multiple tasks instead of the proportional relation that Haasdijk et al. defined. We evaluate the effects of selection using this weighting scheme on a population of 100 simulated robots in an environment that discourages generalist behaviour.

We show that the proposed weighting scheme can in some cases lead to populations that balance the ratio of specialist behaviour, but only when the selection pressure exerted by the environment and that exerted by task-based selection are similar.

Methods

Haasdijk et al. (2014) introduced MONEE (MultiObjective Nd open-Ended Evolution), an implementation of embodied evolution that adds support for multiple tasks by means of a market mechanism. The experiments in this paper are based on the implementation as presented there, and the following subsection describes that same environment and evolutionary set-up.

Environment and Evolution

The population consists of 100 differential drive robots in an arena with obstacles and two types of puck. The robots pick up pucks by driving over them; when a puck is collected, a replacement puck is immediately placed in a random position in the arena. The robots are equipped with 8 obstacle sensors, of which 6 face forward and 2 face to the rear (identical to the arrangement of sensors in e-pucks); 8 puck sensors for each type of puck are arranged similarly. The robots are controlled by a fully connected feed-forward artificial neural network that takes all sensor activation levels as inputs and connects these directly to two output nodes with sigmoid activation function that define the requested angular and translational velocity.

Evolution

Evolution proceeds as follows: robots move around the arena for a fixed lifetime of 2000 time-steps to evaluate their controller. They then revert to an ‘egg’ state for 200 time-steps. In this state, they do not move but wirelessly receive the genomes and performance metrics of any ‘live’ robots that pass by within communication range. At the end of this phase, the robots select one of the received genomes in a binary tournament using the genomes’ performance (calculated as described in the following paragraph), applies mutation to it and starts a new ‘live’ phase of 2000 time-steps, broadcasting its active genome and its performance to any robots in egg state that it encounters.

Genomes encode the weights of the neural network connections and mutation is implemented as Gaussian perturbation of the weights with $\sigma = 0.1$. The weights are initialised randomly from $\mathcal{N}(0, 1)$.

To avoid synchronised life cycles where all robots are in the ‘egg’ phase at the same time, a small perturbation is added to every lifetime. Thus, evolution occurs in a decentralised and asynchronous manner, and to successfully reproduce, robots must move around the arena so that they encounter robots in egg state (so that they can transmit their genome) and collect pucks (to increase the likelihood of selection).

MONEEE’s Market Mechanism

MONEE implements a market mechanism to accommodate multiple objectives by assigning exchange rates to reflect the relative scarcity of each objective. The value of each objective is inversely proportional to the amount of reward collected for that objective. For an objective $o$, the exchange rate $r_o$ is calculated...
as:

\[ r_o = \begin{cases} \frac{\sum_{t \in T} C_t}{C_o} & \text{for } C_o > 0 \\ 0 & \text{otherwise} \end{cases} \] (1)

with \( C_t \) the total collected reward for objective \( t \) by the received genomes combined and \( T \) the set of defined objectives. Typically, every puck type represents a separate objective and the reward is the number of pucks of that type collected. A genome’s performance is the sum of the rewards for all objectives weighted by their exchange rates.

**Mutually Exclusive Tasks** The experiments described in this paper require tasks that are to some extent mutually exclusive, i.e., that cannot be efficiently tackled together. To this end, the environment discourages generalist behaviour by means of a speed penalty. The movement speed of each robot is limited depending on their task specialisation level. Each robot’s requested translational velocity is multiplied by a specialisation factor that is 1 when collecting only pucks of one type (i.e., specialising in one task) and smaller for generalist behaviour:

\[ v = \left( \frac{\max_{g \in T} C_t(g)}{\sum_{g \in T} C_t(g)} \right)^s \cdot v_{desired} \] (2)

with \( g \) the currently active genome and \( T \) the set of all defined tasks. \( C_t(g) \) are the credits collected for task \( t \) with genome \( g \) being active. The specialisation level \( s \) scales the severity of the penalty: higher values imply a stronger pressure towards specialisation. This simulates a situation as described by Smith (1776), where specialising in one tasks leads to increased performance of that task: foraging for one type of puck is more efficient than foraging for multiple types of puck.

With this scheme, specialist robots can move at full speed, while generalist robots move more slowly, reducing the odds of reproductive success for generalists because they can collect fewer pucks and they encounter fewer opportunities to transmit their genomes. Haasdijk et al. (2014) showed that this penalty scheme results in populations of fast-moving specialists, and that the market mechanism can counter that tendency to some extent by promoting generalist behaviour. Because of the incurred speed penalty, generalist behaviour comes at the price of reduced task performance.

**An Improved Market Mechanism**

We compare the proportional weighting scheme from eq. (1) to a sigmoid-based scheme that amplifies the selection pressure towards rare tasks. In this sigmoid market scheme for objective \( o \), the exchange rate \( r_o \) is calculated as:

\[ r_o = \frac{1}{1 + e^{-k \left( \sum_{t \in T} \frac{C_t}{C_o} - 1 \right)}} \] (3)

with \( k \) a parameter that controls the steepness of the slope of the sigmoid curve. High values of \( k \) result in an abrupt, disproportionate change in valuation for tasks that are over-represented, similar to the step function used by Jones and Mataric (2003). Low values of \( k \) result in a linear market scheme. Note that the exchange rate of the sigmoid scheme is limited between 0 and 1 whereas the exchange rate of the original proportional market scheme is an unlimited value \( \geq 1 \).

Table 1 shows an example performance calculation with three genomes \( A, B \) and \( C \) for \( k = 1 \) and \( k = 1000 \). Genome \( C \) collected 2 more pucks than genome \( B \). With \( k = 1 \), genome \( C \) has the highest performance because of its higher total puck count. With \( k = 1000 \), the exchange rate favours rare pucks more strongly. Therefore genome \( B \) has the same fitness as \( C \) even tough it collected fewer pucks in total because it collected more rare pucks.

The pseudo-code of the fitness calculation that assigns a fitness to each genome with the proportional or sigmoid market scheme is depicted in algorithm 1.

**Analytical methods**

**Specialisation level** To analyse the level of specialisation in this particular two-puck set-up, we define the metric \( \gamma \) based the ratio of type 1 pucks collected over the total number of pucks collected as follows:

\[ \gamma = \left| \frac{T_1}{T_1 + T_2} - 0.5 \right| \] (4)

A \( \gamma \) close to 0 indicates that and equal number of both types of puck were collected. If mostly pucks of one type were collected, this value will be close to 0.5.

The \( \gamma \) metric is used to express the specialisation level of an individual\(^1\) as well as the task distribution in a population of robots by calculating the ratio of the pucks collected by

---

\(^1\)We will use the terms *individual* and *genome* interchangeably
productive success and reports employs Fisher’s exact test to determine reproductive success indicates a high selection pressure. It based on the assumption that a high probability of a nonrandom relation between an individual’s performance and its reproductive success indicates a high selection pressure. It based selection pressure is based on the relation between reproductive success and the number of pucks collected.

Experimental Set-Up

The experiments were implemented in the simple 2D simulator RoboRobo (Bredeche et al., 2013) written in C++. Figure 1 shows an image of the arena with robots, pucks and obstacles as used in the experiments.

The arena contains two types of puck, 150 of each type, spread throughout the arena. Each puck type represents a separate task. The experiments were run with a specialisation level $s = 3.0$ cf. eq. (2). Haasdijk et al. (2014) showed that at a specialisation level of $s = 3.0$ the proportional market scheme fails to balance the task proportion: all robots in the population specialise in the same task.

To investigate the effect of the proportional and the sigmoid market schemes, a first set of experiments compares runs of MONEE with the market mechanism disabled (the value for each collected puck as fixed at 1, with the proportional market scheme cf. eq. (1) and with the new sigmoid market scheme cf. eq. (3) with $k = 10,000$. This setting of $k$ approaches the step function used by Jones and Mataric (2003). We report the results of 10 replicate runs with different random seeds for each treatment, each run lasting 1,000,000 time-steps.

A second set of experiments considers different settings for $k$, ranging from 1 to 1000. For this set, we report the results of 10 replicate runs with different random seeds for each setting, each run lasting 2,000,000 time-steps.

Results and Analysis

Market Scheme Comparison

Figure 2 shows the results of the three market schemes with one dot representing the results of one run. The horizontal axis shows $\gamma_{pop}$ and the vertical axis the mean $\gamma_{ind}$ for each population. The colour of the dots indicates the total number of pucks (both types combined) collected by the population. All metrics were calculated over the individuals that completed their ‘live’ state in the last 1000 time steps of the run.

With the environment enforcing specialisation and without a market mechanism to counteract this (left panel), the robots evolved behaviour that picks up one type of puck — the same for all robots in a population — and steers clear of pucks of the other type. This avoids the speed penalty and so the robots could pick up more pucks than they could with

---

Algorithm 1: MONEE’s market algorithm

```
T ← set of defined objectives
G ← set of received genomes

foreach t ∈ T do // count credits
    foreach g ∈ G do
        C_t ← C_t + C_t(g)
    end
    C ← C + C_t
end

foreach t ∈ T do // exchange rates
    if market then
        if sigmoid market then
            r_t = \frac{1}{1 + e^{k_C(g)}} // cf. (3)
        else
            r_t = \frac{\sum_{i \in C_t} g_i}{C_t} // cf. (1)
        end
    end

foreach g ∈ G do
    foreach t ∈ T do
        g.fitness ← g.fitness + (r_t · C_t(g))
    end
end
```

all the robots in the population. This is notated as $\gamma_{ind}$ and $\gamma_{pop}$, respectively.

Individuals with a high $\gamma_{ind}$ (close to 0.5) specialise in one task, individuals with low $\gamma_{ind}$ (close to 0) are generalists. High values of $\gamma_{pop}$ indicate that the population as a whole focusses on only one task, neglecting other tasks, while populations that distribute their effort equitably will have low $\gamma_{pop}$. Note a population with high $\gamma_{pop}$ can only consist of individuals with high $\gamma_{ind}$ (i.e., specialised individuals). Populations with low $\gamma_{pop}$ can consist of generalist individuals or a balanced proportion of specialist individuals. If $\gamma_{pop} \approx \gamma_{ind}$, the population homogeneously consists of individuals at a common specialisation level.

When reporting the specialisation level for a run, we consider all the individuals that ended their life cycle in the last 1000 time steps as the population representing that run.

Selection pressure Haasdijk and Heinermann (2016) present an appropriate method to quantify selection pressure based on the assumption that a high probability of a nonrandom relation between an individual’s performance and its reproductive success indicates a high selection pressure. It employs Fisher’s exact test to determine $p_{random}$, the probability a nonrandom relation between performance and reproductive success and reports $-\log(p_{random})$ as a metric.

Similar to the analysis of selection pressure in Haasdijk et al. (2014), we consider selection pressure deriving from the environment separate from that deriving from the task. The environmental selection pressure is measured by quantifying the relation between an individual’s reproductive success and the distance it covered during its lifetime. The task-based selection pressure is based on the relation between reproductive success and the number of pucks collected.
Figure 2: Mean individual specialisation $\gamma_{\text{ind}}$ vs. population specialisation $\gamma_{\text{pop}}$ for selection without a market scheme, with the proportional and with the sigmoid market scheme. Individuals that collected no pucks at all were ignored. The colour indicates how many pucks were collected by the population. Each dot represents one run; the data of the last 1000 time steps was used to represent the final stage of a run.

generalist behaviour. Consequently, the populations consisted of specialists of one type and focussed on one type of puck exclusively.

The proportional market scheme (middle panel) led to reduced specialisation levels on population as well as individual level. Typically $0.25 < \gamma_{\text{ind}} < 0.35$, indicating somewhat generalist behaviour, so that the robots incurred a speed penalty and picked up fewer pucks than specialist robots. Out of these ten runs, two outliers with a more balanced population of less specialised individuals were observed.

The sigmoid market with a high $k$ results in populations that consist of semi-generalist individuals (typically with $0.15 < \gamma_{\text{ind}} < 0.2$) that achieve good balance of effort ($0.0 < \gamma_{\text{pop}} < 0.1$). In a few cases where more specialised individuals occur the population is less balanced.

Overall, the task performance is comparatively low: on average, the robots collected 696.6, 469.3 and 260.3 pucks in the final 1000 time steps in runs without market mechanism, with the proportional and with the sigmoid market mechanism, respectively. An analysis of variance on the number of pucks collected showed a significant effect ($F(2, 27) = 40.67$, $p < 0.001$). Post hoc comparisons using the Tukey HSD test indicated that the number of pucks collected without a market mechanism was significantly ($p < 0.001$) higher than that with the proportional market scheme, which in turn was significantly ($p < 0.001$) higher than the number of pucks collected with the sigmoid market scheme.

**Sigmoid Market With Different Values for $k$**

Figure 3 shows the results of 32 runs with different values of $k$. It is laid out just as fig. 2. Values of $k$ range from 1 (amounting to an almost linear weighting scheme) to 1000 (a step function that values only the rarer task).

The left panel shows that a linear puck exchange rate cannot counteract the environmental drive towards specialisation and leads to a population that almost exclusively focusses on one task. Both $\gamma_{\text{pop}} > 0.4$ and $\gamma_{\text{ind}} > 0.4$. The result is very similar that without any weighting of the tasks cf. fig. 2. The evolved behaviour avoids the speed penalty and so the number of pucks collected is relatively high.

The individual specialisation $\gamma_{\text{ind}}$ decreases with increasing $k$, although even with $k = 1000$ one run still ends with many specialists ($\gamma_{\text{ind}} > 0.3$). For intermediate values of $10 \leq k \leq 100$, lower values of $\gamma_{\text{pop}}$ were observed in runs with high $\gamma_{\text{ind}}$. We observe a trend where the ratio between $\gamma_{\text{ind}}$ and $\gamma_{\text{pop}}$ decreases, indicating that the populations increasingly are more balanced than the individuals. Well-balanced populations consisting of specialists (i.e., populations that combine low $\gamma_{\text{pop}}$ with high $\gamma_{\text{ind}}$) are not observed. When $k = 1000$ most populations have a balanced task distribution and consist of generalist individuals.

The generalist behaviour with higher values for $k$ implies that the number of pucks collected decreases as shown by the darker coloured dots for higher colours of $k$. On average, the robots collected 759.31, 426.59, 333.43 and 306.16 pucks in the final 1000 time steps in runs with $k$ values of 1, 10, 100 and 1000, respectively. An analysis of variance on the task performance showed that the difference in number of pucks collected with different values of $k$ is significant ($F(3, 123) = 117.4$, $p < 0.001$). A post hoc Tukey test showed that the decrease in the number of pucks collected between increasing values of $k$ is significant for all settings in between $k = 1$ and $k = 100$ ($p < 0.005$ in
all cases). However, the number of collected pucks was not significantly different between the runs with $k = 100$ and $k = 1000$ ($p > 0.75$).

Following Haasdijk et al. (2014), we analyse the task-based selection pressure (deriving from the explicit selection on number of pucks collected) with the environmental selection pressure (deriving from implicit selection on distance covered) in fig. 4. The plots show that reproductive success mostly depends on the number of collected pucks, and less on the distance travelled: the explicit task-based selection exerts a higher pressure than the implicit environmental selection. This is in line with Haasdijk et al.’s findings.

The selection pressure, in particular that related to collecting pucks, decreases with increasing $k$; this may be an additional reason that the populations with high $k$ collect fewer pucks.

**Follow-up Experiment With Equalised Selection Pressure**

Interestingly, runs with $k \geq 100$ show a much smaller difference between environmental and task-based selection pressure, although this difference is larger near the end of the runs. This led us to hypothesise that set-ups where the task-based and environmental selection pressure are more or less equal could achieve a population that balances the effort over the tasks with individual specialist behaviour. To test this hypothesis, we performed a set of additional experiments where the selection pressures were equalised.

To achieve equalisation, the procedure to select a genome when a robot’s ‘egg’ phase ends was modified: with 50% likelihood, the standard selection scheme based on task performance applied, while in the other case selection was random. Thus, in half the cases, an individual’s reproductive success depended only on its ability to spread its genome and the number of pucks collected was immaterial. Other than this change, all settings were kept as for the experiments with sigmoid market scheme with $k = 1000$.

Figure 5 shows the resulting selection pressures, with the standard setting included for ease of comparison. The plots show that introducing a 50% likelihood of random selection was successful in equalising the task-based and environmental selection pressure.

Figure 6 shows $\gamma_{ind}$ and $\gamma_{pop}$ for 32 replicate runs with equalised selection pressure. A Kruskal-Wallis H test showed that there was a statistically significant difference in $\gamma_{pop}$ between the runs with and without random selection ($H(1) = 17.104, p < 0.001$). The results show that equalised selection pressure in some cases led to populations that combined individual specialisation ($\gamma_{ind} > 0.3$) with balanced behaviour at global level ($\gamma_{pop} < 0.2$). However, many runs resulted in populations with balanced effort and generalist behaviour or in populations with poorly balanced effort consisting of specialists. An even task distribution on population level cannot be guaranteed: the populations reach arbitrary values of $\gamma_{pop}$. There is no control over the exact task distribution of the population.

**Discussion and Conclusions**

The comparison of task distributions without market, with proportional and with sigmoid market showed that when the environment strongly favours specialist behaviour the sigmoid market clearly outperforms the proportional market in balancing the task distribution of a population. The sigmoid market leads to a balanced population of individuals even with a very high degree of enforced specialisation by the environment.

However, favouring rarely tackled tasks in a local fitness evaluation with a strict exchange rate leads to generalist behaviour—individuals perform both tasks. In these particular experiments, that meant that balanced populations
have decreased task performance because generalist individuals suffer a speed penalty that reduces their effectiveness.

If genomes causing all types specialist behaviour would be present when selecting a genome, the sigmoid market would favour the genomes that specialise in the rare task. However, if only one type of specialist is present, the sigmoid market mostly favours task-balanced individuals as there is a high chance that they perform the rarer task.

This suggests that the local sampling during the ‘egg’ phase rarely considers sets of specialists for all tasks, at least in the early stages of evolution when few robots manage to spread their genomes.

Investigating different $k$ settings supports this trend. Higher values of $k$, strongly favouring rare tasks, cause more
generalist individuals to occur. With low $k$, the sigmoid market has little influence on specialisation. It seems that a higher number of collected pucks of one type increases the chances of reproductive success more than the lower value of that puck type decreases it. We observed no intermediate values of $k$ where specialist individuals make up a stable population with balanced task distribution.

The task-based selection pressure is higher than the environmental selection pressure. Increasing focus on rarely fulfilled tasks (i.e., high $k$), decreases the environmental selection pressure until the covered distance has little influence on the number of offspring, implying a reduced influence of the environmental penalty for generalist behaviour. Instead of specialising, the robots take the penalty in their stride, move more slowly and collect fewer pucks. This also reduces task selection pressure with increasing $k$, as the number of offspring depends more on the puck price than just the number of collected pucks. Consequently, the discrepancy between task-based and environmental selection pressures is smaller for higher values of $k$.

Follow-up experiments that introduced random selection in 50% of the cases showed some promise in achieving balanced population by equalising the selection pressure. In some cases, this led to balanced populations of specialist individuals, but it also led to populations that focus on one task exclusively. An explanation could be as follows: random selection encourages specialists that do not incur the environmental speed penalty, and this in turn can cause the local samples of genomes collected in ‘egg’ state become large enough to contain a mix of specialists. If this is true, a gradual adaptation of the selection mechanism as evolution progresses may provide a solution. Further research will have to show if extensions of this scheme, e.g., gradually increasing the proportion of task-based selection, yield more consistent results. Another avenue of research could consider increasing the controller complexity: a set of neural network controllers per robot (one for each task) and a switch between these may enable a solution where the individuals can learn different types of behaviour and explicitly decide to switch between them. Finally, further work could investigate whether these results generalise to situations with more than two tasks (i.e., multiple puck types).

References


