On the Emergence of Indexical and Symbolic Interpretation in Artificial Creatures, or What is this I Hear?

Angelo Loula[1,2], Ricardo Gudwin[2] and João Queiroz[3*]

[1] Informatics Area, Department of Exact Sciences, State University of Feira de Santana (UEFS), Brazil
[2] Department of Computer Engineering and Industrial Automation, School of Electrical and Computer Engineering, State University of Campinas (UNICAMP), Brazil
[3] Institute of Arts and Design, Federal University of Juiz de Fora (UFJF), Brazil
[*]queirozj@pq.cnpq.br

Abstract

Communication processes rely on the production and interpretation of representations, thus an important issue is to understand what types of representations are involved during the emergence of interpretations. Here we present an experiment to evaluate conditions for the emergence of interpretations of different representation types. To design our experiment, we follow biological inspirations and a theoretical framework of representation processes. Our results show that different interpretations process can emerge depending on the adaptation cost of cognitive traits and on the availability of cognitive shortcuts.

Introduction

The emergence of semiotic competences (morphosintax, grammaticality, semantics, pragmatics) has been studied through various computational perspectives, including embodied robotics, animats, synthetic ethology, and others. Particularly, virtual simulations have been used extensively to model and simulate the emergence of different types of representations (for a review of works, see Nolfi and Mirolli, 2010, Christiansen and Kirby, 2003, Wagner et al. 2003). Here we propose a synthetic experimental protocol to examine the conditions underlying the emergence of two types of representations (symbols and indexes) in a community of artificial creatures able to interact through communication processes. Empirical constraints come from evidences in studies of animal communication as e.g. the minimum brain model for animal communication, proposed by Queiroz & Ribeiro (2002), which provided us biological inspirations to develop our algorithms.

Despite the many works on the emergence of communication in a community of artificial creatures, there are still important open questions that need further exploration. Particularly, based on the fact that representations can be of different types and that communication processes rely on the production and interpretation of representations, an important issue is to understand what types of representations are involved during the emergence of interpretations in a community of artificial creatures.

In the next section, we will briefly review related work on the emergence of communication and representations processes. Then we present the theoretical and empirical constraints that guided our computational model and simulation. Next, we present our ALife experiment and its results, and, finally, we outline our conclusions and point to future perspectives on the study of the emergence of different representation types.

Related work

To illustrate the open issue of understanding the semiotic process of interpretation in communication events, we bring forward two representative works that simulate the emergence of communication in a community of artificial agents.

Floreano and coleagues (2007) studied the evolutionary conditions that might allow the emergence of a reliable communication system in a community of simulated robots, relying on biological motivations on animal communication. The robots could use a visual signal, turning on or off a light ring, to communicate with other robots about the position of food source. They found that if selection acts on group level instead of individual level, or if members of a community are genetically similar, a reliable communication system could emerge. The robots simulated in this experiment were controlled by artificial neural networks, with a direct connection between the input layer and the output layer. Floreano and coleagues did not discuss how was the light signal interpreted by the robots, or what it represented to them, but, from the neural controller architecture, we can infer that any light signal received was directly mapped to a displacement speeed, so the robot blindly reacted to it without relating to what it could represent, until it finally reached the food source itself.

Cangelosi (2001) is one of the few works to actually propose the emergence of different modalities of representations in a experiment on the evolution of communication. In an experiment with artificial creatures in a grid word, Cangelosi (2001) simulated the emergence of communication systems to name edible and poisonous mushrooms. He had also relied on biological motivations to define a food forage goal for the creatures. In typifying communication systems, Cangelosi (2001) distinguished between signals, which have direct relation with world entities, and symbols, which in addition are related to other symbols, and built two experiments to study the evolution of each type. The simulated creatures were controlled by a 3-layer neural network with an input layer that receive the visual and auditory sensory data, an intermediate layer that joint together sensory data and an output layer that defined movement and the emission of a signal. In his experiments, the neural networks were both evolved and trained in various tasks, and, in the end, a shared communication system emerged, involving signals and symbols, according to Cangelosi. But he did not described how were these signals and symbols interpreted by the creatures, i.e. if a signal heard was first mapped to a mushroom as its referent, and then to an action, or if it was mapped to an action, with a referent being associated with it. Since the intermediate neural layer might develop either solution, it is not possible to infer what could have happened.

Besides Floreano et al (2007) and Cangelosi (2001), other works have studied the emergence of communication traits and the acquisition of vocabulary or language among artificial agents (see Nolfi and Mirolli, 2010, Christiansen and Kirby, 2003, Wagner et al. 2003). But we have not found works that have studied the emergence of different types of interpretations processes and differentiated the interpretation processes that emerged.

Theoretical and Empirical Constraints

Computational models and simulations are based on different tools that are heavily influenced by meta-principles (theoretical constraints) and biological motivations (empirical constraints) in the design of the environment and the morphological definitions of sensors, effectors, cognitive architecture and processes of the conceived systems and scenarios. This theoretical basis influences modeling on different degrees depending on how it constrains the model being built and what decisions it leaves to the experimenter. Depending on the theoretical framework, this allows us to test the various factors influencing semiotic onto-phylogenetic processes, such as the differences between innate and learned communication systems, the adaptive role of compositional languages, the adaptive advantage of symbolic processes, the hypothetic substrate of these processes, the mutual influences between different semiotic competences and low level cognitive tasks (attention, perceptual categorization, motor skill), and the hierarchical presupposition of fundamental kinds of semiotic competences operating on symbol-grounding processes.

Sign-mediated processes, such as the interpretation of representations in communicative contexts, show a remarkable variety. A basic typology (and the most fundamental one) differentiates between iconic, indexical, and symbolic processes. Icons, indexes, and symbols are differentiated on how the sign relates to what it refers to, its object (Peirce 1958; see Ribeiro et al. 2007). They match, respectively, relations of similarity, contiguity, and law between sign and object. Icons are signs that stand for their objects by a similarity or resemblance, no matter if they show any spatio-temporal physical correlation with an existent object. In this case, a sign refers to an object in virtue of a certain quality which is shared between them. Indexes are signs which refer to their objects due to a direct physical connection between them. Since (in this case) the sign should be determined by the object (e.g. by means of a causal relationship) both must exist as actual events. This is an important feature distinguishing iconic from indexical sign-mediated processes. In the other hand, spatio-temporal co-variation is the most characteristic property of indexical processes. Symbols are signs that are related to their object through a determinative relation of law, rule or convention[1]. A symbol becomes a sign of some object merely or mainly by the fact that it is used and understood as such by the interpreter, who establishes this connection.

Communication is a process that occurs among natural systems and as such we can employ empirical evidences on building our synthetic experiment. Animals communicate in various situations, from courtship and dominance to predator warning and food calls (see Hauser, 1997). To further explore the mechanisms behind communication, a minimum brain model can be useful to understand what cognitive resources might be available and process underlining certain behaviors. Queiroz and Ribeiro (2002) described a minimum vertebrate brain for vervet monkeys predator warning vocalization behavior (Seyfarth et al 1980). It was modeled as being composed by three major representational relays or domains: the sensory, the associative and the motor. According to such minimalist design, different first-order sensory representational domains (RD1s) receive unimodal stimuli, which are then associated in a second-order multi-modal representation domain (RD2) so as to elicit symbolic responses to alarm-calls by means of a first-order motor representation domain (RD1m).

The theoretical descriptions and biological evidences described above guided the design of our computer experiment. We were interested in studying the emergence of indexical and symbolic interpretation competences, so, to start of, we needed to specify the requirements for each and also how to recognize each of them in our experiment. Indexical interpretation is a reactive interpretation of signs, such that the interpreter is directed by the sign to recognize its object as something spatio-temporally connected to it, so for our creatures to have this competence, they must be able to reactively respond to sensory stimulus with prompt motor answer. In the minimum brain model, this corresponds to an individual capable of connecting RD1s to RD1m without the need for RD2. But a symbolic interpretation undergoes the mediation of the interpreter to connect the sign to its object, in such a way that a habit (either inborn or acquired) must be present to establish this association. Thus, in symbolic interpretation, RD2 must be present once it is the only domain able to establish connections between different representation modes. Thus, our artificial creatures must be able to receive sensory data, both visual and auditory, in its respective RD1s, that can be connected directly to RD1m, defining motor actions (Type 1 architecture), or connected to RD1m indirectly, through the mediation of RD2, that associates auditory stimulus to visual stimulus acting as a associative

[1] Differently from Cangelosi's (2001) definition of symbol, based on Deacon's approach (1997), Peirce (1958) did not require symbols to be related to each other to be called symbols.

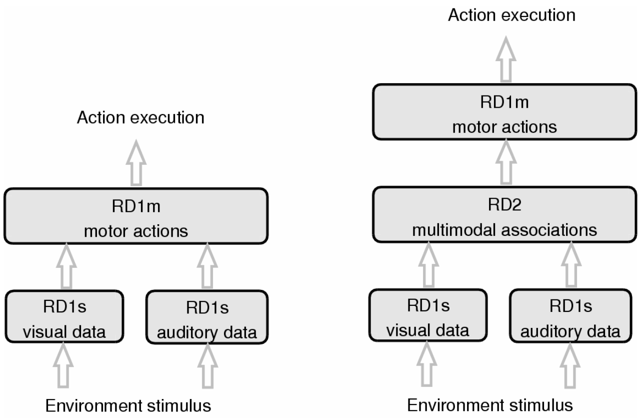

Figure 1: Possible cognitive architectures for representations interpretations. Left: Type 1 architecture, RD1s are connected directly to RD1m. Right: Type 2 architecture, data from visual RD1s and auditory RD1s can be associated in RD2 before connecting to RD1m.

memory module (Type 2 architecture) (figure 1). To evaluate what conditions might elicit each response type – indexical or symbolic –, we implemented these two possible cognitive processing paths as mutually exclusive paths: either the creature responds to auditory events indexically and reactively responds with motor actions, or the creatures responds to auditory events symbolically and associates them with a visual stimulus and responds as if that was really seen. For an external observer, which only watches the information available to the creature and its motor responses, these means changes in the interpretation process.

Building the Experiment

After specifying the brain model requirements and defining the phenomena of interest, we need to set up the scenario where we can test the conditions for both semiotic processes to emerge. To do so, we rely on the empirical evidences of animals vocalizing for food quality, recruiting other group members to feed, and so we designed an experiment where creatures are selected by artificial evolution for their foraging success. Lower quality resources are scattered throughout the environment and a single location receives highest quality resources. One creature (vocalizer) is placed fixed in this high quality resource position, vocalizing a sign continuously. At start, the other creatures (interpreters) do not know how to respond appropriately to sensory inputs and neither recognize the sign vocalized as a sign. But an evolutionary process of variation and selection is applied, with the hope to evolve individuals to better accomplish the task of food foraging. During the evolutionary process, for each start-up conditions, we observe the emergence of indexical or symbolic interpretation for the vocalizations.

The environment is a 50 by 50 grid world (figure 2) and there are 20 positions with only one resource unit each. There is also one position with 500 resource units, where an immovable vocalizer creature is also placed. The vocalizer's sole behavior is to produce a fixed vocal sign, reproduced at every instant. Fifty interpreter creatures are randomly placed in this grid and are capable to visually sense food up to a distance of 4 cells and auditory sense vocalizations up to a distance of 25 cells. This sensory range difference models an environment where vision is limited by the presence of other elements such as vegetation, restraining far vision such as in a open field. The creatures can either see a resource and its position (ahead, left, right, back) or hear a vocalization and its position, if any is within range. Interpreter creatures have a limited repertoire of action: move forward, turn left, turn right, collect resource, or do nothing; and are controlled by (genetically based) Mealy finite state machines (FSM), with up to 4 states (see figure 3). An FSM was chosen as the control architecture because it is quite simple and direct to analyze how it is functioning, permitting direct identification of the processes underlying the creatures' cognition. The creatures always respond to visual inputs with one of the motor actions, and can also respond to auditory input with a direct motor action (a reactive, indexical process) (Type 1 architecture). Alternatively, they can also choose to establish an internal association between the heard stimulus and the visual representation domain (Type 2 architecture). This internal association links what is heard with the view of a collectible resource, i.e. the creature can interpret the sign heard as a resource and act as if the resource was seen. Additionally, the creature may also ignore the sign heard, interpreting it as nothing and acting as if no sensory data was received.

At start, creatures are controlled by randomly constructed FSMs, and are allowed to live for 100 iterations for a trial, trying to collect resources. Artificial evolution selects individuals for their foraging success (number of resources collected in all trials). The 10 best individuals, i.e. the 10 individuals that collected the most resources, are allowed to

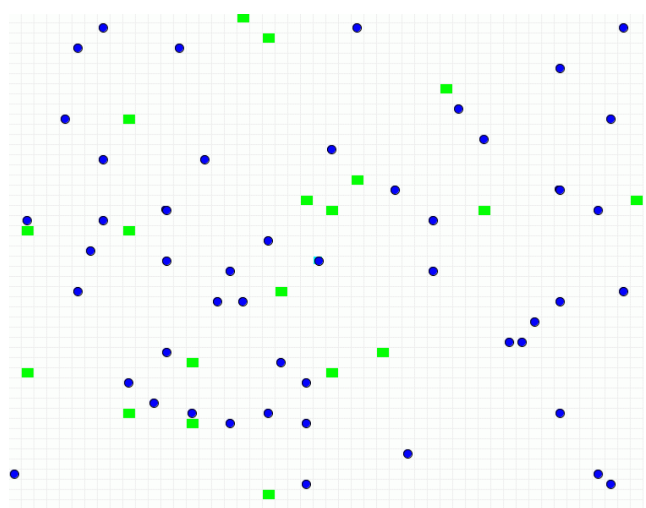

Figure 2: The grid environment. Creatures are blue circles, low quality resource positions are in green cells, and high quality one in the cyan cell in the center.

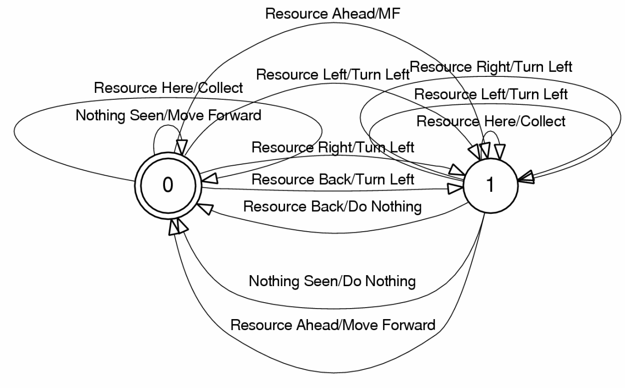

Figure 3: An example of a FSM that controls the creatures. The circles are states and a double circle marks the start state. Arches represent transitions and are labeled according to the sensory event and the action to take over when that event occurs.

breed and make up the next generation. These 10 individuals are copied to the next population and the 40 remaining individuals are a product of mutations (including a cognitive architecture type mutation) and crossovers of the FSMs of the best individuals.

The mutations can be of changing an action for a sensory event in a state, changing the next state after a transition, changing the start state and add or remove a state. The number of mutations is selected from a Poison probability distribution with an expected value of 2. The crossover exchanges states and transitions originating from the selected states between two FSM in a uniform way. All FSM undergo a correction process to fix error that might occur during these operations, such as a transition pointing to a non-existing state.

Every generation undergoes the 10 trials for 500 generations, but, in the first 200 generations (cycle 1), the vocalizer creature is not present and interpreters do not have an auditory sensor, but in the 300 subsequent generations the vocalizer creature is present and interpreters are able to hear (cycle 2). At the start of cycle 2, all creatures are set to ignore the vocalizations, as if it was not relevant, but a small mutation probability is set for changing the kind of response to vocalizations which can be of reacting to them by moving to or to link it with the view of a resource. This corresponds to a change to a Type 1 cognitive architecture (indexical) or to a Type 2 cognitive architecture (symbolic). Besides the probability of going from Type 1 architecture to Type 2 architecture is lower than the other way around, to simulate the fact that such a significant cognitive change is not that easy to happen.

We are interested in observing the overall adaptation process to the foraging task, and are specially focused on the interpretation process, related to the cognitive architecture type, that might result.

Results

To evaluate conditions that might conduct to either an indexical interpretation or to a symbolic interpretation of vocalizations (or even no interpretation at all), we first ran the experiment as described above and observed the evolutionary process and its final result, to see what kind of vocalization response and what type of cognitive architecture would prevail and consequently what type of interpretation process would be chosen. In figure 4, we present the fitness of the best individual, the mean fitness of the 10 best individuals and the mean fitness for the population. In just a few generations, best individuals where able to collect more than 200 resource items and then their foraging success oscillates around 300 items until the end of cycle 1.

Checking the FSM controlling the creatures, by generation 50, they can almost correctly respond to the view of a resource: if it is ahead, move forward, if in the left side, turn left, if at resource, collect resource, but still with bad responses when resource is at right side or at back. And when nothing is seen, they move forward. The oscillations in amount of items collected are due to the random start position of individuals.

At the end of cycle 1, at generation 200, the best individual responds properly to the view of resource, but maybe not optimally. This individual responds to the view of resource in the right with a turn to left, but since it also responds to the view of resource with a turn to the left, the final behavior allows the creature to go in the direction of the resource. If a resource shows up at right it turns left, and then the resource is at its back, so it turns left again, and the resource ends up at the left side now and it turns left once more and then moves forward to collect the resource.

After generation 200, cycle 2 starts, and a vocalizer is placed in the high quality resource position, emitting continually a vocal call. At first all creatures are set to ignore anything heard, so they interpret this as nothing at all. We can observe from figure 3 that the population evaluation rapidly increases and, in generation 210, the best individual reached an amount of resources collected around 800. The individuals adapted fast to the presence of new information in the environment, that enabled them to more easily locate the high quality resource position. The evaluation of the best individual also oscillates much less compared to cycle 1. This is because the start position does not affect as much the individual ability to find the high quality resource position, once the hearing sensor has a much greater range then the visual sensor. But we are interested particularly in the type of response the individual has to vocalizations, whether it was an indexical interpretation, a symbolic interpretation or interpreted as nothing. Figure 5 exhibits the type of response the individuals had along the generations.

In cycle 1, the vocalizer is not present and individuals are not able to hear. But in cycle 2, their hearing sensor is functional and hearing stimulus are received, but all individuals start with a default behavior of ignoring data coming from the hearing sensor and act as if no sensory data is available. In a short period, alternative responses to hearing a vocalization appear in the population, and by generation 205, the population is equally split with all three kinds of response: *indexical* response, *symbolical* response and *ignore*

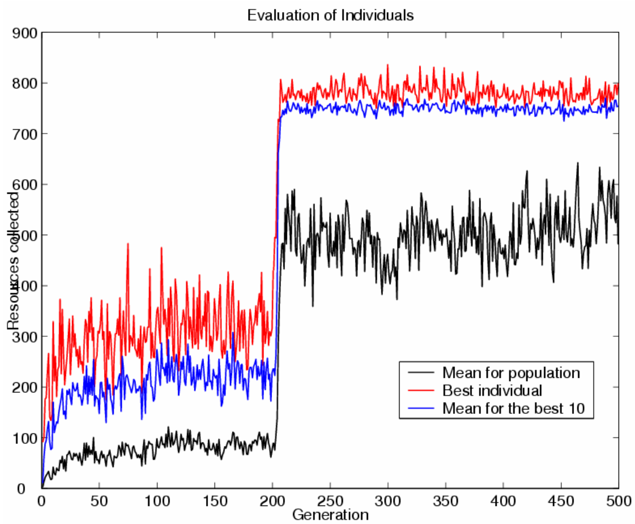

Figure 4: Evaluation of individuals along the generations for the first experiment.

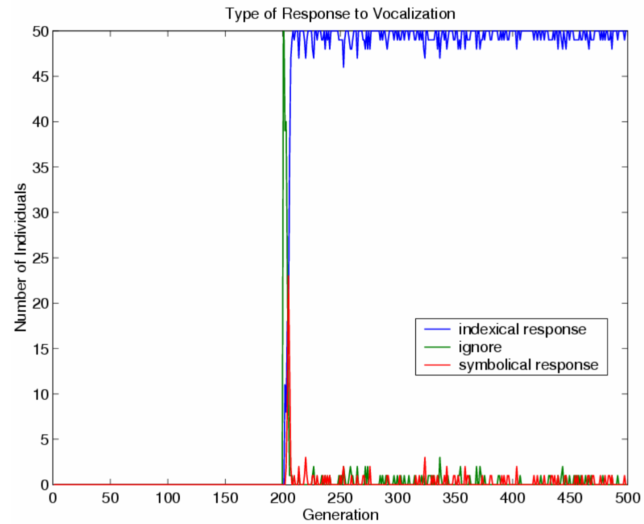

Figure 5: Response type of individuals along the generations for the first experiment.

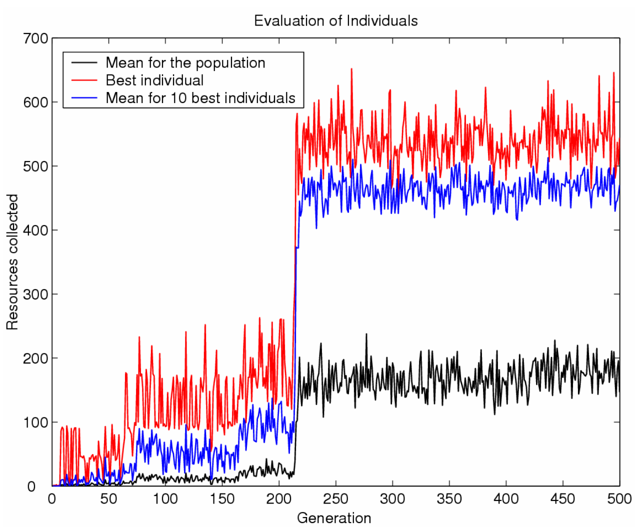

Figure 6: Evaluation of individuals along the generations for the second experiment.

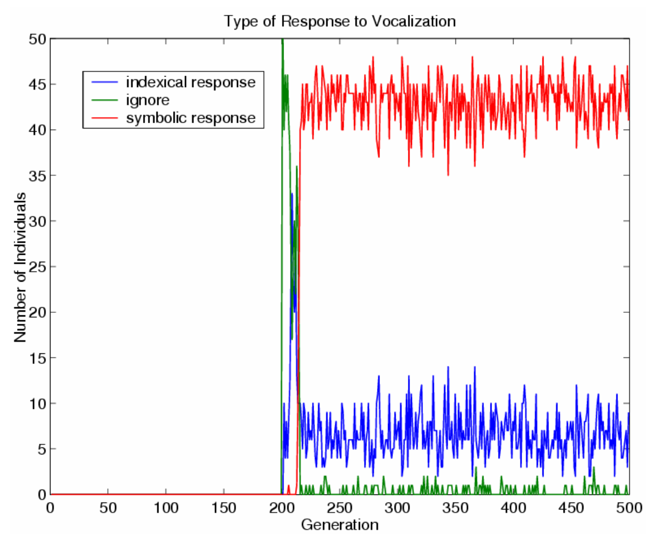

Figure 7: Response type of individuals along the generations for the second experiment.

response. This means, first, that the *ignore* response has severely declined, and, second, that the other two are rising but tied. In a closer look at generation 205, we can see that the best individual is one that responds indexically and collected 728 resource units, and the best individual with symbolic response collected 691 items. However, the mutation operator that changes a Type 1 cognitive architecture (indexical) to a Type 2 cognitive architecture (symbolic) has a quite low probability of happening, and once learning to coordinate sensory data with correct moves is an easy process in this context, as we can see from the fast adaptation in cycle 1, and, moreover, moving from Type 2 architecture to Type 1 is more probable than the other way around, adaptations involving *indexical* response stabilize faster and take over all individuals, exactly what happened after generation 210.

To further test our computational model, we started a new set up for our experiment, where actions coordination in RD1m would be harder to acquire. For that, we impose a restriction that before any movement (moving forward and turning), the creature had to 'prepare' itself by having a null action (do nothing response). To appropriately coordinate its actions then, the creature must use its internal states (finite states machines are capable of dealing with internal states), to 'remember' whether a preparatory action was taken to then take. This makes the task of coordinating sense data and appropriate actions harder.

After simulating these conditions, it can be noticed that it took longer, in cycle 1, for the creatures to evolve an adequate behavior to collect food. By generation 50, for example, the best individual was still not able to move itself around when no resource was seen, it was only able to collect a resource when it was placed in front, to the left, or exactly at a resource position. Only after generation 160, the creatures started to move forward when no resource was seen, instead of staying still when nothing is seen. Comparing to the previous experiment, this new challenge considerably required more effort for adaptation. The amount of resources collected by the creatures is also lower then in the previous experiments, due to fact that they spend a lot of the iterations 'preparing' movements (figure 6).

After cycle 2 starts in this second experiment, we can notice that the amount of resources collected by the creatures grows almost as fast as in the same transition in the first experiment. By generation 217, around 550 resources were collected by the best individual. But the vocalization interpretation evolution was not as smooth as in the first experiment (figure 7).

In the start of cycle 2, only indexical responses appear as an alternative to ignore heard signs, and by generation 212 the population is split between ignoring the vocalization and indexically responding to it with a direct action. But even though the vocalization helps finding the high quality resource, an indexical response to it is quite faulty, providing bad actions as responses. By generation 213, the first creatures start responding symbolically to the vocalization, interpreting it as if a resource was seen, and reusing the already acquired behavior in cycle 1. The symbolic response take over the population after 20 generations and is adopted by the majority of the population. Nevertheless, we can see that this response preference is not as stable as the indexical response in experiment 1, because it is more probable to go from a symbolic response to a indexical response then the other way around. But all 10 best individuals in each generation, after this convergence, are interpreting the vocalization symbolically.

Discussion

These two experiments allow us to see conditions that might guide the emergence of indexical or symbolic interpretation. In the first experiment, the acquisition of *indexical* competence, for associating arbitrary signs directly to expected motor responses is a cheap process and prevails in the population, even though the creatures already acquired the ability to coordinate visual sensory data with actions during cycle 1, and reusing this ability for auditory data would seem faster. This is due to the relative ease of learning a new ability, in face of the low probability to acquire the ability of *symbolic* response.

In the second experiment, the cost of coordinating sensory data and actions is higher, and the adaptation of symbolic responding to vocalizations does act as a viable *cognitive shortcut*, that will use the already costly acquired trait of coordinating RD1s visual and RD1m, so there is no need to learn a new coordination again. We propose that a symbolic interpretation process can happen if a cognitive trait is hard to be acquired and the symbolic interpretation of a sign will connect it with another sign for which the creature already has an appropriate response.

One further test we ran (to be described in a future work) was of removing cycle 1 from the second experiment and let the simulation start at cycle 2, with the vocalizer placed in the high quality resource and all creatures able to hear, but starting with random FSMs. It would be expected that since there was no acquired trait a symbolic response would no prevail, but surprisingly the creatures spend quite a few generations ignoring any sign heard. Only after they are able to almost adequately coordinate visual data with actions, they start interpreting the vocalizations, and they do it symbolically.

Conclusion

The emergence of interpretation processes in computational models is an open issue in Artificial Life experiments. Even though there has been already many experiments on the emergence of different traits of communication systems, the research area still lacks studies on the modalities of processes underling the interpretation of the signs been communicated, and on the conditions that might conduct to the emergence of different modalities of interpretation.

Here we proposed a synthetic experiment to examine the conditions for the emergence of symbolic and indexical interpretation processes. Simulated creatures could interpret available vocalizations in three ways: not interpreting it, interpreting it indexically or interpret it symbolically. From the results obtained, we can conclude that indexical interpretation can emerge when the acquisition of a direct coupling of sensory and motor domains is a cheap process, and symbolic interpretation of signs can emerge as a cognitive shortcut across different sensory modalities, when coordinating representations and actions directly is a costly trait to acquire.

These are initial experiments on the study of conditions for the emergence of different modalities of interpretation processes. Other possible set ups for our experiment will make certain connections faulty (like the connection between RD1s visual and RD1m) and test the robustness of this competence and of it being used as a cognitive shortcut. Furthermore, another experiment will also be built in a scenario where all creatures can hear each other and also vocalize, with no immovable creature, and test not only sign interpretation processes but also sign production processes.

Acknowledgments. R.G. acknowledges FAPESP for funding support. J.Q. thanks the Brazilian National Research Council (CNPq) and the State of Minas Gerais Foundation for Research Support (FAPEMIG).

References

Braitenberg, V. (1984). Vehicles: Experiments in synthetic psychology. Cambridge, MA: MIT Press.

Cangelosi A. (2001). Evolution of communication and language using signals, symbols and words. IEEE Transactions on Evolutionary Computation. 5(2), 93-101.

Christiansen, M.H., & Kirby, S. (2003). Language evolution: consensus and controversies. Trends in Cognitive Sciences, 7 (7), 300-307.

Deacon, T. (1998). The Symbolic Species: The Co-Evolution of Language and the Brain. New York: W. W. Norton & Company.

Floreano, D., Mitri, S., Magnenat, S. and Keller, L. (2007) Evolutionary Conditions for the Emergence of Communication in Robots. Current Biology, 17 pp. 514-519.

Hauser, M. D. (1997), The Evolution of Communication, Cambridge, MA: MIT Press

Nolfi Stefano, Mirolli Marco (2010) Evolution of Communication and Language in Embodied and Situated Agents, Berlin: Springer Verlag.

Peirce CS (1958) Collected Papers of Charles Sanders Peirce. Harvard University Press, Cam-bridge, Mass..

Queiroz, J. & Ribeiro, S. (2002). The biological substrate of icons, indexes, and symbols in animal communication: A neurosemiotic analysis of vervet monkey alarm calls. In Shapiro, M. (Ed.) The Peirce Seminar Papers 5 (pp.69–78). Berghahn Books, New York.

Ribeiro S, Loula A, Araújo I, Gudwin R, Queiroz J (2007) Symbols are not uniquely human. Biosystems 90:263-272..

Seyfarth, R., Cheney, D., & Marler, P. (1980). Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication. Science, 210, 801–803.

Wagner, K., Reggia, J. A., Uriagereka, J., & Wilkinson, G. S. (2003) Progress in the simulation of emergent communication and language. Adaptive Behavior, 11(1):37-69.