All-for-one or One-for-all - how Differential Reward Schemes affect the Evolution of Cooperative Behavior

Clifford Bohm[1,2] and Arend Hintze[1,2,3]

[1]Department of Integrative Biology, Michigan State University, East Lansing, MI 48824
[2]BEACON Center for the the Study of Evolution in Action, Michigan State University, East Lansing, MI 48824
[3]Department of Computer Science and Engineering, Michigan State University, East Lansing, MI 48824
cliff@msu.edu

Abstract

Introduction

Collaborative behavior within groups of closely related organisms is one of the more intriguing phenomena we observe in nature. This collaborative behavior can come in many different forms. For example, insects often form colonies in which a single or few reproductive individuals (queens) extend their phenotype (Dawkins, 2016) with thousands of sterile workers who's sole pursuit is to help the queen to generate more offspring. Dominance hierarchies among vertebrates (Drews, 1993) in which resources are distributed asymmetrically provide another example (although these are less extreme, since often lower ranked individuals may reproduce). We will refer to behaviors which distribute resources unequally as *all-for-one* behaviors. The evolutionary advantage of all-for-one strategies can be explained if the long term success of a group is based on the offspring of one or a small number of individuals belonging that group.

The opposite of this behavior is seen in groups that collaborate in such a way that resources are distributed equally (or near equally) to every member of that group. Bacterial biofilms (Dickschat, 2010), symbiosis, group transport (McCreery and Breed, 2014), and collaborative hunting (for example Stander 1992) are examples of what we will call *one for all* behaviors. Often these groups preform only as well as their weakest link (bridge building in ants is one example (Reid et al., 2015)). The evolutionary advantage of these all-for-one strategies can be explained if the long term success of that group is based on the offspring of many individuals (perhaps because of a long or costly developmental processes). All-for-one here should not be confused with selfless behavior where an organism sacrifices themselves for the good of the group. We may could chosen the term all-for-all in place of one-for-all, but feel that one-for-all conveys the idea more clearly.

What evolutionary conditions lead some species to all-for-one behaviors and others to one-for-all behaviors and how do these behaviors differ? Here we test the outcome of enforcing three different selection criteria (BEST,WORST,AVERAGE, see below) on co-evaluated clonal groups and observe that these different selection methods influence not only evolved behavior, but whether that behavior is all-for-one, one-for-all, or something in between.

We used MABE (Bohm et al., 2017) to evolve digital agents controlled by Markov Brains (Edlund et al., 2011; Marstaller et al., 2013) (for summery see methods; for a detailed technical description see Hintze et al. 2017). The reproductive success of individuals in a group was based on the behavior of the best individual in the group, worst individual in the group, or the average performance of the entire group. In general, we will show that selection for BEST results in all-for-one, selection for WORST results in one-for-all and selection for AVERAGE may result in either one-for-all or all-for-one.

These results are applicable to the evolution of group or swarm robotic controllers as we demonstrate not only the ability to evolve cooperation, but that by applying different selection methods we can affect that type of strategies which result. A typical example of a one-for-all task is a search and rescue mission where success for the group should be based on the ability for one individual to locate a target. However, a swarm of agents that create a communication network (Varga, 2016) would be more efficient with a one-for-all strategies where network traffic is distributed evenly among all agents.

Materials and Methods

Populations of agents were evolved over a number of generations on a foraging task. Agents performance was evaluated in groups of eight where each agent received a score based on individual performance. Four conditions were considered. In the first three, the eight members in each group were clones, copied from a single member of the population. In the BEST and WORST conditions the score assigned to the clonal groups parent was the score of highest or lowest preforming clone in the group. In the AVERAGE condition the score assigned to the clonal groups parent was the average of the group. The fourth condition INDV was a control using non-clonal groups and each agent was assigning a score based on their own performance.

Berry World, a cooperative foraging task

Berry World is a general purpose foraging world designed for MABE. Here we describe berry world as it was configured for this paper. The world is a 2D grid of size 16×16, surrounded by an impenetrable wall. Agents have inputs that allow them to detect the contents (resources, other agents and walls) of the cell they are occupying as well as the two cells directly in front of them and the cells in front and to the left and right (i.e. five cells in total). Agents are evaluated sequentially for 1000 updates and during each update each agents inputs are set and then they may move forward, turn left or right 45 degrees, collect the resource in their current location or take no action. An attempt to move into a cell occupied by another agent or a wall will simply fail. Each cell of the grid can contain either a wall or, up to one agent and up to one food resource. Food resources come in two types which we call *red berries* and *blue berries*. Once a berry is collected from a cell and the agent leaves that cell the cell is filled with a random (red or blue) berry. Note that the only way that a cell can not have a berry is if there is an agent in that cell and that agent collected the berry already.

When agents collect a berry they receive a reward of one point for doing so. A task switching cost of 4.0 is deducted every time an agent collects a type of berry that is different from the type the agent collected last (similar to (**?**)).

Generally speaking, when the task switching cost is low, it is better to ignore berry color and maximize collection. When the cost is high, it is better to collect only one type of berry and when surrounded by the other kind either wait for another agent (who is presumably collecting the other type) to 'flip' a local resource or to preform a random search for a new spot to forage. At middling costs, it may be best to develop a strategy with limited switching. We chose a cost of 4.0 because previous work has demonstrated that this is at the high end of middling.

When non-clonal agents have been evolved in groups at high task switching cost in this environment in the past the result is usually either speciation (in particular, a red collector species and a blue collector species) or a single species which evolved to collect resources of the type which they started on. When we forced all agents to start on the same color cell, then we observed speciation in almost all cases. In these tests agents initially are placed in random cells containing red berries. Balanced populations of red and blue collector agents preformed well as red collectors focused on collecting only red berries generate higher concentrations of blue and blue collectors tend to generate higher concentrations of red. In the results we demonstrate that the tendency to speciation with non-clonal groups in Berry World is seen here as well.

For clonal groups experiencing this task the problem is significantly harder. They can not offload color choice to their genome (see below). If everyone in the group were to only consume one type of berry, that berry type would quickly be depleted due to the effects of randomized replacement. In order to thrive in this environment as a group, a better strategy then "Eat all berries of one color" must emerge.

Brain World interface

MABE (the evolutionary framework we are using to generate these results) provides methods to generate different types of worlds and brains such that any type of brain (MB, Artificial neural networks, Cartesian Genetic Programming, to name only some) can be evaluated in any world. This requires that the world state can be described as a list of numbers which a brain uses as inputs and that the brain output can also be described as a list of numbers. The "berry world" which was used here generates a world state list with 18 values (2 resource types for each of five cells and other agent and wall information for 4 cells) and expects a brain output list with 3 values (two values control movement and 1 controls berry collection).

Markov Brains Markov Brains (MB) are networks of deterministic logic gates that convert inputs inputs into outputs. In addition to gates, a Markov Brain is comprised of input values, output values, internal memory and wires which carry information between the brains components. A genome (an evolvable string of numbers) is used to encode the topology of this network as well as the function of each logic gate each time an offspring is generated. Genomes are scanned for start codons (pairs of numbers which signal the beginning of a gate gene). Some number of sites following the start codon are then read and used to describe a gate which specifies the logic of the logic gate and how it wires to inputs, outputs, and internal memory (**?**).

Population and Replication setup

Populations of 200 agents were evolved for 10,000 generations. 350 replicates were run for BEST,ALL and WORST conditions. 350 replicates were run for the INDV condition.

Reproduction and Mutations

During each generation all of the agents in the population were evaluated and assigned a score. This score was then used to select parents using a proportional selection process (i.e. roulette wheel Moran process (Moran, 1958)). The parents genomes were copied and mutated, resulting an a new generation. Genomes were initialized to have 5000 sites and could be altered by point mutations (.5% per site) and copy and deletion mutations of 128 to 512 sites (.002% per site each).

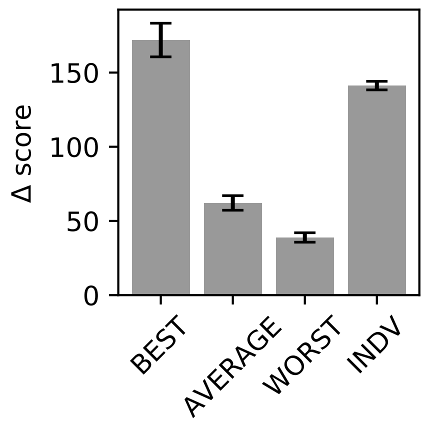

Figure 1: Shown is the average of the difference between highest scoring agent and the average of the remaining agents in that group for each condition (BEST, AVERAGE, WORST, INDV). Error-bars indicate the standard error.

Results

We found that after 350 replicates for each condition (and 50 of the INDV control condition) were run for 10,000 generations, a significant number of replicates in all but the WORST conditions achieved non-trivial results and that many replicates appear to have reached a local optima, as indicated by a plateau in fitness over time (data not show). To determine which replicates were significant, we choose a score cutoff score of 120 for the BEST and AVERAGE conditions and 30 for the WORST condition. Replicates where average score falls below the cutoff value are discarded. After culling, 66 replicates of the BEST condition, 39 of the AVERAGE, 7 of the WORST, and all 50 of the INDV remained. In many cases the discarded replicates generated trivial strategies such as agents in a group collected only one resource type.

We find that each selection regime results in different group level behavior. The BEST selection regime leads to the highest difference in collected food between the best agent and the average of the other agents in the group. This is followed by the INDV selection regime. AVERAGE and WORST show a much lower difference between the best agent in the group and the rest, with the WORST having the smallest difference (see Figure 1).

It is clear that the variation seen in BEST is significantly higher then WORST. This results from the fact that while some BEST strategies are balanced (having about the same number of red collectors as blue), many of the best strategies were biased and involve unequal distribution of resources. In these cases one or a two clones collected one color berry while the remaining group collected the other. In these cases the clones in the minority are able to find berries of their preferred color with less search. This strategy clearly can not work for WORST since the clones in the majority would incur high search costs or be required to preform costly switches. WORST must therefore rely on balanced strategies. AVERAGE has less requirements as these clones must simply collect as must resource with the least number of switches possible. Which member of the group does exactly what work is not important. AVERAGE is able to leverage both biased and balanced strategies.

Division of labor

How are the agents determining what role they should play in the group. Simply put, each agent must decide if they will collect red, collect blue, or collect some red and some blue. How do they choose what to collect and if and when they should switch? Each replicate represents a unique path of evolution and we have observed many different strategies develop in each condition. Even when we see similar behaviors evolve they often have subtle differences and as with other examples of convergent evolution, even though two organisms may appear outwardly similar, they may be structurally different. Here we do not attempt to explain all of the behavior that we have witnessed. Instead, we rely on an automatic classification methods to demonstrate that the behaviors are significantly and objectively different. With the results of the classification, we then attempt to draw broad conclusions about the identified classes.

Individual Selection and Speciation A simple and effective strategy for berry world is to have groups comprised of four red collectors and 4 blue. In the INDV condition, we observe that the organisms offload this choice to genetics. The vast majority of replicates in this condition had two qualities. First, an analysis of the phylogeny demonstrated that these populations were speciated for long periods of time (see Figure 2) and second that the agents rarely switch which color berry they are collecting. When randomly picking eight individuals from a population containing 50% red-collectors and 50% blue-collectors over 70% of possible groups (182 out of 256 possible combinations) contain at least three agents of each species. When agents employ a never switch behavior they will benefit if they are in the minority in a group. This creates a "fit when rare" situation where if one type of collector becomes more common in the population, they will on average collect less resource and since in the INDV case each agent is evaluated on their own performance, they will have lower reproductive success. Not only does genetics provide a means to offload decision making, but it is also self balancing.

Classification of Evolved Clonal Behaviors Unlike INDV, the agents in the other conditions can not rely on a genetic switch to determine their berry preferences. More over since these agents are all identical clones, they must have

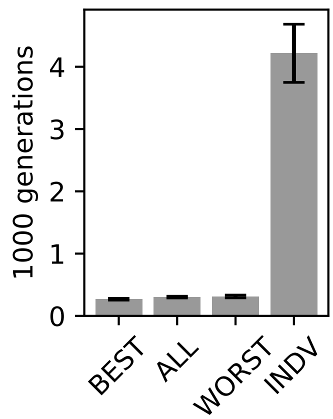

Figure 2: The average time between the last generation and the most recent common ancestor for to the members of the last generation for all four conditions (BEST, AVERAGE, WORST, and INDV). Error-bars indicate the standard error.

some mechanism which allows them to choose different behaviors or they would all execute identical strategies. Rather than subjectively attempting to determine the evolved strategies for the three clonal conditions by hand we employed k-means clustering. K-means clustering requires a feature vector describing the behavior of each replicate. For each replicate we evaluated each of the eight highest performing agents 30 times. From this we could derive the food preference for each agent (that is the type of food which that agent collected more of) and using these values, the preference ratio for each group evaluation. We called the berry type preferred by more agents, the "primary food" while the other color was the "secondary food" (i.e. a preference ration of 3/5 would indicate that 3 agents preferred one type of berry and 5 the other). Figure 3 shows a composite of all of the data collected from the clonal conditions. Note that the preference ratio shows no values at 5/3, 6/2, 7/1 to 8 since these have been remapped to 3/5, 2/6, 1/7 and 0 respectively.

From these four parameters (primary food, secondary food, switches, and preference ratio) we determine the feature vector for each evaluation. When clustering this data ($k = 5$, different numbers of clusters did not improve the results, data not shown) we can derive so called centroids for each cluster, which represent 5 idealize behaviors (see Figure 4).

all-for-one strategies

When rewarding clonal groups using only the payoff the BEST agent received, we get groups of agents that evolve an all-for-one strategy (see Figure above). However, this can be implemented in different ways. We therefore compare how -the X plots - we find -¿

Describe strategies using quant data and figures

one-for-all strategies

see above just the opposite

you have 3 pages +/- 1

Discussion

We showed that different selection regimes (AVE, BEST, WORST) can be used to control the kinds of strategies developed by evolutionary processes.

As expected, rewarding groups by BEST results in all-for-one strategies, while WORST encourages one-for-all strategies to evolve. What surprised us what the diversity of solutions that implement these strategies.

We found bla and bla, but also bla and bla. It is important to mention that agents need to use

In either case (one-for-all or all-for-one) the performance of the group might need individuals to perform different tasks to achieve the common goal. This generates a new problem. When using this selection regime in a genetic algorithm we face this question: are the individual of the groups selected clones or randomly chosen genetically distinct individuals from the population? In the later case genetic differences could be used to determine who does what. In the case of clones, this selection regime might require agents to behave differently even though every individual has the same brain. Fortunately, our computational system was already capable to overcome this issue (**?**).

Acknowledgements

References

Bohm, C., C G, N., and Hintze, A. (2017). Mabe (modular agent based evolver): A framework for digital evolution research. In *Proceedings of the European Conference on Artificial Life*, pages pp–76. MIT Press.

Dawkins, R. (2016). *The extended phenotype: The long reach of the gene*. Oxford University Press.

Dickschat, J. S. (2010). Quorum sensing and bacterial biofilms. *Natural product reports*, 27(3):343–369.

Drews, C. (1993). The concept and definition of dominance in animal behaviour. *Behaviour*, 125(3):283–313.

Edlund, J. A., Chaumont, N., Hintze, A., Koch, C., Tononi, G., and Adami, C. (2011). Integrated information increases with fitness in the evolution of animats. *PLoS Computational Biology*, 7(10):e1002236.

Hintze, A., Edlund, J. A., Olson, R. S., Knoester, D. B., Schossau, J., Albantakis, L., Tehrani-Saleh, A., Kvam, P., Sheneman, L., Goldsby, H., et al. (2017). Markov brains: A technical introduction. *arXiv preprint arXiv:1709.05601*.

Marstaller, L., Hintze, A., and Adami, C. (2013). The evolution of representation in simple cognitive networks. *Neural computation*, 25(8):2079–2107.

McCreery, H. F. and Breed, M. (2014). Cooperative transport in ants: a review of proximate mechanisms. *Insectes Sociaux*, 61(2):99–110.

Moran, P. A. P. (1958). Random processes in genetics. In *Proceedings of the Cambridge Philosophical Society*, volume 54, page 60.

Reid, C. R., Lutz, M. J., Powell, S., Kao, A. B., Couzin, I. D., and Garnier, S. (2015). Army ants dynamically adjust living bridges in response to a cost–benefit trade-off. *Proceedings of the National Academy of Sciences*, 112(49):15113–15118.

Stander, P. E. (1992). Cooperative hunting in lions: the role of the individual. *Behavioral ecology and sociobiology*, 29(6):445–454.

Varga, M. (2016). Fixed-wing drones for communication networks.

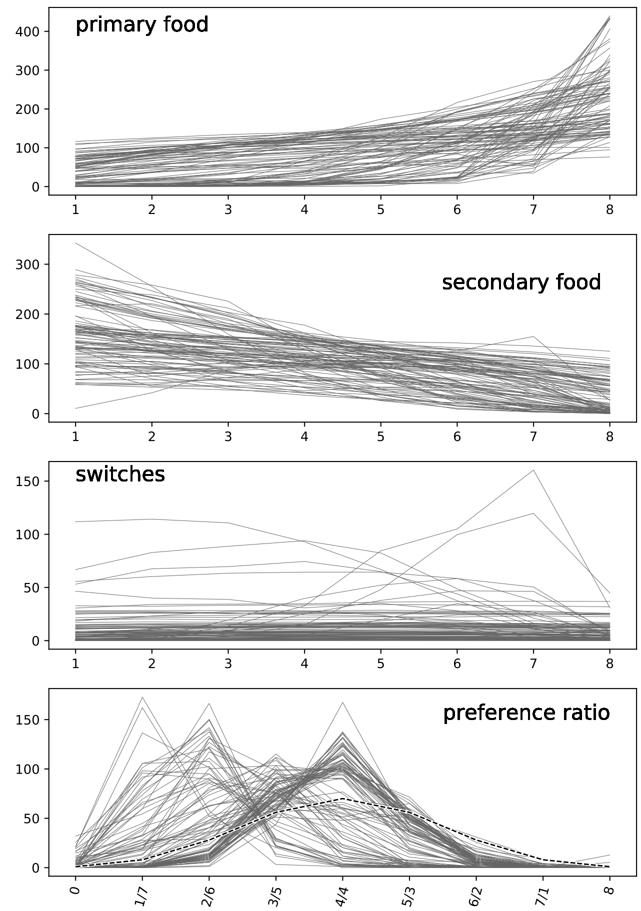

Figure 3: Overview of behavior for all replicates. The top two panels show the primary and secondary food collection for each agent in the group for all replicates. The third panel shows the number of switches for each agent in the group for all replicates. The bottom panel shows the distribution for each replicate of how many agents in the group prefer the primary food over the secondary. Observe that the primary food collected is preferred by only a few agents in the group

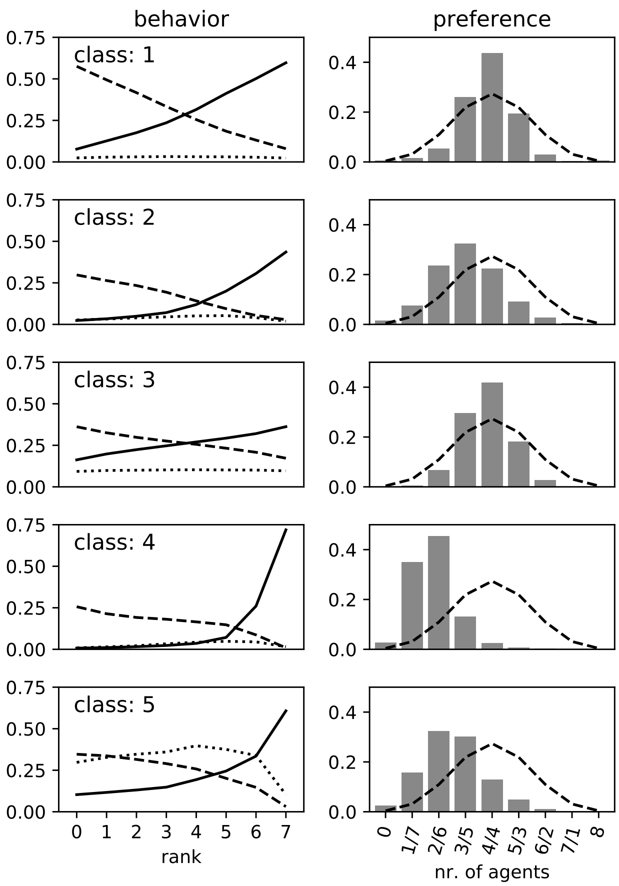

Figure 4: The five classes of strategies as a result of k-means clustering. Each row describes the characteristic feature vector of each strategy class (centroid). The left column (behavior) shows normalized values for primary food collection of each agent as a solid line, the secondary food collected as a dashed line, and the number of switches as a dotted line. The right column shows the food preference ratio for the primary food as gray bars, and the expected distribution if it would be generated randomly.