Signal drives genetic diversity: an agent-based approach to speciation

CONFERENCE PAPER · JANUARY 2015

DOWNLOADS
42

VIEWS
32

3 AUTHORS:

Olaf Witkowski
The University of Tokyo
10 PUBLICATIONS  1 CITATION

Geoff Nitschke
University of Cape Town
40 PUBLICATIONS  108 CITATIONS

Takashi Ikegami
The University of Tokyo
158 PUBLICATIONS  1,386 CITATIONS

Available from: Olaf Witkowski
Retrieved on: 06 July 2015
Signal drives genetic diversity: an agent-based approach to speciation

Olaf Witkowski, Geoff Nitschke and Takashi Ikegami

University of Tokyo, Japan
olaf@sacral.c.u-tokyo.ac.jp

Introduction

Species diversification is generally thought to emerge from space barriers, which isolate individuals from each other long enough so that they diverge significantly from each other, each adapting to their own ecological niche. However, the search for ecological opportunities alone does not account for all cases of speciation (Schluter 2000, Rundell & Price 2009).

Signaling is observed between individuals of species having an advantage to communicate their identity via species-specific signals, leading to reproductive isolation (Nevo et al. 1987, Carlson & Arnegard 2011).

Species may in turn diversify the signal modality, promoting speciation (Feulner et al. 2009, Carlson et al. 2011).

The ability of individuals to enter an open communication niche can trigger species diversification (Arnegard et al. 2010), which can later lead to further ecological and morphological divergence.

Many hypotheses have been proposed for the cause of signaling diversification. The first candidate is adaptation to an ecological niche, but the signal must be shown to emerge in specific environmental conditions, which is difficult to set up experimentally. Genetic drift, leading the population to randomly diverge in signal structure, is difficult to test empirically (Coyne 1992, Kawecki et al. 2012). Finally, sexual selection can also be responsible for the signal structure to branch out, as a result of divergence in mate choice between individuals of different groups (Hopkins & Bass 1981, Hanika & Kramer 2005, Boul et al. 2007).

Although signaling relation to speciation has been shown in nature, its mechanisms of emergence remain unclear. The lack of empirical data and the difficulty of experimental setups give an incentive for a computer simulated experiment. In our computational model, we use an agent-based approach to highlight the dynamics of signal versus species diversification.

The results show a similarity of dynamics, with certain differences, with the model presented Witkowski & Ikegami (2014), which this paper extends in order to study speciation. The model indeed leads to a swarming behavior, which consequently leads to specific phylogenetic structures.

Model

The utilized model extends Witkowski & Ikegami (2014), which proposed an asynchronous simulation evolving a swarming behavior based on signaling between individuals. We here adapt this model by adding a sexual reproduction scheme between agents biased positively towards agents sharing similar genetic material, that is to say, individuals with similar genes have more chance to successfully reproduce together.

A population of agents is simulated in a three-dimensional space of $600 \times 600 \times 600$, gaining a vital amount of energy from a resource gathering task. A set of 20 food spots are placed randomly in the environment, and each moved around at every 1000 iterations. By getting close to one of those food spots, the agents gain more energy, allowing them to compensate for the energy losses due to their movement and their signaling. If an agent’s energy drops to zero, it is removed from the simulation. Also, only agents with an energy higher than 4.0 are allowed to reproduce. In this regard, the energy also represents the agent’s fitness, and both terms are here used interchangeably.

The agent’s position is determined by three floating point coordinates between 0.0 and 600.0. Each agent is positioned randomly at the start of the simulation, and then moves at a fixed speed of 1 unit per iteration. Every iteration, the agent’s new velocity $\vec{c}_t$ is obtained by rotating its velocity vector at the previous time step $\vec{c}_{t-1}$ by two Euler angles: $\psi$ for the agent’s pitch (i.e. elevation) and $\theta$ for the agent’s yaw (i.e. heading). The rotation is determined by the two motor output values of the neural controller $o_1$ and $o_2$, determining respectively the acceleration in $y$ and $z$ in the agent’s inertial frame of reference, while the norm of the velocity is kept constant. The agent’s position $\vec{x}_t$ is then updated according to its current velocity with $\vec{x}_t = \vec{x}_{t-1} + \vec{c}_t$.

All individuals are blind, in the sense that they don’t see either the food patches or the other agents around them. Their interaction is limited to their exchange of signals between each others. Every agent is capable of sending signals of varying values and varying intensities (both values and intensities are encoded as floating point values ranging from 0.0 to 1.0). Each agent is also a directional communication sensor allowing it to detect signals produced by other agents in a 60-degree frontal cone. The distance to the source propor-
tionally affects the intensity of a received signal, and signals from agents above a 100 distance are ignored.

Each agent detects both the average value and the total intensity of signals produced in the cone of reception. Those are distinct values, since the intensity of a signal decreases with the distance from the emitter, whereas the value of the signal remains the same. First, the intensity sensor receives a float for the intensity, equal to the sum of every signal emitted within range, divided by the distance, and normalized between 0 and 1. Second, the value sensor receives a float for the average value of the signals from the agents in range.

The agent’s neural controller is implemented by a modified Elman artificial neural network with 2 input neurons, 10 hidden neurons and 3 output neurons controlling the two motors and the communication signal emitted by the agent. The hidden layer is given a form of memory feedback from a 10-neuron context layer, containing the values of the hidden layer from the previous time step. All nodes in the neural network take input values between 0 and 1. All output values are also floating values between 0 and 1, the motor outputs are then converted to angles between $-\pi$ to $\pi$. The activation state of internal neurons is updated according to a sigmoid function. The weights of each connection in the neural network, comprised between 0 and 1, are stored in an array. That array, constituting the agent’s genotype, is then evolved using a specific genetic algorithm described below.

The agents reproduce by mating with individuals that are close to them, resulting in the creation of offspring with a success probability proportional to the similarity of their genotypes. The genotype of an agent defines its size and the range of signals it is able to produce. The values of produced signals are modulated by each agent’s decision, modeled by an artificial neural network with the weights encoded in each agent’s genotype.

Every new agent is born with an energy equal to 2.0. In the course of the simulation, each agent can gain or lose a variable amount of energy. At iteration $t$, the fitness function $f_i$ for agent $i$ is defined by $f_i(t) = \frac{r(t)}{d_i(t)}$, where $r$ is the reward value and $d_i$ is the agent’s distance to any food spot. The reward value is controlled by the simulation such that the population remains between 100 and 1000 agents, and as close as possible to 500 agents. All the way through the simulation, the agents also spend a fixed amount of energy for movement ($0.01$ per iteration) and a variable amount of energy for signaling costs ($0.001 \times \text{signal intensity}$ per iteration).

The weights of every connection in the neural network (apart from the links from hidden to context nodes, which have fixed weights) are encoded in genotypes and evolved through successive generations of agents. Each weight is represented by a unique floating point value in the genotype vector, such that the size of the vector corresponds to the total number of connections in a neural network. The simulation uses a genetic algorithm with overlapping generations to evolve the weights of the neural networks. Whenever an agent accumulates 4.0 in energy, it becomes able to mate. When a potential mate is within 100 units of distance, they have a probability to mate proportional to the similarity of their genotypes, calculated by euclidian distance. In case of success, an infant individual is created, using a 2-point crossover, with a 5% mutation in the genotype, and added in the neighborhood of the parents, at a distance lower than 100 units to at least one of them. The parents’ energy is decreased by 2.0 and the new replica’s energy is set to 2.0.

The parameter values used in the simulations are summarized in Table 1.

![Figure 1: Architecture of the agent’s controller, a recursive neural network composed of 2 input neurons ($I_1$ to $I_2$), 10 hidden neurons ($H_1$ to $H_{10}$), 10 context neurons ($C_1$ to $C_{10}$) and 3 output neurons ($O_1$ to $O_3$). The input neurons receive the average signal value and the total intensity from the other agents. The output neurons $O_1$ and $O_2$ control the agent’s motion, and $O_3$ controls the signal it emits.](image)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial/average number of agents</td>
<td>5000</td>
</tr>
<tr>
<td>Maximum number of agents</td>
<td>1000</td>
</tr>
<tr>
<td>Minimum number of agents</td>
<td>100</td>
</tr>
<tr>
<td>Agent maximum age</td>
<td>50000 iterations</td>
</tr>
<tr>
<td>Maximum agent energy</td>
<td>100</td>
</tr>
<tr>
<td>Maximum energy absorption</td>
<td>1 per iteration</td>
</tr>
<tr>
<td>Maximum neighborhood radius</td>
<td>100</td>
</tr>
<tr>
<td>Map dimensions (side of the cube)</td>
<td>600</td>
</tr>
<tr>
<td>Reproduction radius</td>
<td>100</td>
</tr>
<tr>
<td>Initial energy (newborn agent)</td>
<td>2</td>
</tr>
<tr>
<td>Energy for replication (threshold)</td>
<td>4</td>
</tr>
<tr>
<td>Cost of replication (parent)</td>
<td>2</td>
</tr>
<tr>
<td>Survival cost</td>
<td>0.01 per iteration</td>
</tr>
<tr>
<td>Signaling cost</td>
<td>0.001 per intensity signal per iteration</td>
</tr>
<tr>
<td>Range of signal intensity</td>
<td>[0, 1]</td>
</tr>
<tr>
<td>Range of neural network (NN) weights</td>
<td>[0, 1]</td>
</tr>
<tr>
<td>Ratio of genes per NN weight</td>
<td>1</td>
</tr>
<tr>
<td>Gene mutation rate</td>
<td>0.05</td>
</tr>
</tbody>
</table>

### Results

Throughout the simulation, the progressive formation of clusters of agents is observed. Each cluster is formed of individuals of similar genotypes, producing similar signals.

In Figure 3, we show the phylogenetic tree, represented radially. At the center of the graph is the root of the tree, which corresponds to time zero of the simulation, from which start
Figure 2: Visualization of the simulation with 500 initial agents and 10 food spots (after $10^5$ iterations). The agents’ colors represent the signal values they are producing, ranging from 0 (blue) to 1 (red). The goal locations are represented as green spheres on the visualization.

The 500 initial branches. As those branches progress outward, they create ramifications that represent the descendence of each agent. The time step scale is preserved, and the segment drawn below serves as a reference for $10^5$ iterations. Every fork corresponds to a newborn agent. Therefore, every “fork burst” corresponds to a period of high fitness for the concerned agents.

To study the heterogeneity, we classify the set of all the generated genotypes with a principal component analysis or PCA (Pearson, 1901). In practice, we perform an orthogonal transformation to convert the set of weights in every genotype into values of linearly uncorrelated variables called principal components, in such a way that the first principal component $PC_1$ has the highest possible variance, and the second component $PC_2$ has the highest variance possible while remaining uncorrelated with $PC_1$. In Figure 4, the PCA results on a typical long run of the simulation, over 100000 iterations, are visualized as a biplot of the two principal components. On the plot, the genotype of each individual present in the simulation is represented as one circle. The radius of each circle represents the average number of neighbors around the agent during its lifetime. Finally, the color shows the iteration in which the agent is born, ranging from light green for the earliest time steps, to bright red towards the end of the simulation.

Figure 3: Phylogenetic tree of agents created during a run. The center corresponds to the start of the simulation. Each branch represents an agent, and every fork corresponds to a reproduction process.

Figure 4: Principal component analysis (PCA) on the genotypes of all agents throughout the simulation. Each circle represents one individual, and the color corresponds to the iteration of birth, ranging from green (early simulation) to red (late simulation).

Discussion and Conclusion

The simulations were aimed at demonstrating the mechanism of signal diversification and organism speciation. For that purpose, a difference between value and intensity of signals was introduced, in order to create the possibility for creation of communities sharing different signals. However, the preliminary results of the research indicate results of speciation reminiscent of Witkowski & Ikegami (2014), in which the genetic structure is created only based on the signal-based swarming of individuals, suggesting a degenerated version of those results. However, it is noted that the clusters observed here are not moving dynamically across the map, and clearly gather together individuals of similar genotypes.

The analysis on the genotypes of the emerging clusters of agents shows similarity between the genotypes of one cluster.
towards the end of the simulation. The phylogenetic tree, as well as the clusters of the PCA showing the whole range the colors, suggest that the formation of the clusters led to a genetic drift in each cluster. The tree rapidly goes into several main branches corresponding to the different clusters, which correspond to the different genotypes clusters observed in the population.

It should be noted, as it the case for the model presented in Witkowski & Ikegami (2014), that the simulated population has strong heterogeneity due to the asynchronous reproduction schema, which can be visualized in the phylogenetic tree. Such heterogeneity may suppress swarming but the evolved signaling helps the population to form and keep swarming. The simulations do not exhibit strong selection pressures to adopt specific behavior apart from the use of the signaling. Without high homogeneity in the population, the signaling alone allows for interaction dynamics sufficient to form swarms, which proves in turn to be beneficial to get extra fitness as mentioned above.

The results suggest that by coordinating in clusters, the agents enter an evolutionary neutral space, where little selection is applied to their genotypes. The formation of swarms acts as a shield on the selection process, as a consequence allowing for the genotypes to drift. This relaxation of selection can be compared to a niche construction, in which the system is ready to adapt to further optimizations to the surrounding environment. This can be examined in further research by the addition of a secondary task.

In this model, like in Witkowski & Ikegami (2014) the population of genotypes seems to progressively reach the part of the search space that corresponds to swarming, as it helps agents achieve a higher fitness. The behavioral transition between non-swarming and swarming happens relatively abruptly, and can be caused by either the individual behavior improving enough or the population dynamical state satisfying certain conditions, or a combination of both. The latter one is highlighted by the variable amount of time necessary before swarms can reform after the positions have been randomized, thus illustrating the concept of collective memory in groups of self-propelled individuals. Indeed, although one agent’s behavior is dictated by its genotype, the swarming also depends on the collective state of the neighborhood. Couzin et al. (2002) brought to attention that even for identical individual behaviors, the previous history of a group structure can change its dynamics. In the light of that fact, reaching the neutral space relies on more than just the individual’s genetic heritage.

References


