Description: Beskrivning: hd_vertikal_farg WORD.wmf

Thesis report

Master thesis in Microdata Analysis

Business Intelligence Program

School for Technology and Business Studies

Dalarna University

Description: StreckTitle

All for one and one for All

Author: Bamshad Shirmohammadi

Supervisor: Professor Arend Hintze

Examiner:

Dalarna University

791 88 Falun

Sweden

Tel 023-77 80 00

Subject:

Course code: MI4002

Points: 15/30hp

Date of the examination: 20XX-XX-XX

**Abstract**

**Contents**

[Introduction 1](#_Toc60586508)

[Background 1](#_Toc60586509)

[Purpose 4](#_Toc60586510)

[Literature review 4](#_Toc60586511)

[Material and Methods 4](#_Toc60586512)

[Experiments 5](#_Toc60586513)

[Data description 6](#_Toc60586514)

[LOD.csv 6](#_Toc60586515)

[Movement.csv 6](#_Toc60586516)

[Beep.csv 7](#_Toc60586517)

[Methods 7](#_Toc60586518)

[LOD analyzer 7](#_Toc60586519)

[Movement analyzer 8](#_Toc60586520)

[Beep analyzer 9](#_Toc60586521)

[Results 9](#_Toc60586522)

[How agents adapt over the course of evolution 10](#_Toc60586523)

[Comparison of the different scenarios 12](#_Toc60586524)

[Comparison of clone and not clone groups 14](#_Toc60586525)

[Agents' behaviors 17](#_Toc60586526)

[Do agents communicate for a reason? 18](#_Toc60586527)

[Discussion 19](#_Toc60586528)

[Conclusions 20](#_Toc60586529)

[References 21](#_Toc60586530)

# Introduction

The state of the art machine learning technique to optimize neural network AI controllers (agents) is back-propagation and specifically deep learning. This method assumes that solutions and examples for such solutions are known. A different approach to this is evolutionary computation and genetic algorithms. Instead of rewarding proper responses to specific situations and hoping for proper generalization, a genetic algorithm uses an objective function. This objective function evaluates the total performance of a system and can select controllers that perform better. Over many generations such system optimizes controllers, often finds more creative solutions [1], and most importantly does not require knowledge about specific actions but only the desired outcome. This method works well when optimizing individual controllers as their objectives do not conflict with the objectives of other controllers. When agents need to interact or work in teams [2], the situation becomes much more complicated, and often the goals of the individual are contrary to the success of the group. Imagine self driving cars. Optimizing them to reach their goal as fast as possible might induce negative side effects. How to overcome such issues? Group level selection plays an important role in evolution. Often individuals do not act alone, but in groups. As such they can achieve together more than alone. Collaborative hunting is one of those examples, where individuals are not selected individually but as a group [3]. This group level selection scheme typically pools the resources the group collected and redistributes them back equally, which implies that groups are evaluated by their average performance. What if this scheme is altered? Imagine a group is rewarded according to the performance of its best individual (all for one) or its worst performer (one for all)? In the all for one case, you would pool all resources on one individual, and in the case of the one for all you would distribute the resources as fairly as possible. The big question is, which of the two groups collected the most resources in total? This research will use neuro-evolution where agents are simulated in a virtual environment and controlled by so called Markov Brains. Group level selection regimes such as the ones described above will be tested with respect to their effect on individual and group level performance.

## Background

Societies are depending on the collaboration of their members. This collaboration is the basis of the economy, health care system, and education among others. Obviously, such institutions benefit from everyone contributing to them via taxes. However, since they are carried by the society, individuals who for example evade taxes can still benefit from them while also benefiting from the resources they did not contribute due to cheating.

To study this problem social science, experimental economics, and game theory use the "Public Goods Game." [citation] In this game, the participants are given equal amounts of money. They then can decide to withhold this money (defect) or contribute (cooperate) into a public pot. The total contribution of the all participants will be multiplied by a synergy factor. This synergy represents gains that can only be achieved by pooling resources. The now larger amount of money in the pot will then be divided equally among all the players - regardless of them having contributed or not in the first place. Considering the game rules, a player who is careful about the social benefits; will contribute money, and those who are selfish and only consider their personal benefits will not. The scenario that individuals only think about their gain exploiting the contribution of others is called "Tragedy of the Commons". [citation]

The question is now how people can be incentivized to be collaborative? In other words, how we can avoid the tragedy of the commons?

If we see it from a government perspective, we would consider incentives to motivate people or punishment to avoid selfish approaches. But for a broader historical perspective, we have to ask where cooperation within groups comes from initially? We find that Darwinian Evolution Theory also struggles to explain why organisms cooperate [Nowak five rules for the evolution of cooperation nature 2006 article], regardless organisms, including humans, evolved to cooperate. To discuss how natural selection is the solution for the collaboration problem, let's first examine how we humans in agriculture select the best genes for reproduction. In this example, a farmer, based on a certain criteria (such as the biggest plants) selects the desired ones for reproduction.



First Generation

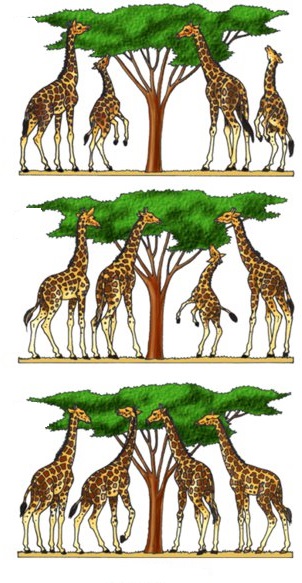
Second Generation

Third Generation



Figure 1. Human selection

In the above picture, the farmer harvests only big cabbages in the last generation by choosing only the biggest for reproducing. Nature does the same to all organism, and only lets some of them survive through history, and the rest will become extinct.



First Generation

Second Generation



Third Generation

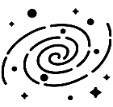


Figure 2. Natural selection

So, nature acts as the farmer and lets only the tall giraffes survive and produce offspring as they had an advantage in eating leaves on high trees. We call this *natural selection*. In the context of natural selection, the individuals' ability to survive and reproduce is called *fitness (w).* You may ask how it comes that we have a variety of different species if only the fittest survives? Then all organisms should look like each other. The answer lies within the two other concepts of mutation and selection. Mutation means sudden and significant changes in the species genetics that results in offspring being different from their parents. Suppose we take a family of giraffes and divide them into two distinct islands which one of them has tall trees, and the other has short trees. If there were no mutation in nature, regardless of the islands' differences, the giraffes after generations would not be different while both groups are from the same family. The mutation causes significantly taller and shorter offspring in the two groups. In the short tree island, the smaller ones will survive and reproduce, and the opposite happens on the other island. Eventually, after a long time by natural selection and mutation, we will have two types of giraffes (short and tall) from the same routes. Consequently, the natural environment presents organisms with different ways to be the fittest, called niches. Those niches, together with inheritance, variation, and natural selection leads to the biodiversity we observe today. Darwin's *evolution* is, therefore, a mechanism composed of inheritance, variation (random mutation), and selection. It thus leads to an adaptation of organisms that fit their environment better.

However, evolution is inherently selfish, as its rewards happen on a short time scale. Reaping rewards now, and making more offspring immediately, will always outcompete saving rewards and reproducing later. Cooperation needs the investment of multiple partners, but from the tragedy of the commons, we know that the defectors fair better. Still, we find different biological mechanisms that allow cooperation to evolve. Among others (kin selection, green beard effect, reciprocity [Nowak 2006]) the one that most likely gave raise to multicellularity is group level selection [citation]. Here, not the individual reproduces, but the entire group benefits from the rewards the entire organism receives. Thus, the most prevalent choice for incentivizing cooperation is group level selection.

Since evolution already solved the problem of cooperation before, in this study, I would like to optimize groups of agents (robots) using a genetic algorithm to cooperate with each other. These agents are controlled by evolvable Neural Networks, specifically Markov Brains [Markv Brains technical introducton 2017 or 2018]. Previous work has already performed preliminary research in this direction. The MABE (Modular Agent Based Evolution) Framework [cite MABE] to run computational experiments has been developed before, and group level selection reward functions has been tested before. I will build in these systems, and test different forms of group level selection, as well as new payoff schemes seeking to improve cooperation. This will be compared to none group level selection.

## Purpose

This research seeks to improve the way we train groups of AI controllers (agents) to perform better individually and in teams at the same time. While this is a basic research question in optimization of neural networks using genetic algorithms, it has direct applications to robotics and other autonomous AI decision making systems that need to work in groups.

As explained in the introduction, the thesis is in the sequence of other works using Markov's brain. Based on the literature review on other works in this chain of researches and also the other similar studies, I decided to work on the impact of incentives or rewarding schemes on team working of the AI controllers. According to the literature review, group level selection has been shown to improve cooperation within groups of agents. Here I introduce a new set of fitness criteria where the effort of the group is not measured by average performance of the group, but instead by either the worst performer or the best performer. This work on *minimum* and *maximum* reward schemes fills a particular *research gap* as it has not been tested before.

## Literature review

# Material and Methods

This research will use the MABE (C++ Modular Agent Based Evolution Framework [4]) to implement virtual test environments. Agents are controlled using Markov Brains [5] which are a particular evolvable type of neural network. After replicating evolutionary experiments are completed, data will be analyzed and visualized.

## Experiments

Simple example

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  |  |  | |  |  |  |  |  | |  |  |  | |  |  |  |  |  | |  |  |  | |  |  |  |  |  | |  |  | Icon  Description automatically generated | | Icon  Description automatically generated | Icon  Description automatically generated | Icon  Description automatically generated |  |  | |  |  |  | |  |  |  |  |  | |  |  |  | |  |  |  |  |  | | Turn 1 Scores | | | | | | | | | | A1 0 | | | A2  0 | | A3  0 | | A4  0 | | | |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  |  |  | |  |  |  |  |  | |  |  |  | |  |  |  |  |  | |  |  | Icon  Description automatically generated | |  | Icon  Description automatically generated |  |  |  | |  |  |  | |  |  |  |  |  | |  |  | Icon  Description automatically generated | |  |  |  | Icon  Description automatically generated |  | |  |  |  | |  |  |  |  |  | | Turn 2 Scores | | | | | | | | | | A1  1 | | | A2  0 | | A3  1 | | A4  0 | | | |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  |  |  | |  |  |  |  |  | |  | Icon  Description automatically generated |  | |  | Icon  Description automatically generated |  |  |  | |  |  |  | |  |  |  |  |  | |  |  | Icon  Description automatically generated | |  |  |  | Icon  Description automatically generated |  | |  |  |  | |  |  |  |  |  | |  |  |  | |  |  |  |  |  | | Turn 3 Scores | | | | | | | | | | A1  1 | | | A2  0 | | A3  2 | | A4  0 | |   First generation’s genetics: A1: G1 A2: G2 A3: G3 A4: G4  IDs: A1: 1 A2: 2 A3: 3 A4: 4 |
| |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  |  |  | |  |  |  |  |  | |  |  |  | |  |  |  |  |  | |  |  |  | |  |  |  |  |  | |  |  | Icon  Description automatically generated | | Icon  Description automatically generated | Icon  Description automatically generated | Icon  Description automatically generated |  |  | |  |  |  | |  |  |  |  |  | |  |  |  | |  |  |  |  |  | | Turn 1 Scores | | | | | | | | | | A1 0 | | | A2  0 | | A3  0 | | A4  0 | | | |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  |  |  | |  |  |  |  |  | |  |  |  | |  |  |  |  |  | |  |  |  | | Icon  Description automatically generated |  | Icon  Description automatically generated |  |  | |  |  |  | |  |  |  |  |  | |  | Icon  Description automatically generated |  | |  | Icon  Description automatically generated |  |  |  | |  |  |  | |  |  |  |  |  | | Turn 2 Scores | | | | | | | | | | A1 0 | | | A2  1 | | A3  0 | | A4  1 | |   Second generation’s genetics: A1: G1\*V1 A2: G3\*V2 A3: G3\*V3 A4: G3\*V4  IDs: A1: 5 A2: 6 A3: 7 A4: 8 | |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  |  |  | |  |  |  |  |  | |  |  |  | |  |  | Icon  Description automatically generated |  |  | |  |  |  | |  |  |  |  |  | |  | Icon  Description automatically generated |  | |  | Icon  Description automatically generated |  |  |  | |  |  |  | |  |  |  | Icon  Description automatically generated |  | |  |  |  | |  |  |  |  |  | | Turn 3 Scores | | | | | | | | | | A1 0 | | | A2  1 | | A3  0 | | A4  2 | | |
| |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  |  |  | |  |  |  |  |  | |  |  |  | |  |  |  |  |  | |  |  |  | |  |  |  |  |  | |  |  | Icon  Description automatically generated | | Icon  Description automatically generated | Icon  Description automatically generated | Icon  Description automatically generated |  |  | |  |  |  | |  |  |  |  |  | |  |  |  | |  |  |  |  |  | | Turn 1 Scores | | | | | | | | | | A1 0 | | | A2  0 | | A3  0 | | A4  0 | | | |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  |  |  | |  |  |  |  |  | |  |  |  | |  |  |  |  |  | |  |  | Icon  Description automatically generated | | Icon  Description automatically generated | Icon  Description automatically generated | Icon  Description automatically generated |  |  | |  |  |  | |  |  |  |  |  | |  |  |  | |  |  |  |  |  | |  |  |  | |  |  |  |  |  | | Turn 2 Scores | | | | | | | | | | A1 1 | | | A2  1 | | A3  1 | | A4  1 | |   Third generation’s genetics: A1: G3\*V2 \*V5 A2: G3\*V4\*V6 A3: G3\*V4\*V7 A4: G3\*V4\*V8  IDs: A1: 9 A2: 10 A3: 11 A4: 12 | |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  |  |  | |  |  |  |  |  | |  | Icon  Description automatically generated |  | | Icon  Description automatically generated |  |  | Icon  Description automatically generated |  | |  |  |  | | Icon  Description automatically generated |  |  |  |  | |  |  |  | |  |  |  |  |  | |  |  |  | |  |  |  |  |  | |  |  |  | |  |  |  |  |  | | Turn 3 Scores | | | | | | | | | | A1 1 | | | A2  2 | | A3  1 | | A4  1 | | |

**Game settings:**

Variable settings:

|  |  |
| --- | --- |
| Reward mode | Meaning |
| 0 | Individual reward |
| 1 | Mean score (What they got on average) |
| 2 | Maximum score (What their best performer received) |
| 3 | Minimum score (What their worst performer received) |

|  |  |
| --- | --- |
| Group mode | Meaning |
| 0 | Clone |
| 1 | Four different organisms |

Constant parameters:

Other 11 points such as field's dimension, …

## Data description

### LOD.csv

|  |  |
| --- | --- |
| Column name | Explanation |
| Generation | Generation number |
| ID | ID of each agent |
| Score | This field based on rewarding scheme value has the following meanings:   |  |  | | --- | --- | | Reward mode | Score Meaning | | 0 | Individual score | | 1 | Mean score of the four agents | | 2 | Max score of the four agents | | 3 | Minimum score of the four agents | |
| rawScores | Scores of all the group members |
| ownScore | The own score of the agent which is selected in the line of decent |

### Movement.csv

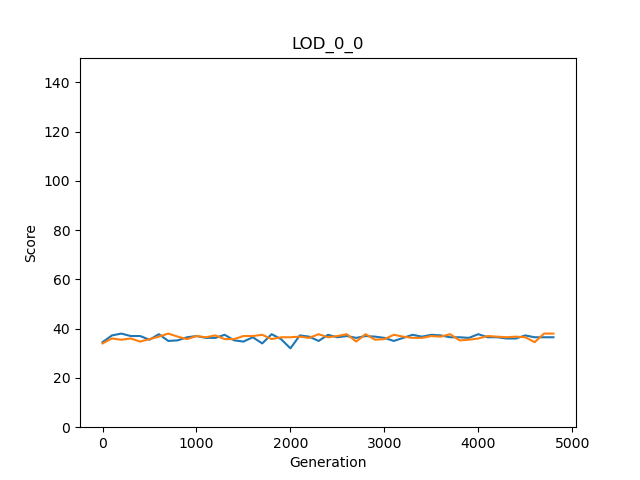
|  |  |  |
| --- | --- | --- |
| Column name | Explanation | |
| T | Turn number | |
| X | Position in X axis | |
| Y | Position in Y axis | |
| D | Direction | Meaning |
| 0 | Up |
| 1 | Right |
| 2 | Bottom |
| 3 | Left |
| E | Number of grasses that the agent gathered. It can be named as energy as well. | |
| A | Action type | Meaning |
| 0 | Do nothing |
| 1 | Turn left |
| 2 | Turn right |
| 3 | Move forward |
| 4,5,6,7 | giving grasses to an agent or putting it on a tile. |
| B | Beep (yes or no) | Meaning |
| 0 | No |
| Any other number | Beep |

### Beep.csv

## Methods

### LOD analyzer

Testing: using 0.25% of data for LOD\_0\_0\_0 and LOD\_0\_0\_1

Graph test (step 3)

LOD statistics test (step 4)

Step 4.1

Firstly, we have checked if we calculated the minimum, maximum, sum, and average of each row correctly. For this case, we tested two of the rows from LOD\_0\_0\_0 and LOD\_0\_0\_1 manually.

Step 4.2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| rewardGroupModes | meanOwnScores | meanMinimums | meanMaximums | meanAverages | meanSums |
| LOD\_0\_0 Replicate 0 & 1  (Excel) | 36.41326531 | 25.83673469 | 48.3775510 | 36.413265 | 145.653061 |
| LOD\_0\_0  Replicate 0 & 1  (Python) | 36.41327 | 25.83673 | 48.37755 | 36.41327 | 145.6531 |

### Movement analyzer

Testing:

First, we should make sure that the percentages for each movement file are calculated correctly. For this proposal, we selected two of the files and compared the results using python and excel.

The next testing part is about the averages, which again we compared python and the Excel results.

|  |  |  |  |
| --- | --- | --- | --- |
| Reward scheme and  group mode | Mean of *giving or putting grasses* percentage | Mean of *do nothing*  percentage | Mean of *beep* percentage |
| Movement\_1\_1 Excel | 0.002 | 2.0775 | 21.978 |
| Movement\_1\_1  Python | 0.002 | 2.0775 | 21.978 |
| Movement\_2\_1  Excel | 2.006 | 0.8965 | 21.8615 |
| Movement\_2\_1  Python | 2.006 | 0.8965 | 21.8615 |

### Beep analyzer

# Results

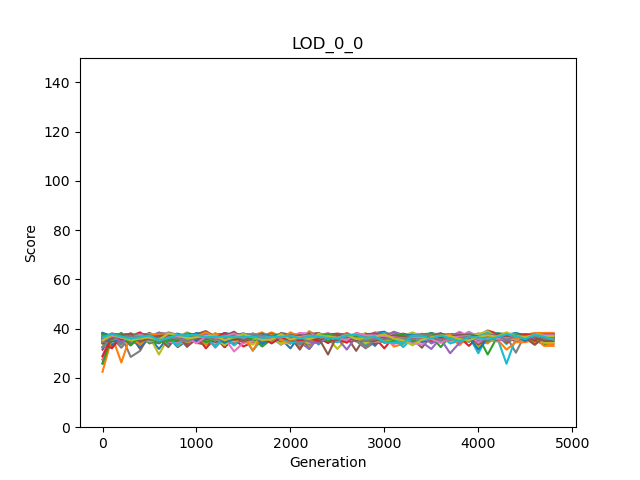
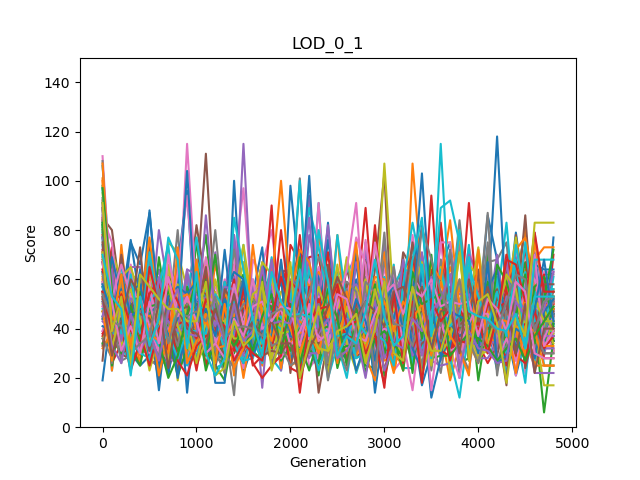
The question to be addressed here is about payoff distribution within groups. Specifically, groups of cloned agents were compared to groups composed of randomly selected agents from the population. Secondly, the payoff scheme was varied: Distributing the mean payoff the group obtained to everyone (mean), identifying the payoff the best performer in the group had and rewarding all agents of the group with that performance (max), and lastly, awarding each group member according to the least performer in the group (min). While group level selection, as present in clonal groups, is believed to improve cooperation, here additionally, payoff according to the worst (min) or best (max) performer is tested, with the hope that it might reveal an even better incentive for cooperation.

For this purpose all conditions were tested using an evolutionary computational model, testing each condition in 100 replicate evolutionary runs. The results of these experiments are organized into five sections: \_\_ , \_\_ , \_\_, \_\_, \_\_,

## How agents adapt over the course of evolution

Agents were created randomly at the beginning and had 5000 generations to evolve given the different experimental conditions. Over the course of evolution, their performance is expected to improve according to the selection criteria.

Individual | Clone

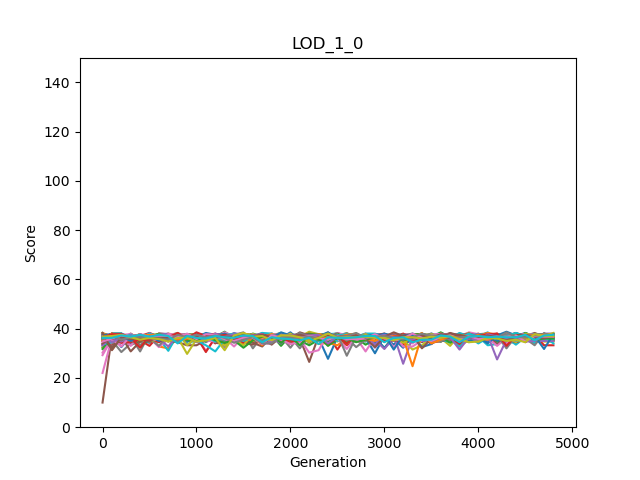
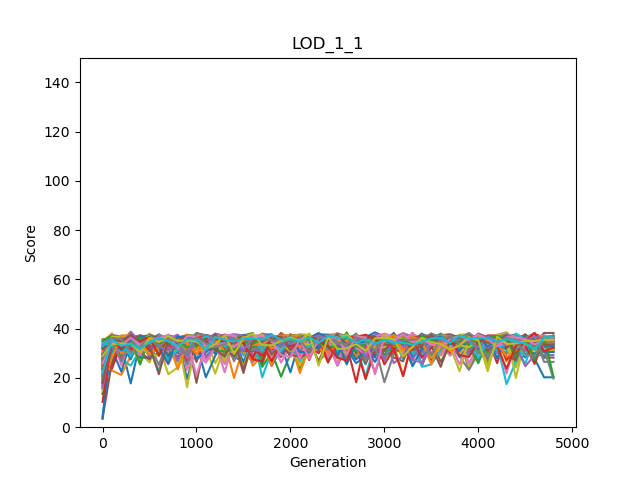


Individual | Not Clone

Figure 3. Score variations of the individual rewarding scheme

When rewarding individuals according to the amount of gras collected by themselves, and agents are clones, I find little improvement of performance and the performance remains roughly constant over the course of evolution (see Figure 3 left). When the group is not composed of clones little performance improvements are observed, but performance varies highly during evolution (see Figure 3 right). Comparing these two different scenarios suggests that the not-clonal group evolves better performance than rewarding clones. Additionally, higher variance during evolution seems to have a beneficial effect on performance but this needs to be explored further comparing more conditions (see XYZ).

Mean | Clone

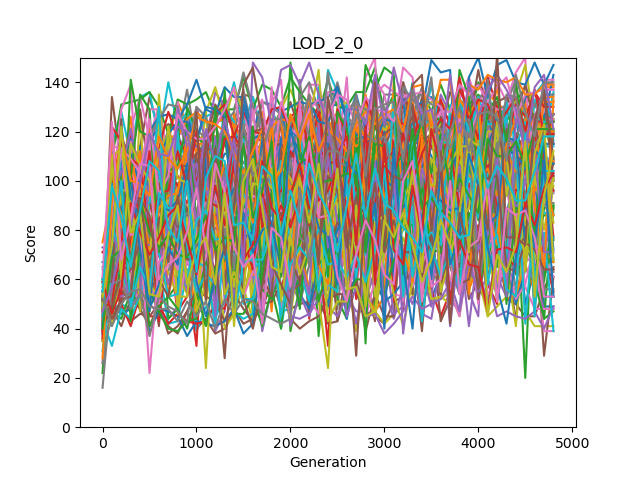
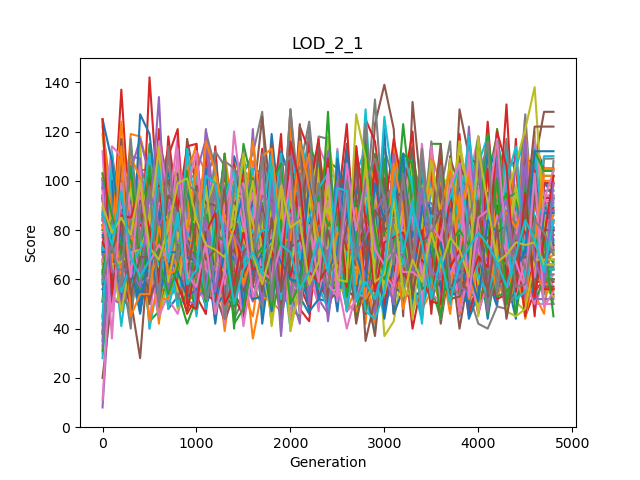


Mean | Not Clone

Figure 4. Score variations of the mean rewarding scheme

When rewarding the group members not according to their own performance, but according to the average performance of the group, the results differ. While clonal groups again show little improvement over 5000 generations, the not clonal groups show some improvement but both conditions have little variation during adaptation (see Figure 4 left and right).

Maximum | Clone

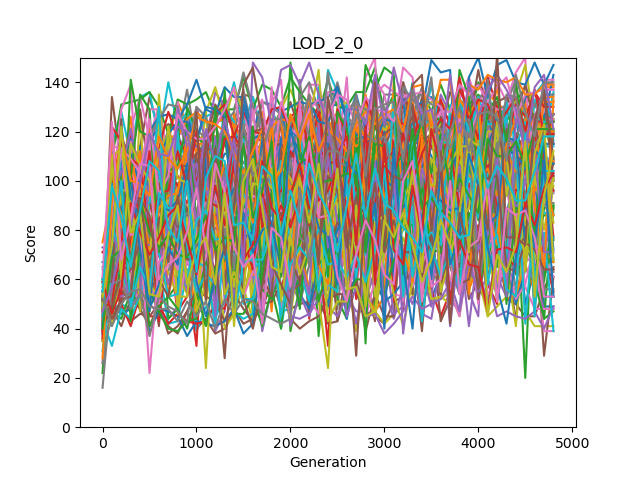
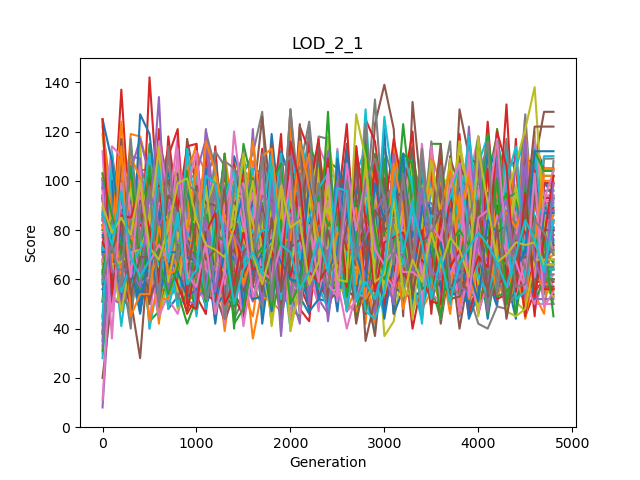


Maximum | Not Clone

Figure 5. Score variations of the maximum rewarding scheme

The max reward scheme identifies the individual of the group who collected the most grass, and now rewards all agents in the group according to that maximum performance. Thus, the group should evolved a strategy that pools resources on one individual or at least avoid that individuals compete about resources. Both, clonal and not clonal groups show that their performance improves over generations while also showing a high degree of variance throughout (see Figure 5 left and right). The clonal group seems to have a higher variance as well as a higher performance in the end.

Minimum | Clone



Minimum | Not Clone

Figure 6. Score variations of the minimum rewarding scheme

The last condition, identified the individual who collected the least grass and then rewards all group members according to that performance. This reward scheme should prevent agents to not participate in collecting grass at all, as that would result in a payoff of zero. At the same time, not one individual should collect overly much grass, instead agents should devise a strategy to keep the amount of grass collected by everyone equal. Like in the maximum rewards scheme, we find an improvement in performance of 5000 generations, as well as a high variance within the adaptive periods (see Figure 6 left and right).

In the next section, the performance of agents after evolution will be compared. Further, the observation that some conditions leads to high variances in performance, while others do not, seem interesting as well, and should be investigated further. Is it possible that those evolutionary runs that show a high variance explore more options and thus have a higher chance to find better solutions?

## Comparison of the different scenarios

As we have discussed in the experiment section, we kept all the parameters constant except the rewarding scheme and how groups are composed (clone or not clonal) resulting in eight different experimental conditions. We expect different experimental conditions to affect individual behavior and how the groups as a whole performs. Thus the criteria compared are the average payoff of the individual across all 100 replicate experiments with the same condition. Further, the average performance of the group member who collected the least (mean minimum) and the most (mean Maximum), the average amount the groups collected (mean Averages), and finally the average of what groups collected in total (mean sums).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| rewardGroupModes | Mean  OwnScores | Mean  Minimums | Mean  Maximums | Mean  Averages | Mean  Sums |
| LOD\_0\_0  Group mode: Clone  Reward mode: Individual | 36.4275510 | A picture containing shape  Description automatically generated26.0175510 | 47.4516326 | 36.4275510 | 145.710204 |
| LOD\_0\_1 3  Group mode: Not Clone  Reward mode: Individual | 44.7497959 | 20.1351020 | 52.0906122 | Logo  Description automatically generated36.5244898 | Logo  Description automatically generated146.097959 |
| LOD\_1\_0 2  Group mode: Clone  Reward mode: Average | 36.4372959 | Logo  Description automatically generated25.8336734 | 47.5702040 | A picture containing shape  Description automatically generated36.4372959 | A picture containing shape  Description automatically generated145.749183 |
| LOD\_1\_1  Group mode: Not clone  Reward mode: Average | 47.0285714  Logo  Description automatically generated | 11.6457142 | Logo  Description automatically generated56.7748979 | 33.9260204 | 135.704081 |
| LOD\_2\_0  Group mode: Clone  Reward mode: Maximum | 91.9987755  Logo  Description automatically generated | 2.70897959 | Logo  Description automatically generated91.9987755 | 30.8365306 | 123.346122 |
| LOD\_2\_1  Group mode: Not clone  Reward mode: Maximum | 67.5675510  **A picture containing shape  Description automatically generated** | 1.56244898 | **A picture containing shape  Description automatically generated**75.5930612 | 31.0185204 | 124.074081 |
| LOD\_3\_0 1  Group mode: Clone  Reward mode: Minimum | 30.7763265 | Logo  Description automatically generated30.7763265 | 42.3973469 | Logo  Description automatically generated36.4311734 | Logo  Description automatically generated145.724693 |
| LOD\_3\_1  Group mode: Not clone  Reward mode: Minimum | 34.1008163 | 23.0051020 | 44.5406122 | 33.0433673 | 132.173469 |

Table . Comparison of the different scenarios. Observe that "own score" does not imply the amount of grass collected, but what that agent was awarded due to the reward scheme. For example, and agent might have collected more grass than it received as a reward in the minimum scheme, since another agent collecting less, defined that reward.

There are five measurements for each scenario in the above table, and they are ranked using gold, silver, and bronze medals. The performance of individuals is different from who well each member in a group fairs. Thus, the average minimal performance could be understood as an indicator for the entire group. Obviously, the reward that every group member receives will always be equal or better than the amount of grass the worst performer collected. Similarly, the total performance (Mean Averages, and Mean Sums) of the group reflects how well groups work together.

With respect to the total number of grass collected (Mean Averages, and Mean sum), we find that the not clonal group with individual payoff fairs best, followed by clonal groups with average payoff. Interestingly, the third best performance is achieved by using clonal groups awarded by the minimum. This confirms that this reward scheme is indeed capable of driving groups to perform optimally.

However, when considering how well every member of the group performs, we have to consider the mean Minimum payoff (which is an individual payoff). The clonal minimal rewarding scheme optimizes this parameter the best (place 1 Mean minimum) followed by Clonal groups with individual payoff, and Clonal groups with average payoff. While maybe not surprising that the reward scheme that selects for highest minimal payoff optimizes this parameter the best, it still confirms that this reward scheme is very capable of not only optimizing the least amount each agent collects, but more importantly it also compares very well on the total number of grass collected within the entire group (third place in Mean Averages and Mean sum).

When ignoring how well the group performs, but focusing on a single individual to collect the most grass, we find the clonal maximum reward scheme to perform best, followed by not clonal groups maximum, and not clonal groups with average payoff. Again, this might not be surprising as the reward function selects for a single group member to perform optimally even incentivizing others to give their collected grass to said best performer.

To sum it up, when we select clone agent and set the reward scheme as minimum, we get a high average score, while ensuring that all group members also fair well. In other words, this combination *trains* our agents to *perform better both individually and in teams at the same time*!

## Comparison of clone and not clone groups

This subsection will compare the clones and not clone agents' performance in different rewarding schemes.

Figure . comparison of clone VS Not clone modes in different rewarding schemes based on the mean of minimum scores

Based on the above figure, the minimum scores are always higher in clone mode relative to different individuals. So, if we are interested in having good minimum scores, we should use clone mode.

Figure 8. comparison of clone VS Not clone modes in different rewarding schemes based on the mean of maximum scores

Except for the maximum rewarding scheme, the top score is higher when we have different agents. Therefore, if we do not consider setting a rewarding scheme, various agents are better than clones for maximum score.

Figure 9. comparison of clone VS Not clone modes in different rewarding schemes based on the mean of own scores

As in the previous graph, except for the maximum rewarding scheme, the own score is higher when the agents are not clone. Therefore, if we do not consider setting a rewarding scheme, different agents have better performance in this case.

Figure 10. comparison of clone VS Not clone modes in different rewarding schemes based on the mean of average scores

There is no clear pattern; however, when clone modes have more significant averages, the differences are more prominent. So, we can say that for average, the clones are better.

Figure 11. comparison of clone VS Not clone modes in different rewarding schemes based on the mean of the sum of scores

Again, there is no clear pattern, and when clone modes have a bigger sum, the differences are more significant. As the average for total sum also clones are slightly better.

In this section's first four figures, the bars separated using rectangles show the scores when the rewarding scheme and score are matched. For example, in the first figure, the vertical axis is for the mean of the maximum score, and the two bars that are for the maximum rewarding scheme are isolated from others using a rectangle. If we check these bars, we can see that clones, in most cases, more carefully listen to us and follow the rewarding scheme relative to different agents.

## Agents' behaviors

In the comparison of the different scenarios, we have found the best three combinations that were 1. Clone and minimum, 2. Clone and average, and 3. Not clone and individual. In this part of the results, we will check which actions are the reason behind success or failure in our agents' performance. We have worked on three actions that are explained in the next table.

|  |  |
| --- | --- |
| Action name | Explanation |
| Beep | Beep is simply the way of our agents' communication. It can have different values. When it is zero, they do not use a beep, and the other values have their meaning. In this thesis, I only worked on two statuses that are using or not using the beep. |
| giving or putting grasses | We have allowed the agents to give their energy (grass) to other agents or simply put it on an empty tile. In this study, we have considered both cases as one action. |
| do nothing | As its name suggests, it means that the agent does nothing in its turn. |

Table 2. Agents' actions definitions

|  |  |  |  |
| --- | --- | --- | --- |
| Reward scheme and  group mode | Mean of *giving or putting grasses* percentage | Mean of *do nothing*  percentage | Mean of *beep* percentage |
| Individual  Clone | 0 | 0.7005 | 13.4225 |
| Individual 3  Not clone | 0 | 0.136 | 24.147 |
| Average 2  Clone | 0 | 0.2965 | 17.156 |
| Average  Not clone | 0.002 | 2.0775 | 21.978 |
| Maximum  Clone | 24.948 | 3.304 | 53.814 |
| Maximum  Not clone | 2.006 | 0.8965 | 21.8615 |
| Minimum 1  Clone | 0 | 0.963 | 51.2315 |
| Minimum  Not clone | 0 | 8.1085 | 9.5385 |

Table 3. The Agents' actions

I have highlighted the top three values for each action in the above table. Two of the best combinations (reward scheme and group mode) are matched with the highest values of beep usage, and also the third one is for maximum and clone with two gold medals. So we can say, using beep has a positive impact on agents' performance. Minimum and not clone is our worst combination without any medal, and it has the highest value of do nothing so, we can conclude that doing nothing is not a good strategy. However, we should also consider that this action's second-highest value belongs to maximum and clone agents with the maximum score's gold medal. Therefore, do-nothing action is the right way to have a leader with a very high score. Because the percentage of giving the energy to other agents or putting it on an empty tile is way bigger in case of maximum and clone, it is a useful action if they want to have an agent with a high score. To sum it up in general, usage of beep is a positive action, do nothing and giving or putting the energy can only be useful when we want to have a team leader.

## Do agents communicate for a reason?

In the previous section, we have identified that the usage of beep, in general, has a positive impact on our agents' performance. We wanted to double-check our observation, and therefore, we have done two experiments in a way that agents were allowed to use beep in one of them, and in the other one, they were muted.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Reward  And Group | Beep  Mean Minimums | Mute  Mean Minimums | Beep  Mean Maximums | Mute  Mean Maximums |
| Individual  Clone | 26.09 | 25.24 | 47.19 | 46.11 |
| Individual  Not clone | 26.958 | 26.954 | 47.53 | 47.43 |
| Mean  Clone | 26.14 | 25.27 | 47.45 | 46.20 |
| Mean  Not clone | 25.41 | 25.45 | 46.97 | 46.99 |
| Maximum  Clone | 1.56 | 3.74 | 92.88 | 34.00 |
| Maximum  Not clone | 25.13 | 25.57 | 46.89 | 47.41 |
| Minimum  Clone | 31.51 | 8.72 | 41.49 | 15.19 |
| Minimum  Not clone | 23.72 | 23.46 | 43.85 | 43.47 |

Table 4. Beep vs. mute for the mean of the minimum and maximum scores

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Reward  And Group | Beep  Mean Averages | Mute  Mean Averages | Beep  Mean Sums | Mute  Mean Sums |
| Individual  Clone | 36.40 | 35.40 | 145.61 | 141.63 |
| Individual  Not clone | 36.97 | 36.96 | 147.88 | 147.84 |
| Mean  Clone | 36.52 | 35.46 | 146.10 | 141.86 |
| Mean  Not clone | 35.88 | 35.92 | 143.55 | 143.69 |
| Maximum  Clone | 30.00 | 16.24 | 120.03 | 64.97 |
| Maximum  Not clone | 35.75 | 36.21 | 143.03 | 144.87 |
| Minimum  Clone | 36.46 | 11.89 | 145.84 | 47.56 |
| Minimum  Not clone | 33.44 | 33.15 | 133.77 | 132.61 |

Table 5. Beep vs. mute for the mean of the average and sum scores

The last two tables are the results of beep and mute experiments. In each comparison, I have highlighted the greater value. For instance, in individual and clone for the mean of minimums, those who had beep option had better performance relative to muted agents. Based on these comparisons in general, beep agents has better scores than mute ones in all the measures (minimum, maximum, average, and sum)

# Discussion

The argument of satisfying the initial thesis aim is the first point that I will elaborate in this section. As I have explained in the introduction, we want to improve the way we train groups of AI controllers (agents) to perform better *individually* and in *teams*. To find a solution, we have defined and implemented a game for a group of AI agents, and we tested their performance in different scenarios. The details of the game and its setting parameters are explained in the material and methods; however, the experiments' variable parameters are group mode and rewarding scheme. Group mode can be set to either clone (which the agents are identical) or not clone (with different agents). The rewarding scheme is the tool for changing the team-working approach of our agents. It has four types: a) *minimum* that they should help the weakest member, *b) maximum* which everyone supports the leader, *c) average* when they must perform to have a good mean and final *d) Individual* which they think about their payoff. We have two variables; the first one has two options and the second one four; therefore, we will have eight combinations, which we call the game scenarios. The agents' performances in the eight scenarios are measured and evaluated based on *a) minimum score* indicating the lowest score among the group members, *b) maximum* for the highest score*, c) an average* of agents scores, and *d) total payoff,* which is the sum of all the agents' scores. Based on our observation, when the agents are clone, and the rewarding scheme is set to minimum, firstly the minimum score is the highest among all the combinations, and secondly we have a good score for average and total payoff. Therefore, using this combination (minimum and clone) we are sure that all the agents have a relatively high score, which means each individual has a good performance and also at the same time the total payoff is high. So, we could find a method that caused our AI controllers (agents) to perform better individually and in teams. However, we should mention that this conclusion is based on our experiments, and to generalize it, we should do more experiments using different settings such as a different brain or game. Also, we do not want to simulate human collaboration using AI and find a solution for human collaboration conflicts, and it is just about the Markov brain with its specific features.

If agents should try different methods through their evolutionary journey? Based on our observation, a high variance of performance during evolution correlates with high performance in the end. So, we can interpret it as proof of testing different methods that positively impact their performance because, when their scores are very different in different generations, they have probably tested multiple ways of collaborating. This conclusion is clearer in the case of individual rewarding schemes that clone agents stick to a score of 40 from the beginning, whereas not clones score varies from almost zero to 120 (see figure number 3). As we mentioned in the first table in the individual rewarding scheme, the not clones have way better performance. So maybe the not individual are braver and test more different methods which lead to their success. Indeed, we only have the variations in the scores, but we should check their behavior in detail on future works to make the conclusion.

If you check our results, you can see that the groups of clone agents have better performance than the groups that have different agents. This higher performance is because clones have the same logic, and therefore, all of them at the same time will find the best solution, whereas in different agents, it takes longer time, and maybe one of some of them never find the best method. However, the individual rewarding scheme is the exception, and the not clones are better. Probably it is because the clones must always take care of each other, although the rewarding scheme does not suggest it. Identifying the reasons behind this type of unexpected behavior needs more studies, which is considered one of the thesis's future works.

At this point, I will explain the agents' behavior in different scenarios. We have three actions for the agents that are beep, giving or putting grass, and do nothing. Beep is the tool of communication for the agents; they can give their grass (energy) to other agents or put it on an empty tile and just do nothing in their turn. According to our experiments' results, in general, usage of beep is a positive action; doing nothing and giving or putting the energy can only be useful when they want to have a team leader. So, the beep for all the cases is positive; for example, either we want a higher maximum or minimum; if they use the beep, it causes better outcomes. We expected it because communication is one of the main reasons for group work success. We did more experiments and compared their performance when they are muted and when they can talk, and that comparison proved the effectiveness of the communication (beep). This beep can have various values, and we only checked it as a binary variable that zero means silence and any other value means talking. Coding the beep value is done by Markov's brain and understanding this language or decoding these variable values is fascinating research that we considered future works. As I mentioned before, doing nothing and giving or putting the energy is only useful for the maximum rewarding scheme. When they give the energy to someone else or put it on an empty tile so another agent takes it, they sacrifice for others' benefit. I believe this action is an instance of *altruistic behavior.* We can also consider the do nothing as an altruistic behavior because when they do not try to gather grass, others can gain more energy. So our agents have altruistic behavior, but what about the other side and selfish actions? Indeed we do not have any special action indication of selfish behavior; however, maybe there is a correlation between agents' movement and their benefits. For example, suppose the four agents initial locations are in the corners of the map; if an agent first moves to the center and gather grass to (avoid others take them) and then come back to his corner and gather the local area grass, I believe it is a kind of selfish behavior. I did not work on this aspect, and it requires a specific analysis of the agents' location during the game, which is also one of the future research works.

# Conclusions

# References

|  |  |
| --- | --- |
| [1] | L. Joel, C. Jeff and M. Dusan, "The surprising creativity of digital evolution," in *Artificial Life Conference Proceedings 14, pages 76-83. MIT Press*, 2017. |
| [2] | H. Arend and M. Masoud, "Evolution of autonomous hierarchy formation and maintenance," in *Artificial Life Conference Proceedings 14, pages 366-367. MIT Press*, 2014. |
| [3] | B. Gergely and S. Szabolcs, "Beneficial laggards: multilevel selection, cooperative polymorphism and division of labour in threshold public good games," *BMC evolutionary biology,* 2010. |
| [4] | B. Clifford and H. Arend, "Mabe (modular agent based evolver): A framework for digital evolution research," in *Artificial Life Conference Proceedings*, 2017. |
| [5] | H. Arend, A. E. Jeffrey, S. O. Randal, B. K. David, S. Jory, A. Larissa, T. S. Ali, K. Peter, S. e. Leigh och G. Heather," Markov brains: A technical introduction". |