Università degli studi di Siena

Dottorato in Scienze Cognitive

Ciclo XVII

An Artificial Life Approach to the Evolution of Language: Preliminary experiments

Marco Mirolli

Tutors:

Prof. Cristiano Castelfranchi

Prof. Sandro Nannini

a

A mamma e babbo,

per tutto quello che hanno sempre fatto per me

Abstract

Human language is one of the most complex phenomena that we know. Today's research on language is subdivided in a number of different disciplines, which study the various aspects of language indipendently the one to the other and with different tools, concepts and theoretical frameworks. In this thesis I present some artificial life simulations which represent a first step into a more unified study of language and its relationships with human cognition.

Looking at human language from the point of view of artificial life not only provides new tools for addressing old questions, but also changes the kind of questions we can and want to address in the first place. First, the artificial life perspective tends to view any human phenomenon as in continuity with the animal world. This results in a tendency to consider language, first of all, as a communication system. Secondly, the artificial life perspective on human language tends to shift the focus of research from syntax to semantics and pragmatics, and from laboratory-like tasks to the evolutionary and developmental emergence of language.

The major contributions of this thesis derive from the application of artificial life tools like neural networks, genetic algorithms, agent-based computational simulations and the concepts of complex dynamical systems theory to the study of (a) the evolutionary emergence of a simple communication system and (b) how language can affect cognitive capacities, in particular categorization.

In the first series of simulations I study two related problems posed by the evolution of any communication system (language included): (1) how can communication emerge given that it requires the symultaneous presence of two independent abilities, namely speaking and hearing, and (2) which are the adaptive factors that can favour the emergence of a communication system which provides benefits only for one of the two actors of communication, in this case the hearer.

Regarding the problem of co-evolution between speakers and hearers, my simulations show that the historical emergence of communication can be promoted by what I have called 'producer biases', that is, the spontanous tendency of organisms to produce behaviors (signals) that systematically co-vary with features of the environment which are relevant for organisms' survival and reproduction. Furthermore, the detailed analysis of the simulations also suggest one possible mechanism which can lead to a producer bias: namely, the cognitive pressure towards good signal production due to the need for organisms to internally represent stimuli according to adaptively relevant categories.

With respect to the adaptive problem posed by possible conflicts of interests between speakers and hearers, I discuss a number of factors that can favour the emergence of a communication system which benefit only hearers. The first is kin selection: such an altruistic communication system can emerge if organisms interact preferentially among kin. The second factor is cultural evolution due to 'docility': if what the organisms inherit genetically are not behaviors, but the propensity to learn them from their parents, then this propensity can be the basis for learning the communication system irrespective of whether it confers an advantage to the speaker or to the hearer. Finally, the last factor is talking-to-oneself: if the communication system is not only used in social contexts, but also individually, in particular as an aid to memory, then the emergence of communication is favoured because talking-to-oneself poses a selective pressure towards good signal production.

After having discussed how the evolution of even a very simple communication system can be favoured by talking-to-oneself, in the second set of simulations I turn my attention to the study of the possible consequences of talking to oneself for human cognition. In this context, I develop an original neural network model of early language acquisition (learning to map the first words with internal representations of their 'referents'). This model provides the possibility of studying the effects of language on internal representations of objects, both when the linguistic input comes from other individuals and when it is self-produced, as in private or inner speech. The results show that both the social and the private uses of language can improve individual categorization. This finding supports the idea that language is not only a powerful communication system, but also a cognitive tool which substantially transforms human cognition. According to this view, most of the peculiar characteristics of human cognition depend on the internalization of the social cognitive aids provided by adults to the developing child through language. In the final part of the thesis I develop this point by providing some ideas on the possible influences that language can have in several cognitive domains: learning, abstraction, memory, attention, problem-solving, voluntary control, and mental life. I consider each of these domains as a possible line for future research.

Apart from the specific contributions described above, the main goal of this thesis is to demonstrate that the artificial life approach to language is both feasible and promising, and can give an important contribution to the scientific understanding of human beings.

Contents

Abstract v

Preface xi

Acknowledgments xiii

List of Figures xv

I. Introduction 1

1. General introduction 3
- 1.1. Aims of the thesis 3
- 1.2. Thesis outline 5

2. Artificial Life and language 7
- 2.1. What is artificial life? 7
- 2.2. Artificial life and complex systems 9
- 2.3. Artificial life vs. artificial intelligence 11
- 2.4. An ethological perspective to language 14
- 2.5. The synthetic approach 17
- 2.6. Diminishing the role of disciplinary boundaries 21
- 2.7. Changing the focus of the research on language 23
- 2.8. Setting the stage: the adaptive function(s) of language 25

II. Simulations 31

3. Producer biases in the evolution of communication 33
- 3.1. Two problems in the evolution of communication 33
- 3.2. Relevant literature 35
  - 3.2.1. The biological view 35

3.2.2. Computational modeling 37
3.3. The model 38
3.4. Cognitive and Genetic Factors in the Evolution of Communication . . . 41
3.5. Analyzing evolutionary dynamics 43
3.5.1. Understanding evolutionary dynamics by zooming in evolution . 43
3.5.2. The population as a dynamical system 46
3.5.3. Analyses 51
3.5.4. Results 52
3.6. The no-cognitive-pressure and no-communication simulations 57
3.6.1. Simulations 57
3.6.2. Results 58
3.7. Discussion 61
3.7.1. Artificial life simulations 62
3.7.2. Evolution of communication 64

4. Adaptive factors for the emergence of communication 67
4.1. Altruism in the evolution of communication and language 67
4.2. Relevant literature 68
4.2.1. The biological view 68
4.2.2. Computational modeling 70
4.3. Simulations 71
4.3.1. Kin selection 71
4.3.2. Docility 73
4.3.3. Talking to oneself 76
4.4. Discussion 77

5. Language as an aid to categorization 81
5.1. Language and cognition 81
5.1.1. The classical view 81
5.1.2. Challanges to the classical view 82
5.1.3. Language as a cognitive tool 85
5.2. State of the art 87
5.2.1. Empirical evidence 87
5.2.2. Computational modeling 92
5.3. A neural network model of early language acquisition 94
5.3.1. The model 94
5.3.2. Results 97
5.4. How language affects categorization 98
5.4.1. Private and inner speech 98
5.4.2. Analyzing internal representations 99

5.5. Discussion 102
  5.5.1. Language and categorization 102
  5.5.2. Implications for language evolution 104

III. Beyond 107

6. The influences of language on human cognition 109
6.1. Introduction 109
6.2. Learning 110
6.3. Abstraction 112
6.4. Memory 114
6.5. Attention 115
6.6. Problem solving 116
6.7. Voluntary control 117
6.8. Mental life 118

7. Conclusion 121

Appendices 125

A. Neural Networks 127
A.1. Natural and artificial neural networks 127
A.2. Learning algorithms 130
A.3. Different kinds of networks 132
A.4. Representations (Space-state semantics) 134

B. Genetic algorithms 137
B.1. Natural and artificial selection 137
B.2. Algorithms 138
B.3. The geometrical interpretation of evolution 142

Bibliography 145

Preface

Some of the work presented in this thesis has already been published. In particular, parts of the three Chapters which constitute Part II have been re-adapted from Mirolli and Parisi (2004, 2005a,b, 2006); some of the ideas developed in Chapter 2 have been re-adapted from Mirolli (2006); and parts of Chapter 6 have been re-adapted from Parisi and Mirolli (2006). Here is the list of the published papers:

- Mirolli Marco (2006) Vita artificiale e linguaggio, Sistemi Intelligenti, 1: 133–142
- Mirolli Marco, Parisi Domenico (2004): Language, altruism, and docility: How cultural learning can favour language evolution. In J. B. Pollack, M. Bedau, P. Husbands, T. Ikegami and R. A. Watson (Eds.) Artificial Life IX: Proceedings of the Ninth International Conference on the Simulation and Synthesis of Living Systems. Cambridge (MA), MIT Press: 182–187
- Mirolli Marco, Parisi Domenico (2005a): How can we explain the emergence of a language that benefits the hearer but not the speaker? Connection Science, 17(3-4): 307–324
- Mirolli Marco, Parisi Domenico (2005b): Language as an aid to categorization: A neural network model of early language acquisition. In A. Cangelosi, G. Bugmann and R. Borisyuk (Eds.) Modelling Language, Cognition and Action. Proceedings of the 9th Neural Computation and Psychology Workshop. Singapore, World Scientific: 97–106
- Mirolli Marco, Parisi Domenico (2006): Talking to oneself as a selective pressure for the emergence of language. In A. Cangelosi, A.D.M. Smith and K. Smith (Eds.) Proclofnameeedings of the 6th International Conference on the Evolution of Language. Singapore, World Scientific: 214–221
- Parisi Domenico, Mirolli Marco (2006): The emergence of language: How to simulate it. In C. Lyon, C. Nehaniv and A. Cangelosi (Eds.) Emergence of Communication and Language. Berlin, Springer Verlag

Acknowledgments

The years of my PhD have been very exciting, but also quite tough. During these years I have come across a great number of people who have influenced and/or helped me in various ways. I have no hope to mention all the ones who would deserve to be acknowledged, and I apologize for the omissions. In any case, I have to thank at least the following people:

- Domenico Parisi, for being a great master, a patient 'boss', an excellent collaborator, a thoughtful advisor, and a friend. Domenico represents for me – as for others before me – a paragon. And not only from the intellectual, but also from the humane side.
- Federico Cecconi, for his humor, his cheerfulness, and for being the best teacher of programming of the world. He taught me how to do simulations with neural networks and genetic algorithms through C++ programming in just three lessons which worth a thousand.
- Angelo Cangelosi, for all his useful advices and the practical help. Probably he never realized it, but my staying in Plymouth has been quite important for me, and I am really grateful for his support during that period.
- Cristiano Castelfranchi, for having helped me in various ways, and for not having got too much upset when I betrayed him for passing to the neural side.
- Fabrizio Gambassi, Gianluca Baldassarre, Luca Tummolini, Fabio Paglieri, Alberto Acerbi, Paolo Turrini, and Davide Marocco, for all the time I spent with them discussing all kinds of academic, political, and personal affairs.
- Stefano Nolfi, for the peace with which he coordinates the ECAgents project and supervises the LARAL group at the ISTC, CNR in Rome. If the LARAL group is one of the best places one can imagine for doing research, a great share of the merit is his own.

- All the people (past and present) of the LARAL group, for providing a wonderful environment to work in. Every time you have a problem, of any kind, you can find someone who is willing to help you. Besides the ones I already mentioned, this is a fairly incomplete list of the people of the ’LARAL family’ which I have to thank (in alphabetical order): Francesca Acerra, Raffaele Bianco, Anna Borghi, Raffaele Calabretta, Massimiliano Caretti, Andrea Di Ferdinando, Tomassino Ferrauto, Stefano Ghirlanda, Onofrio Gigliotta, Diana Giorgini, Francesco Mannella, Gianluca Massera, Giovanna Mazzapioda, Orazio Miglino, Dimitri Ognibene, Massimiliano Schembri, Vito Trianni, and Stefano Zappacosta. Among them, special thanks are due to Stefano Ghirlanda, for his help with the micronet neural network library, and Gianluca Massera, for all the help he has been giving to me about any kind of informatic stuff.

- The other members of the ISTC in Rome which make it a friendly and lively place to work in.

- The Plymouth friends, for making my staying there rewarding, notwithstanding all the difficulties. In particular, I have to thank Eduardo, Joao, Lea, Carmen, Lari, and Ulrich.

- Tommaso and Giorgio, Marina and family, and all the other people who have guested me during these years, for their kind and generous hospitality.

- The European Science Foundation and the Future and Emerging Technologies program of the European Community, for having founded my research during these years.

- All the reviewers of the papers I have submitted, and all the people with whome I have discussed my work in conferences, workshops and meetings, for all their valuable comments, critics, and suggestions.

- All who have been close to me during the difficult moments.

- My syster Barbara and my grandmother Bruna, for all the help and affection they have provided to me in all these years.

- My parents, Maura and Massimo, to which this work is dedicated. I owe to them practically everything.

- My wonderful daughter Marta, for being there. And for tolerating all the time I am not there.

List of Figures

3.1. The environment 38

3.2. The neural network. Arrows represent connections from all the neurons of the sending group to all the neurons of the receiving group 39

3.3. Average fitness (gray line) and communication system quality (black line) of 10 replications of the simulation. 41

3.4. (a) Average fitness (gray line) and communication system quality (black line) of one replication of the simulation. (b) Zooming in the 200 generations (from generation 400 to generation 600) enclosed in the dotted lines of graph (a). The dashed lines of graph (b) divide macro-evolutionary phase, while the dot-dashed lines subdivide this evolutionary period in the 11 sub-phases described in table 3.1. See text for details 43

3.5. 6 kinds of movement of a two-dimensional dynamical system (top), and the respective diagrams of the displacement after delays (bottom). (a) A fixed point generates a white diagram, because the displacement is always null. (b) A random walk generates a gray diagram, which tends to get darker in the upper part. (c) A limit cycle generates a dark diagram with horizontal white lines, with the distance between the lines corresponding to the period. (d) An erratic movement around an attractor generates a uniform light gray diagram. (e) A point which moves around an attractor, jumps away and then returns generates an oblique and a vertical dark line (corresponding to the jump) on a light gray diagram. (f) A jump between two attractors generates two light grey triangles, corresponding to the periods in which the point is around the two attractors, divided by a darker parallelogram, which corresponds to the distance between the two attractors 47

3.6. Diagrams of the displacement of the centroid of the population after delays for the entire system (a) and the three functional sub-systems: non-communicative (b), communicative-input (c) and communicative-output weights (d). Lines divide the 200 generations under analysis in the evolutionary phases described in table 3.1. See text for details. . . . 48

3.7. Evolutionary dynamics during generations 400-600. (a) Average variances of the non-communicative, communicative-input and communicative-output genes. (b) Displacement of the non-communicative, communicative-input and communicative-output populational centroids. (c) Coordinates of the two centroids of the signals used for edible and poisonous mushrooms. (d) Quality of the communication and representation systems of the population and average quality of individuals' systems. (e) Gray-scale bitmap of the genes of the best individual of each generation and (f) of the populational centroid (genes are grouped as non-communicative (non-C), communicative-input (C-in) and communicative-output (C-out)). The lines on or under the graphs correspond to the evolutionary phases described in table 3.1. 53

3.8. Comparison between the neural network used in the base-line and no-communication simulations (a) with that used in the no-cognitive pressure simulation (b). Connection weights are grouped according to their roles in communication: non-communicative (thick arrows), communicative-output (thin arrows) and communicative-input (broken arrows). 58

3.9. Communication system quality (a) and average fitness (b) of the base-line, no-communication and no-cognitive-pressure simulations. Average results of 50 replications. 59

3.10. Average variances on the non-communicative and communicative-output weights of the base-line simulations (a) and average variances on the communicative-output weights of the base-line, no-communication and no-cognitive-pressure simulations (b). Average results of 50 replications of the simulations. 61

4.1. Comparison between the base-line and of 4 kin selection simulations with different values of p (1%, 5%, 10% and 100%): (a) communication system quality and (b) average fitness. Average results of 10 replications of the simulations 72

4.2. Communication system quality, average fitness and average value of the docility gene of the docility simulation. Average results of 10 replications of the simulation 75

4.3. Comparison between the base-line (BL) and the talking-to-oneself (TTO) simulations with respect to average fitness and quality of the communication system. Average results of 10 replications of both simulations . . 77

5.1. The modular neural network 95

5.2. (a) Number of errors in the categorization and linguistic imitation tasks before (0) and after (4500) the first stage of learning. (b) Number of errors in the naming and comprehension tasks before (4500) and after (5000) the second stage of learning. 98

5.3. Internal representations of objects (category clouds) in the conditions of no-learning (a), no-language (b), social language (c), private speech (d), and inner speech (e). See text for details. 100

5.4. Experimental conditions for testing internal representations: (a) no-learning and no-language; (b) social language, (c) private speech, and (d) and inner speech. The process of activation is divided up into a sequence of discrete time steps. Numbers indicate connections that are involved in each time step. See text for details. 100

A.1. Some examples of neurons' activation functions. Linear: $y = x$. Step: $y = \frac{segno(x)+1}{2}$. Sigmoid (or logistic): $y = \frac{1}{1+e^{-x}}$. Hyperbolic tangent (tanh): $y = \tanh(x)$. 129

B.1. Examples of single-point (a) and double-points (b) cross-over 141

B.2. An example of a fitness landscape for genotypes composed by only two genes (g1 and g2). 143

Part I.

Introduction

1. General introduction

1.1. Aims of the thesis

Human language is one of the most fascinating phenomena that we know. And it is one of the most important characteristics that distinguish human beings from other animals. In their seminal work on the major transitions in the evolutionary history of our planet, evolutionary biologists John Maynard-Smith and Eors Szathmary regard the evolution of language as the last major transition (Maynard-Smith and Szathmary, 1995; Szathmary and Maynard Smith, 1995). Indeed, the evolution of language has surely played a fundamental role in the subsequent history not only of the human species, but of the hearth as a whole. Without language we wouldn't have had the evolutionary success that we had. In fact, the possession of language is a sine-qua-non condition for the development of all major human achievements, including human religion, society, technology and science. Therefore, it would be an important scientific achievement to clearly understand and explain how language has emerged in human beings' evolutionary history.

But studying human language is very difficult, because language is an incredibly complicated phenomenon which depends on the human brain, on ontogenetic development, and on genetic and glossogenetic (cultural) evolution. All those systems are complex systems, acting at different time scales, which in turn interact with each other in complex ways. So, it is very unlikely that a single simple explanation can be found for the emergence of human language. Indeed, it is very unlikely that a single discipline or a single methodology can suffice for understanding all the amazing features of human language and the mechanisms and factors that generate them.

What we do know is that if we move sufficiently back in time we find ancestors of human beings that lacked language and that all modern humans, except for pathologies, have language. But because of the extreme complexity of human language and because language does not leave fossil traces, we only have speculative theories concerning how and when language has emerged and evolved. In this context, computer simulations can be of help.

Simulations are computer programs and to express our hypotheses and theories as computer programs forces us to formulate them in a more explicit, detailed, and complete way than if they are just expressed in words because, otherwise, it would be impossible to translate them in a computer program. Furthermore, since a simulation's results are the empirical predictions which are derived from the theory incorporated in the simulation, hypotheses and theories expressed as simulations generate detailed empirical predictions in a mechanical and, therefore, uncontroversial way. This allows us to make a profitable use of whatever empirical evidence we do have on the initial emergence and further evolution of human language.

The general aim of this thesis is to provide preliminary simulative experiments towards the simulation of the evolution of language. The general framework on which my simulations are based is that of artificial life, which includes the use of neural networks as simple models of organisms' nervous systems, genetic algorithms as models of evolution by natural selection, and the use of concepts and tools of complex systems theory. From the theoretical point of view, I will consider language primarily as a communication system and, secondarily, as an extraordinary cognitive tool which deeply transforms most of human cognitive functions. Consequently, the simulations I present will be related to (a) general problems in the evolution of communication and (b) the role that language can play in human cognition. In particular, the first series of simulations will deal with the problem of co-evolution between hearers and speakers and with the adaptive problem posed by altruistic communication, while in the second set of simulations I will show how language can affect categorization.

All the simulations presented in this work regard simple communication systems without no syntax or grammar. There are at least two reasons for this, one practical and another theoretical. The practical reason is simply that studying syntax and grammar with artificial life simulations is currently very difficult. There are indeed some interesting works which try to address the evolution of combinatorial systems in populations of artificial agents, but typically this is done at the cost of ignoring the neural, semantic and pragmatic aspects of language, which I take to be of the major importance. The theoretical reason is that taking into account syntax and grammar seems to be un-necessary for addressing the specific problems I address in this thesis. On the contrary, at least with respect to the study of the role that language plays in human cognition, the lack of syntax in my models has a positive theoretical importance. If in order to have a positive influence on cognition language does not need to be syntactic, then it is possible that language started to be used as an individual cognitive aid early in hominid evolution, in particular before the transition from a holistic, a-syntactic proto-language to the compositional, grammatical language that we know today. Hence, in sharp contrast to what is commonly assumed, the use of language as a cognitive tool might have played an important role in the evolution of language itself. One of the main goals of this thesis is to provide evidence that studying the role that language plays in human cognition represents a crucial step in understanding both the evolution of language and the specific characteristics of human cognition.

1.2. Thesis outline

The thesis is structured in three parts and comprises 7 chapters plus two appendices. **Part I** constitutes an introduction to the thesis. **Chapter 2** is devoted to explaining what is artificial life and what are the consequences of adopting an artificial life approach to the study of language. First of all (Section 2.1), I introduce the field of artificial life in general terms, as the study of all living phenomena through their simulation in the computer or their reproduction as physical artifacts. Then (Section 2.2) I try to sort out the general characteristics of artificial life simulations and explain how and why they are related to the field of complex systems. In Section 2.3 I focus on the part of artificial life which is most directly related to the cognitive sciences, that is that part which is devoted to the study of psychological and social phenomena. In this context, I discuss the relationships between artificial life and artificial intelligence. In Section 2.4 I argue for taking an ethological perspective to language. This means trying to address, in a coherent manner, four different kinds of questions: questions about the adaptive functions of language, questions about the neural and psychological mechanisms that subserve language, questions about the ontogenetical development of language in the child, and questions about the phylogenetic evolution of language in our species. In Section 2.5 I discuss the 'artificial' side of artificial life, that is the consequences (advantages and limitations) of studying language through computer simulations. In the next two sections I discuss the two major (most general) consequences of approaching language from artificial life. First, artificial life, as computer simulations in general, has the potential for decreasing the need for disciplinary boundaries in the study of language (Section 2.6). Second, artificial life tends to shift the focus of the research on language from syntax to semantics and pragmatics, and from synchronic to dyachronic issues (Section 2.7). Finally, in Section 2.8 I introduce the simulations of part two by briefly discussing what I claim to be the two fundamental roles that language have played in human evolution: as a powerful communication system, and as a cognitive tool.

Part II constitutes the core of this thesis, in that it contains my original simulative work. It is divided in three chapters, each of which addresses a distinct theoretical question. These three chapters are all structured in the same way. They are opened by an introduction and a brief review of the literature (both empirical and computational) which is relevant for the discussed problem; then the simulations and their results are presented, and finally the chapters are closed by a discussion. The first two sets of simulations regard two related general problems posed by the evolution of any communication system (including language). In **Chapter 3** I study the phylogenetic problem posed by the fact that communication requires two independent but complementary abilities, namely sending useful signals (speaking) and responding appropriately to them (hearing). How then can communication emerge given that mutations that can lead to good speaking or good hearing abilities seem to be adaptively neutral without the complementary ability being already in place? In **Chapter 4** I study the adaptive problem posed by the fact that most of animals' communication systems (and a great part of language use) do not benefit speakers and hearers in the same way. In particular, I address the following questions: how can communication systems which benefit only hearers emerge? Why should speaker send useful signals if they do not get any advantage from doing so? As we will see, one of the possible solutions to this problem is related to one peculiar characteristic of human language, namely the fact that it is not only used socially, for communicating with others, but also individually, for talking to oneself. In **Chapter 5** I begin to study the possible consequences that using language for oneself can have on human cognition. In particular, I develop an original neural network model of early language (lexical) acquisition with which I study the effects of private and inner speech on one of the most fundamental cognitive functions, namely categorization.

In **Part III** I discuss possible directions for future work. In particular, in **Chapter 6** I discuss several of the possible effects that language has on the development of high-level human cognition, while in **Chapter 7** I summarize the major contributions provided by this work and make some conclusive remarks. Finally, the thesis contains two appendices in which I give a general introduction (including some technical details) about neural networks (**Appendix A**) and genetic algorithms (**Appendix B**).

2. Artificial Life and language

> Artificial Life is the study of man-made systems that exhibit behaviors characteristic of natural living systems. It complements the traditional biological sciences concerned with the analysis of living organisms by attempting to synthesize life-like behaviors within computers and other artificial media. By extending the empirical foundation upon which biology is based beyond the carbon-chain life that has evolved on Earth, Artificial Life can contribute to theoretical biology by locating life-as-we-know-it within the larger picture of life-as-it-could-be.
> Christopher Langton: Introduction to Artificial Life, p. 1 (Langton, 1989a)

2.1. What is artificial life?

Broadly speaking, artificial life is the study of living phenomena through their simulation in the computer or their reproduction as real, physical artifacts, such as robots. The birth of artificial life as a field can be estabilished in 1989, when the first International Workshop on the Simulation and Synthesis of Living Systems was organized at at the Los Alamos National Laboratory by Christopher Langton. The goal of that workshop was to put forward an alternative methodology to the study of life. While traditional biology is focused on the carbon-based kind of life-forms that exist on our planet, artificial life's proposed goal was to concentrate on understanding the general principles of life through the synthesis of life-like processes in computers and other artificial media. Since then, the artificial life community has grown a lot and nowadays it can count on:

- an International Society (International Society for Artificial Life, or ISAL),
- an internet web-site (http://www.alife.org),

- a bi-annual International Conference (the International Conference on the Simulation and Synthesis of Living Systems, now at the 10th edition: http://www.alifex.org),
- a bi-annual European Conference (the European Conference on Artificial Life, at the 8th edition: http://www.ecal2005.org),
- an international journal: Artificial Life (published by MIT Press).[1]

Since the general goal of artificial life is to study 'living systems' through their reproduction in artificial media, the field is obviously very broad and etherogenous, in that it covers all the spectrum of living phenomena, from the molecular, to the organismic, to the collective level (some useful introductions to the whole field are Langton, 1989b, 1995; Levy, 1992; Boden, 1996; Bedau, 2003). A great deal of work in artificial life is in fact related to low-level biological phenomena like molecular self-organization (e.g. Kauffman, 1993; Breyer et al., 1998), the construction of artificial cells (e.g. Szostak et al., 2001; Rasmussen et al., 2003), the evolution of the genetic code (e.g. Takagi et al., 2000) and the origin of multicellularity (e.g. Furusawa and Kaneko, 2002). Another important part of artificial life research is devoted to the study of abstract properties and principles related to life, like self-replication (e.g. Sipper, 1998; Ono and Ikegami, 2003), the evolution of complexity (e.g. Adami et al., 2000; Lenski et al., 2003) and dynamical hierarchies (see Lenaerts et al., 2005). Finally, a good deal of work on artificial life is devoted to the study of organisms and their behaviors (e.g. Cliff et al., 1993; Pfeifer and Scheier, 1999; Nolfi and Floreano, 2002) and to collective and social phenomena like collective problem-solving (e.g. Bonabeau et al., 1999; Mondada et al., 2004), economic behavior (e.g. Tesfatsion, 2002), and the evolution of communication and language (two recent reviews are Kirby, 2002; Wagner et al., 2003; recent interesting collections are Cangelosi and Parisi, 2002; Vogt, 2005; Cangelosi, 2005).

With such a broad spectrum of topics, it is no surprise that the field is also very variegate in methods, with all kind of models developed in all kinds of media: software (computer simulations), hardware (robotic experiments) and 'wetware' (i.e. artificial cell construction). Notwithstanding this huge etherogeneity, there are also some principles, concepts and tools which can be considered as the foundamental core of artificial life research. These include the view of living phenomena as *complex, adaptive, dynamical systems*, which rely on processes of *self-organization* which in turn give rise

---

[1] The bi-annual conference on the Simulation of Adaptive Behavior, the related journal Adaptive Behavior and the journal Connection Science are other important reference points for the part of the artificial life community devoted to the study of behaviour. At the national level, the artificial life community has also been growing significantly in the last few years, as demonstrated by the institution of an annual Workshop (at the third edition), the constitution of an 'interest group' within the Italian Association for the Cognitive Sciences, and the recent publication of a special issue of the journal Sistemi Intelligenti devoted to Artificial Life (Baldassarre et al., 2006).

to *emergent phenomena.* As a result, the standard way to study these phenomena is through *agent-based simulations.* This is true not only for software 'artificial life', but also for those parts of artificial life research which deal with hardware and wetware experiments: in fact also those kinds of researches are generally supported and/or complemented with agent-based computer simulations. In the next section I will try to briefly explain the general approach of artificial life, while the rest of this chapter is devoted to discussing the consequences of adopting the artificial life approach towards the study of cognition in general and of language in particular.

2.2. Artificial life and complex systems

Despite the great importance that complex systems research has been gaining in the last decads, there is no commonly accepted definition or classification of complex systems (Gell-Mann, 1995). A tentative definition could be the following: a complex system consists of a *large* number of components, whose mutual interactions produce highly *non-linear* behaviors which result in *emergent* properties. This is no doubt a poor definition, since the concept of an 'emergent property' is not much clearer than that of 'complex system'. A review of the philosophical debate on the concept of emergence is far beyond the scope of this work. Roughly speaking, a system has an emergent property if that property cannot be predicted/deduced/characterized by even a complete knowledge of the behavior of the components of the system itself. In other words, even if you have a perfect knowledge of all the properties of all the individual elements composing a complex system and of the rules that govern the interactions beween those elements, you still cannot predict the macroscopic behavior of the system as a whole.

The reason of this unpredictability is foundamentally due to the non-linearity of the interacions between the parts of the system. In fact, non-linear systems are in general very difficult to study with analytical (mathematical, error-free) tools. Mathematicians and phisicists can solve by analytical means only very simple non-linear system: in most of the cases, the physicist which deals with non-linear phenomena must recover to numerical solutions of the model, that is to computer simulations. And the computer simulation of even very simple non-linear systems is never guaranteed to provide a correct solution, that is a correct prediction of the state of the system at some future time. The reason is two-fold. First, non-linearities can make the system infinitely dependent on initial conditions: even very small differences in the initial conditions can lead to completely different behaviors of the system, because the differences can be indefenitely augmented by the non-lineraities. This means that in order to predict correctly the state of the system at a given time you might need to know *infinitely well* the initial state of the system, which is, of course, impossible. Second, the computer is a digital machine, which can approximate continuous values with only a *finite* precision. Hence, any infinitesime error due to the approximations which are necessarily introduced by the use of the computer can be augmented indefinitely by the non-linearities.

The non-linearity of the interactions between the elements of a complex system renders even a deterministic system foundamentally unpredictable. But most of biological complex systems are (must be considered as) non-deterministic. For example, evolution by natural selection is a due to a combination of Chance and Necessity, as biologist Jacques Monod put it (Monod, 1971). The necessity of natural selection must in fact act upon the variability due to stochastic phenomena (mutations, genetic recombinations and so on). One could philosophically dispute whether the probabilistic nature of biological phenomena is 'essential' – that is it is intrinsic to the phenomena themselves – or just 'epistemic' – that is due to our incapacity to reach a perfect knowledge of the world. But nobody can doubt that our scientific study of biological phenomena requires the introduction of stocasticity. And this constitutes another factor of unpredictabilty which must be summed up to those due to the non-linearity of complex system phenomena. Notwithstanding all those sources of unpredictability, it is just the existence of emergent properties which makes complex systems such an interesting subject of scrutiny. The reason is that in a complex system a global coherence of behavior is reached despite the local non-linear interactions between the elements.

Complex systems are said to be 'dynamical' in that their state changes in time. There are three basic macro-states which a complex system can be in: equilibrium, self-organization and chaos. A system is at equilibrium when it is in a steady state. A system is in a chaotic state when the states of its variables (or elements) change continuously in a completely unpredictable manner. Finally, the most interesting case is when a complex system is 'at the edge of chaos' (Langton, 1990), that is in a state which lies between the equilibrium and chaos. In this condition, a self-organizing process takes place which often results in the generation of higher level structures, with possibly interesting (emergent) properties. A classical example is a heated fluid (Nicolis and Prigogine, 1985). When the heating is small, the fluid fluctuates a little bit, but remains substantially homogeneous. This is the equilibrium state. If the difference between the fluid's temperature and the heating is very large, the fluid enters a turbulent state, which is chaotic. But if the difference of temperature is in a intermediate range, the fluid self-organizes into spatially distributed hexagonal cells. This is the self-organization state, which generates the order at the edge of chaos.

A complex system is 'adaptive' if it is not only the state of the system that changes in time, but also the behavior of its elements and the nature of their interactions. When this happens, a gerarchical organization appears, with the elements at one gerarchical level constituting, considered as a whole, a single element of a greater system at a higher level of the gerarchy. In this way, higher-order dynamics appear, which can lead to still higher order structures. Biological entities are typical complex adaptive systems, with biological molecules composing cells, which in turn aggregate into organs, which constitute organisms, which in turn compose societies and echosystems (Holland, 1995).

The particular properties which characterize complex adaptive systems seem to call for a particular style of computational modeling: namely, agent-based simulations, which form the core of artificial life research. An agent-based simulation in fact studies a domain by simulating active entities – the agents – whose behavior is specified at a low level. The researcher let the agents autonomously interact between themselves and with their environment and studies the global properties of the whole *multi-agent system* which typically emerges from these low-level local interactions. This style of computational modeling can be, and in fact is, applied to all kind of biological phenomena, including animal and human behavior. Such an approach is in sharp contrast with the top-down approach typically adopted in classical artificial intelligence, in which the analized behavior is typically determined (programmed) by the researcher. In the next section I will compare in more detail classical artificial intelligence and artificial life as two approaches to the study of behavior.

2.3. Artificial life vs. artificial intelligence

Traditionally, science has always been trying to understand realty through its systematic observation. This is what can be called the 'analitic approach' to science. But since the advent of the computer in the late 1940's a new kind of approach to science has appeared, one which tries to understand realty by reproducing it. This is what can be called the 'synthetic approach' to science. The rationale for the synthetic approach is twofold. First, there are some aspects of reality that are very difficult to understand by the traditional analytic approach of observing them – or even very difficult to observe in the first place. Second, once you have built a system that reproduces some phenomenon of realty, you have a candidate explanation of that phenomenon in that it is possible – even though by no means certain – that the principles that you have used to build your artificial system are the same principles that underlie the real phenomenon and explain it.

The use of the synthetic approach is today very common in many scientific fields. Science is in fact realizing that almost all aspects of reality are rather difficult to understand by using (only) the traditional analytic approach. Hence, the simulation of all kinds of phenomena with the computer has becoming a very important tool assisting the scientist in his quest for the principles that explain reality. But the synthetic approach is particularly important in Cognitive Science (Bechtel et al., 1998) as it is the at the core of one of the disciplines the contributed to its birth, namely Artificial Intelligence. Artificial Intelligence is in fact the endeveour of understanting cognition by trying to construct artificial systems able to demonstrate intelligent behavior.[2]

From the one hand, artificial life can be considered as a direct descendent of artificial intelligence, since the former has heredited by the latter the (synthetic) method, the epistemological justifications, and several of the bold (and deeply controversial) metaphysical claims.[3] On the other hand, since human beings and their behaviors are phenomena that belong to the living world, artificial intelligence may be considered just as a *part* of Artificial life, namely the part that is specifically concerned in the study of (human) behavior and cognition.

Since this thesis deals with the application of artificial life concepts and methods to the study of communication and language, from now on I will ignore most of the work produced in the artificial life community, which has little if any relevance to the cognitive sciences, and I will focus only on that part of artificial life that deals with organisms and their behaviors. According to the previous characterization, then, this thesis should be about artificial intelligence, considered as the part of artificial life dealing with intelligent behavior. But to say that would be misleading. In fact, Artificial Life and Artificial Intelligence can also be viewed as two very different (and competing) approaches to the (synthetic) study of behavior.

From about the mid 1980s on there have been in fact several important attacks to the basic assumptions made by classical, Good Old Fashioned Artificial Intelligence (GOFAI).[4] The rebuttal of these assumptions has led to the adoption of different principles, the use of different methods, and the engagement with different problems. Those differences have led to a significantly different approach to the construcion of intelligent systems. This new approach can be considered as an alternative to GOFAI and has been dubbed the 'Artificial Life route to Artificial Intelligence' (Steels and Brooks, 1994; Steels, 1994).

Considered as competing approaches to the synthetic study of intelligent behavior Artificial Intelligence and Artificial Life differ in the principles they rest upon, the methods they adopt and the problems they tend to address. This is a very brief list of

---

[2] Of course, I am referring here to the *scientific* side of artificial intelligence research, which used to constitute its core during the infancy of both Artificial Intelligece and Cognitive Science. Nowadays most of artificial intelligence research has a purely *engeenering* spirit. That is, most of artificial intelligence systems are constructed for application purposes, with little, if any, scientific goal in mind. A similar distinction between scientifically vs. technologically oriented research must also be made for the artificial life field. But, in contrast to what happened to AI, most of artificial life is still giuded by scientific goals.

[3] As an example, consider the quote by Chris Langton which open this chapter about 'locating life-as-we-know-it within the larger picture of life-as-it-could-be': this view mirrors the idea that artificial intelligence could be viewed as the study of intelligence in general, without the restriction to our (human) kind of intelligence (see, for example, Newell and Simon, 1976). In the same vein, the classical philosophical debate about 'strong' vs. 'weak' artificial intelligence which comprehends the vast literature on the 'Chinese Nation' (Block, 1978) or the 'Chinese Room' (Searle, 1980) has been mirrored by a debate on strong vs. weak artificial life (see, for example, Pattee, 1989 and Sober, 1992).

[4] The term GOFAI has been conied by John Haugeland in (Haugeland, 1985).

those differences:[5]

- Principles:
    - AI considers cognition as the manipulation of symbols according to formal rules (that is, rules that consider only the *form* of symbols and not their *meaning*); ALife strongly denies this fundamental assumption;
    - AI considers cognition as something happening in the head; ALife considers behavior – and, consequently, cognitive processes – as fundamentally determined by the interactions between an organism and its environment;
    - AI generally ignores any biological characteristics of organisms (nervous systems, the body, internal organs, genes, environment, echoloy...); ALife considers them of great importance, and tries to take them into account as much as possible;
    - AI is based on the metaphor that equates the mind to the computer; ALife refuses this metaphor as profoundly misleading and substitutes the robot to the computer as the icon of the mind.[6]
- Methods:
    - AI traditionally uses symbolic, rule-based systems, such as BDI (Beliefs-Desires-Intentions) systems; ALife typically uses non-symbolic systems, such as artificial neural networks (see Appendix A) and genetic algorithms (see Appendix B);
    - AI typically constructs its systems by top-down, explicit design; ALife uses a softer, bottom-up approach to the construction of intelligent systems by exploiting learning, evolution and self-organization;
- Problems:
    - AI tends to be exclusively interested in the cognitive aspects of behavior; ALife is interested in all aspects of behavior, including non-cognitive aspects such as emotions, motivations, sleeping, and so on;[7]

---

[5] For more comprehensive accounts see Clark (1997, 2001) and Pfeifer and Scheier (1999). See also Parisi (1999, 2005a). For a different perspective, which stresses the similarities rather than the differences between artificial intelligence and artificial life, see Castelfranchi (2006).

[6] See Parisi (2006).

[7] See, for example, Balkenius (1993); Gadanho and Hallam (2001); Mirolli and Parisi (2003). For a general discussion, see Parisi (2004).

- AI is interested only in human-like, high-level cognition; ALife is interested in the behavior of all organisms, and is currently particularly devoted to the study of low-level cognition since it is considered to be more primitive and, conseqently, more fundamental.
- AI has a synchronic view of behavior, that is it is interested only in the mechanisms that explain a given intelligent behavior; ALife has also a dyachronic, genetic, view, that it is interested in understanding also the evolutionary, developmental and historical genesis of behaviors;
- AI is only interested in understanding *how* a given behavior is realized (which are the mechanisms that underlie it); ALife is also interested in *why* that behavior is present in the first place (what is its function in the more general contex of an organism's life, its adaptive value).

So far I have given a general overview of artificial life and of the basic differences between artificial life and artificial intelligence as two competitive approaches to the study of human cognition. In the next sections I will explain, with particular reference to the study of language, which are the consequences in terms of assumed principles, adopted methods and addressed problems of endorsing the artificial life approach. In fact, the artificial life approach has at least four major consequences on the study of language. In fact it implies: a strong biological stance, the adoption of the synthetic approach, a radical decrease in the importance of disciplinary boundaries, and a significant change in the focus of research, that is, in the kind of questions we try to answer. Let's consider these points in sequence.

2.4. An ethological perspective to language

Language is of course a biological phenomenon. More specifically, it is a trait of a particular species of organisms: homo sapiens. This is pretty obvious, but the various disciplines that have been studying language have often ignored the consequences of this very simple fact. As artificial life is committed to taking a strong biological stance, the ultimate goal of the artificial life approach to language is to answer *all* kinds of questions one could ask about a biological trait.

The Nobel prize Nikolas Tinbergen, one of the founder of modern ethology, classified ethological questions in four main categories (Tinbergen, 1963). According to Tinbergen, animal behavior must be understood from four perspectives:

1. functional: what are the effects of a given behavior on organisms' survival and reproduction?

2. mechanistic: what are the mechanisms (e.g. neural or psychological) which determine the behavior?
3. ontogenetic: what are the genetic and environmental factors that contribute to the behavior's development?
4. phylogenetic: how did the behavior evolve during the history of the species?

From the point of view of the study of language, taking this ethological perspecitve, like artificial life does, has in itself several important consequences.

First, trying to answer the functional question means taking an adaptationist stance on language, something on which there is no real consensus. For example, the most eminent living linguist, Noam Chomsky, has been skeptical about the view of language as an adaptation (see, for example, Chomsky, 1988), and this skepticism has led the whole filed of (generative) linguistics to almost ignore any kind of adaptive question. According to Chomsky, language must be considered as an abstract object of inquiry, without any reference to the adaptive function that it could have played in hominid evolution. I will discuss this point in more detail in Section 2.8. Here I want just to emphasize that the artificial life approach, *qua* an ethological approach to language, tends to take an adaptivist position, assuming that understanding the adaptive functions of language is a fundamental piece in the understanding of language in general. Indeed, the major contribution of this thesis is related just to the adaptive question. In fact, I will provide some original ideas and a computational model for demonstrating one fundamental but heavily underestimated function of language, namely its role in improving human cognitive abilities (see Chapter 5 and Chapter 6).

Second, even trying to answer the mechanistic question is not very common. The literature on the neural bases of language is indeed growing very rapidly, as it is true for all the neurosciences in general. But most of the disciplines studying language – like much part of linguistics and psycholinguistics and practically the whole fields of computational linguistics and historical and social linguistics – deliberately ignore what is known about the biological bases of language. On the contrary, the artificial life perspective assumes that a knowledge of the mechanisms that underlay a given behavior are of foundamental importance for understanding that behavior. Consequently, artificial life – at least the kind of artificial life I am arguing for – is committed to taking into account and include in its models the knowledge which is acquired about the neural mechanisms underlaying language. This is clearly not an easy task. Language is a very complex phenomenon produced by the very complex brain of a very complex species of organisms. In fact, the state of the art on artificial life simulations of language does not include much knowledge on the neurobiology of language, and, viceversa, it has still not contributed much to the neuroscientific study of language. Yet, the artificial life perspective on language I argue for is still committed to studying the mechanisms underlying language with biologically founded models, which translates in the committment to using artificial neural network as control systems of simulated organisms (see Appendix A). One of the reasons for this is that the models we are currently using are very simple and their resemblance with real nervous systems is quite poor. Nonetheless, the basic principles of functioning – in short, parallel distributed processing – are modeled, and interesting parallels between the artificial models and real brains can already be made,[8] which supports the general approach and encourages the quest for more and more biologically inspired models. The work reported in Chapter 3 about possible producer biases in the evolution of communication is indeed a clear demonstration of the importance of using biological inspired models like neural network as the control systems of artificial organisms. In fact, as we will see later on, the idea of a producer bias itself has been suggested by the use of neural networks: non-neural models would have not produced the same simulation results and hence would have not lead to the same interesting theoretical hypothesis.

Third, the ontogenetic question is clearly of great importance in the study of language. Indeed, there has been a quite heated debate in the cognitive sciences about the relative role of genetic and environmental factors for the development of language in the child. Linguistists from the Chomskian tradition tend to consider language as an almost completely innate characteristic of our species, and speak about the 'language organon', the 'language faculty' or the 'language instinct' (e.g. Pinker, 1994). On the other hand, the importance of environmental factors and hence of individual learning in the development of language has always been particularly stressed by connectionist research (e.g. Elman, 1990; Elman et al., 1996; Christiansen and Chater, 1999). In continuity with the connectionist tradition, much of artificial life research on language stresses the importance of cultural transmission, trying to reconduct most of language properties not to a genetic printout but to the processes of linguistic acquisition in the child and of cultural evolution in the human societies (e.g. Briscoe, 2002; Kirby, 2002; Steels, 2005). The idea is not to deny the importance of genetic predispositions for acquiring linguistic competence, but rather to acknowledge the foundamental contribution of both ontogenetic and glossogenetic[9] processes to language evolution.

Fourth, the philogenetic question is perhaps the mostly debated one in the artificial life literature. Indeed, it can be argued that much of the renewed scientific interest in theories of language evolution and origin is due to the introduction of computater simu-

---

[8] Two very interesting exempla related to language are Cangelosi and Parisi (2004) and Sugita and Tani (2005).

[9] 'Ontogeny' refers to changes which happen at the level of the individual and at the time scale of individual development. 'Phylogeny' refers to changes which happen at the level of the species and at the time scale of biological evolution. Finally, 'glossogeny' refers to changes happening at the level of a population and at the time scale of cultural (historical) evolution.

lations – mostly of the artificial life kind – into the field (Christiansen and Kirby, 2003). This is demonstrated by the huge number of recent publications devoted to studying the evolution of communication and language through artificial life simulations,[10] and by the impressive percentage of contributes containing computational models which characterizes the International Conference on the Evolution of Language.[11]

2.5. The synthetic approach

The great importance of artificial life for the study of the evolution of language derives principally by the fact that artificial life studies language evolution through computater simulations. Simulations are important new tools for scientific research, for several reasons.[12] The first reason is that simulations are computer programs, and to express hypotheses and theories as computer programs forces us to formulate our hypotheses and theories in an explicit, detailed, and complete way because, otherwise, it would be impossible to translate them in a running computer program. Expressing theories in explicit, detailed and complete ways is of great importance in science, especially for those sciences, like the human sciences, in which theories are usually expressed only verbally. In fact, theories that are expressed only verbally are often ambiguous, vague and incomplete. This is particularly true in the evolution of language field. In fact, because of the extreme complexity of the problem, and because language does not leave fossil traces, theories about the evolutionary origins of language are necessarily wildly speculative. So, in this context, using computer simulations for expressing theories and hypotheses is of great help.

Furthermore, a simulation's results can be considered as the empirical predictions which are derived from the theory incorporated in the simulation. Those empirical predictions are generated in a mechanicam and, therefore, uncontroversial way. This is also in contrast with verbally expressed theories and speculations, whose empirical prediction, if derived at all, are often controversial and always debatable. Hence, the use of computer simulations for the study of the evolution of language is of great value also because it allows us to make a profitable use of whatever empirical evidence we do have on the initial emergence and further evolution of human language.

Finally, another great advantage of using simulations in science, and particularly in

---

[10] See, for example, the collections Cangelosi and Parisi, 2002 and Briscoe, 2002, and the two recent special issues of the Journals Connection Science and Adaptive Behavior: Cangelosi, 2005; Vogt, 2005.

[11] See, for example, the proceedings of the last (6th) edition of the EvoLang conference (Cangelosi et al., 2006).

[12] Here I will just give a very brief overview of the advantages of using simulations in science, particularly for the scientific study of language and its evolution. For a much more comprehensive discussion, see Parisi (2001b). Other interesting discussions about the role of artificial life in scientific research are Taylor and Jefferson (1994); Bonabeau and Theraulaz (1994); Bedau (1999); Di Paolo et al. (2000).

the science of language, derives from the fact that simulations are also virtual laboratories. When you run a simulation and reproduce some real phenomenon your job has not (necessarily) finished. You can also use your simulation for testing in a completely controlled setting various hypotheses on the mechanisms that underlay the phenomenon in question. You can vary the parameters of the simulation and check if and how the results of the simulation change. In this way you can test which variables the your phenomenon is dependent on and in which ways. You can run experiments with different conditions, thus assessing the importance (necessity and/or sufficience) of various mechanisms for the generation of your results. You can re-run the simulations taking new measures in order to better understand the process which led to the final results. Or you can take the outcome of your simulation – like the best evolved organism of an evolutionary process – and run a number of behavioral tests for assessing how the organism solves its task or how much is it able to generalize to novel situations or how much the solution is robust to noise or to damage and so on and so forth. This is the meaninig of using a simulation as a virtual laboratory.

The virtual laboratory has two main differences with respect to the real one: one being an advantage and the other a disadvantage. The advantage is that the virtual laboratory is much more controllable with respect to the real one. A simulation is surely 'opaque' (see Di Paolo et al., 2000), in the sense that it is not immediate to understand what's going on in the simulation. Indeed, a big effort is often required for getting a full understanding of your model's functioning. But it is clear that you can have the full access to whatever you might be interested in. You (the programmer) are the master of your simulation, and so you can ask it whatever question you have in mind. This is not true for real experiments, in which you can control and manipulate only a very limited set of variables. Furthermore, in real experiments there are often many kinds of measures that it would be very important to take but you simply cannot take. Just think, as an example, at measuring the activation of specific neurons in an human brain during a given task: you simply cannot take that measure for clear ethical reasons.

The disadvantage of a simulation with respect to a real laboratory is that the simulation is a *virtual* laboratory. You can run whatever kind of virtual experiments and take every kind of measure, but you can never be sure that the results you obtain from your simulation correspond to reality. This means that simulations can never *substitute* real experiments or empirical data gethering. In other words, the ultimate jury of any scientific theory is the real, not the virtual, world. Nonetheless, the use of simulations as virtual experiments is still of foundamental importance, expecially for those field, like the evolution of language, in which empirical evidence is scarce and real experiments are too difficult or completely impossible to run. You simply cannott reload the tape of evolution and see what happened, or what would have happened in different scenarios. But you can do that in simulations, and you can use your virtual experiments to guide speculations, real-data-gathering, and the interpretation of available empirical evidence.

Finally, simulations do not serve only for testing the internal consistency of a theory and for unfolding the theory's consequences. Making simulations changes in important ways the theory construction process itself. In fact, computer simulations can also be considered as mechanized thought experiments.[13] The activity of the researcher when constructing a simulation and that of the philosopher or the scientist when constructing a though-experiment are similar. Both are just trying to develop and clarify their theories. The difference is that the former is aided by the power of the computer, while the latter can only count on his/her own intelligence.

The thinking activity of the researcher who construct simulations is considerably transformed by the use of the computer. The construction of the simulation is in fact rarely a straightforward process in which the researcher builds a theory, implements it as a computer program, and then runs it and collects the results. The actual process is much more complicated. The first implementations of a simulation are seldom, if ever, satisfying because the results obtained are never the ones you expected. This simple fact is by itself a clear demonstration that the consequences of one's ideas are seldom if ever completely transparent. But even more important are the consequences of this fact. Frustrated by the unsatisfying results provided by the simulation, the researcher is forced to think about why the implementation of his/her theory did not give the expected results. This can be due to several reasons, like bugs in the code or all sorts of technical problems. But one important and very likely possibility is that the theory incorporated in the simulation were not correct: in other words, the consequences of the ideas implemented by the researcher are not the ones that the researcher supposed to be. This situation challanges the researcher to think better, and the continuous going back and forth between the computer and the 'drawing board', though sometimes quite frustrating, is often also very rewarding. In fact, it not only improves the understanding of one's own theory in a way that could hardly be reached with an un-aided thought experiment. It also provides new insights and new ideas which are often fundamental for the improvement of the theory itself. Again, the research presented in Chapter 3 is a very good example of a situation in which it is this continuous interaction with the simulation which permitted me to generate a novel hypothesis (and its related empirical predictions) on the evolutionary origins of communication.

Of course, simulations have also several potential problems. The first, perhaps trivial, but very concrete, is that simulations are computer programs, and computer programs

---

[13] Classical examples of thought experiments are Cartesio's evil demon, Putnam's twin-hearth (Putnam, 1975) and Searle's chinese room (Searle, 1980). For a discussion of thought experiments in science, see Kuhn (1977). For discussions of artificial life as a way of mechanizing thought experiments, see Dennett (1994); Bedau (1998); Di Paolo et al. (2000).

can have bugs. Though this problem may have not much theoretical relevance, it surely has a practical one. More effort should be put in the development of standard computational tools to be reused in the community. This could not only diminish the time wasted in debugging, but it would also prevent the quite unlikely, but still possible and more dangerous cases in which the simulation does give the expected results not because the theory is correct but because of the presence of bugs in its implementation.[14]

Another potential problem of simulations is due to the fact that they are fun. Creating 'artificial life' is really fascinating and the researcher can risk to make a simulation just for its own sake. Obviously, there is nothing wrong in writing programs just for fun: the problem rises when the program is supposed to be written for providing a contribution to science. The best way for avoiding this risk is to pay attention to already existing theories and empirical data on the subject you are dealing with, and to start writing the code only after you have a clear idea about what are the scientific problems you want to address through your simulation.

Another possible risk of working with simulations is that of not being clear about the goals you have in contructing your model. Since artificial life implies the construction of artifacts that are supposed to reproduce reality, it is easy to start thinking about those artifacts as potential applications. This makes often difficult to sort out whether a given simulation is supposed to be judged from the scientific or from the technological point of view, that is, according to how much it helps us in the understanding of reality or according to how much it can help us in the construction of useful applications. Again, as far as one wants to contribute to scientific knowledge, a good practice is to clearly relate the simulative work to the scientific problems the simulation is supposed to investigate.

Finally, I want to spend a few words on a possible misunderstanding about the role of simulations in science. Sometimes there is a resistence in the scientific community not abituated to computer simulations in accepting them as relevant tools for scientific inquiry. This is often due to the fact that simulations appear very abstract with respect to the real world. But this is not a good argument, *in se*, against simulations. Simulations do in fact simplify with respect to reality but this is true for all theories in science. Scientific theories let us better understand the extreme complexity of empirical phenomena just because they abstract with respect to reality and try to capture the essential mechanisms and processes that lie behind the phenomena and explain them.

---

[14] This is one of the reasons why most of the simulations reported in this thesis have been run twice, using both code written by myself and two open-source libraries, one for the neural networks (micronet, developed by Stefano Ghirlanda and available on request), and the other for genetic algorithms (GALib, available from http://lancet.mit.edu/ga). This is also the reason why at the ISTC of Rome, together with Gianluca Massera e Federico Cecconi, we are develping a new fast, modular, easily-extendable and open-source neural network library, called Neural Net FrameWork (NNFW). The code can be downloaded from the project's web page: http://nnfw.berlios.de.

The real problem is that simulations, and theories in general, should make the correct abstractions, that is, they should include the critical entities and factors that explain the phenomena of interest, and leave the rest out. But this, for simulations as for scientific theories in general, can only be judged in each particular case.

2.6. Diminishing the role of disciplinary boundaries

Another important implication of studying language through artificial life models is that artificial life, as computer simulations in general, tends to diminish the role of disciplinary boundaries. The study of language is divided in a great number of different disciplines, each of which is devoted to the study of only one or a very few aspects of language. For example, Chomskyan linguistics studies formal aspects of language like phonetics and syntax; historical linguistics studies the historical languages and their change in time; psycholinguistics studies how language is processed and, partially, the interactions between language processing and other cognitive functions like memory or attention; neurolinguistics studies the biological (neural) basis of language and its pathologies; and so on and so forth.

The fragmentation in the scientific study of language represents a foundamental problem for at least two reasons (Parisi, 2001a). Firstly, different disciplines do not only study different aspects of langauge, but they also tend to use radically different kinds of data, methods and concepts. As a result, disciplinary boundaries are particularly strong and difficult to bridge and this situation is clearly unsatisfying. Language is a unitary phenomenon. Therefore, if the various disciplines that study its different aspects ignore each other and do not communicate and collaborate, our knowledge of the whole phenomenon cannot but be partial and fragmented.

But the major problem is that language, like many other natural phenomena, especially the biological ones, is a *complex* phenomenon, whose aspects influence each other in complex ways and hence depend substantially the one on the other. Consequently, studying an aspect of language while ignoring all the others is probably misleading. In fact, it is likely that the correct understanding of even one single aspect of language (like, for istance, syntax) is impossible (or, at least, extremely difficult) without considering its relationships with the other aspects (for instance, the biological basis of syntax, its development, its emergence in the history of humans and so on). From this point of view, studying language from an artificial life perspective can represent an important progress. Simulations allow the researcher to consider, if not all, at least several different aspects of a given phenomenon. This makes it possible to study the interactions between those different aspects, on which probably depend the general properties of the phenomenon as a whole.

Consider the four Tinbergen's questions discussed in the previous section. Those four questions are clearly related, in that they depend the one from the other. Just for giving an example related to language, the mechanisms underlying the acquisition and use of language have evidently played a major role in its phylogenetic emergence. Conversely, those same mechanisms have evolved themselves, probably in co-evolution with language. Hence, in order to understand them it is very likely that we need to understand the phylogenetic process which made them the way they are now. The same can be said about the relationships between the biological mechanisms that underlay language and the ontogenetic development of the child. Again, it is very likely that an explanation of the adaptive functions of language depends on understanding the phylogenetic co-evolution of the language capacity, the human brain, and other anatomical mechanisms that make language possible, like the human vocal tract. Furthermore, as I argue in Chapter 5, another important font of information for understanding the adaptive function of language is the role of language in the overall development of the child. And so on and so forth. In general, it is quite reasonable to bet – I would say it is almost certain – that the explanation of any aspect of a complex phenomenon like human language is to be found in some other aspect of the same phenomenon. Consequently, strict disciplinary boundaries cannot but constitute a serious problem.

Nonetheless, it is undeniable that there are good reasons for disciplines to exist. Apart from sociological factors, like the important role of academic disciplines in structuring the scientific world (journals, societies, departments etc.), the main reason for disciplines to exist is simply that dividing the scientific study of reality in disciplines and sub-disciplines has been necessary for the development of scientific knowledge itself. It is clear in fact that the development of modern science has been facilitated, or even made possible, by the distinction among different kinds of phenomena and by the disciplinary practice of studying just one particular kind of phenomena focusing on only a small class of potential causes. Furthermore, disciplines seem to be necessary just because of the limitations of the human mind: with the increase of available scientific knowledge, it has become impossible for a single researcher to know everything. And even if one would be aware of all the facts that are necessary for explaining a given phenomenon, it would be extremely difficult for s/he to take all of them into account at the same time when studying that phenomenon. The human mind is simply limited. And this is another reason why computer simulations, especially of the artificial life kind, can be of help. In fact, the computer is not as limited as the human mind. It can process a huge amount of information coming from different fields without getting puzzled, overloaded or confused. Of course, the computer cannot *substitute* the researcher, who has to develop, run and analize the simulation. But in a simulation is the computer who has to manage all the factors that the researcher has put in it, including the interactions between those factors.

But probably the most contribution that simulations can give to solving the problem posed by disciplinary boundaries is that they can play the role of the *lingua franca* in which researchers belonging to different disciplines can start talking to each other. As I noted above, cross-disciplinary studies are difficult because different disciplines tend to work within different experimental frameworks, use different instruments, and deal with different kinds of data. But computer simulations can be (and are indeed) used for all kind of phenomena. Hence, they are a unifying method which can favour collaborations between different disciplines. In fact, since individual minds keep on being limited, it is necessary that researchers coming from different fields collaborate. And (artificial life) simulations have the potential to play a major role in making such collaborations possible.

2.7. Changing the focus of the research on language

Approaching language from an artificial life perspective does not only limit the importance of disciplinary boundaries, it also changes the focus of research (Parisi, 1997). If, on the one hand, the various disciplines tend to study different aspects of language, on the other hand it is also true that certain aspects receive more attention, while other are less investigated, if not almost ingnored. For example, the major focus of most of the studies on language is syntax (how complex linguistic forms are generated through the combination of more simple forms) while semantics (the meaning of linguistic expressions) and pragmatics (the contexts in which and the functions for which linguistic forms are used) are almost ignored.

Studying language through artificial life tends to revert this asimmetry. Artificial life simulations – at least of the kind I am arguing for – do not deal with syntactical aspects of language to a great extent. They focus mostly on semantical and pragmatic aspects: the reason is that in an artificial life simulation you can study linguistic signs together with the brain that produces them, the context in which signs are produced, the adaptive advantages that communication gives and so on. Furthermore, most of traditional research on language assumes only a synchronic perspective: that is, it deals only with what language is now (either at an individual or at a social level), and what are the mechanisms that can explain it now, where 'now' means in the contemporary adult. In contrast with this, the biological perspective of artificial life attributes a great importance also to dyachronic factors, that is, to how a given phenomenon changes through time. As discussed previously, in fact, two of the four Timbergen's questions regard dyachronic aspects: phylogenesis and onthogenesis. As a consequence, artificial life is mostly interested in the study of language origins and evolution, language ontogenetic development, and the cultural mechanisms that underlay the changes in historical natural languages.

The importance attributed to dyachronic mechanisms is related to the epistemological approach known as 'genetic epistemology' after the work of developmental psychologist Jean Piaget (Piaget, 1972). The basic idea here is that diachronnic aspects do not constitute only important objects of inquiry in themselves. They are also acknowledged to be foundamental for the understanding of the phenomenon under scrutiny – in our case, language – as a whole. In other words, it is assumed that it is extremely difficult – if not impossible – to understand language, both at the social and at the individual level, without understanding how (and why) language has become what it is now, both in the philogenetic and in the ontogenetic sense.

Obviously, I do not want to sell artificial life as the real and only solution for each and every problem. In fact, it must be admitted that many of the problems that affect the more traditional study of language affect also the study of language through artificial life. In particular, the intrinsic complexity of language makes it extremely difficult, at least today, to take into account, simultaneously, all of its important aspects. In other words, even the artificial life researcher must focus on some specific aspects, ignoring or (over-)simplifying all the rest. It is true that, contrary to the traditional methods used by the sciences of language, a simulation can potentially include a great number of different factors related to different kinds of empirical evidcence (as argued by Parisi, 2001b). But it is also true that this theoretical possibility is quite difficult to realize in practice. The reason is twofold. First, as discussed above, the researcher who do the simulation is not, and can hardly be, an expert in all the disciplines related to language. Consequently, he or she comes to ignore a great number of relevant facts simply because (s)he does't know them. Second, the state of the art in computational models of the nervous system (neural networks), natural selection (genetic algorithms) and the interactions between organisms and their environment (embodied and robotic simulations) is still in its infancy, and it does not yet allow for very complex simulations. In fact, when constructing a simulation, the researcher encounters a huge and frustrating number of both scientific and technical problems, like the lack of good models of temporal processing in neural networks, or the lack of powerful and plausible learning algorithms, or the limits of current genetic algorithms. Furthermore, another extremely relevant practical problem is constituted by the great amount of computational time that is required for artificial life simulations. And the richer your model, the longer it takes your simulation. Consequently, all these problem do not allow the researcher to include in the simulations all the aspects that (s)he would like to.

Still, artificial life simulations are already giving a foundamental constribution to all the biological sciences, and in particular to the study of communication and language. Furthermore, the importance of simulations is likely to increase in the future as the available computational power increases, technical problems are solved, more complex and biologically plausible models of nervous systems, natural selection and embodied cognition are developed, and new collaborations between artificial life researchers and more traditional scholars take place.

2.8. Setting the stage: the adaptive function(s) of language

The first point to be clarified when discussing the evolution of language is whether we are to take and adaptationist stance or not, since doubts on the value of an adaptivist perspective on language have been casted by eminent linguists, like Noam Chomsky (Chomsky, 1972, 1975, 1988) and Massimo Piatteli-Palarini (Piattelli-Palmarini, 1989). Adaptation is the process by which an organism changes in order to fit its environment. An adaptationist account of a trait is one which investigates the role of natural (and cultural) selection in the process of adaptation. Basically, the idea is to sort out the pressures which, by acting on genetic and cultural transmission, favour organisms which are more adapted to their environment, such that over time a population of such organisms comes to fit its environment rather well. A *pan-adaptivist* view is one in which all aspects of an organism are considered as adaptations for specific aspects of the organism's environment.

Adaptationists aprroaches have been criticized by the eminent biologist Stephen J. Gould, who, in a series of publications, strongly argued against the risk of pan-adaptivism, that is the assumption that every biological trait is to be considered as an adaptation. In this contest, two concepts were proposed as possible anthagonist to adaptations: 'sprandels' (Gould and Lewontin, 1979) and 'exaptations' (Gould and Vrba, 1982). The term sprandel is related to the role of architectural constrains in shaping the ways in which organisms develop and evolve: ontogenetic development is in fact subject to the laws of growth and form (Thompson, 1917), which restrict in foundamental ways the search space on which natural selection can act. For this reason, particular structures that might appear to be designed for some specific purpose can in fact be the result of architectural constrains on an organism's development: these are sprandels. Exaptations are traits that, after having evolved as sprandels or as adaptations, are put to a new use by the organism that possess them. In other words, the function that an exaptation serves in a given moment is not the function that trait have served (if any) during its own evolution.

Chomsky and Piattelli-Palmarini have criticized adaptationist approaches to the study of language appelling, in a more or less explicit way, to the views of language as an sprandel or as an exaptation. These are a few significant quotations:

> We know very little about what happens when 1010 neurons are crammed into something the size of a basketball, with further conditions imposed by the specific manner in which this system developed over time. It would be a

> serious error to suppose that all properties, or the interesting properties of the structures that evolved, can be explained in terms of natural selection. (Chomsky, 1975, pag. 59)
>
> The answers may well lie not so much in the theory of natural selection as in molecular biology, in the study of what kinds of physical systems can develop under the conditions of life on earth. (Chomsky, 1988, pag. 167)
>
> ... innate, very specific, and highly abstract structures governing language and cognition may be seen as spandrels, that is, biological traits that have become central to our whole existence, but which may well have arisen for some purely architectural or structural reason ... or as a by-product of evolutionary pressures driven by other functions. (Piattelli-Palmarini, 1989, pag. 19)

Since no concrete and specific proposals have been made about which could be the physical (architecureal or structural) constrains that would render language the result of the laws of form and development, the quoted statements appear nothing more than the result of a profound suspiciousness in the power of evolutionary explanations in general. This is confirmed by the following early statement by Chomsky:

> ...it is safe to attribute this development [of language] to natural selection, so long as we realize that there is no substance to this assertion, that it amounts to nothing more than a belief that there is some naturalistic explanation for these phenomena. (Chomsky, 1972, pag. 97)

This supsiciousness is probably due to some general misconceptions about evolutionary biology, and, in particular, about the relative roles of natural selection, sprandels and exaptation in evolution (for a similar discussion, see Pinker, 2003). Gould's notions of sprandels and exaptations (together with the notion of evolution happening through 'punctuated equilibria' rather than through gradual adaptation, see Gould and Eldredge, 1993) have often been popularized, firstly by Gould himself, as substantial revolutions in evolutionary theory. This has had the effect of forming a strong misconception, outside the field of biology, about the presumed vacuousness of adaptive thinking. This view is completely misleading. Rather then having disconfirmed Darwin's theory of evolution by natural selection or having redimesioned its importance, Gould's ideas are better considered just as little emendaments of the same theory. In fact, they have been accepted within evolutionary biology as just other factors to be considered within the adaptivist framework (for a detailed and comprehensive discussion on this topic see Dennett, 1995).

Nobody in evolutionary biology (nor in the field of language evolution) is a pan-adaptivist. Everyone agrees that there are traits in organisms whose explanation is to be found in developmental or physical or architectural constrains, and others which are just due to genetic drift. Nonetheless, it must be ackowledged that adaptationist thinking has a foundamental and un-sostitutable importance in biology.[15] Indeed, natural selection remains the *only* known force which is be able to produce complex adaptive design: that is, an organism's feature which manifests non-random organization and which permits to achieve improbable goals that increase the organism's chances of survival and reproduction (beside Dennett, 1995, see also Williams, 1966 and Dawkins, 1986).

Human language – and the brain and other anatomical organs which are required for acquiring and using it – is no doubt a very complex and non-random human trait. Furthermore, it is clearly at the core of humans' adaptive style. Hence, it is not only reasonable, but perhaps compelling, to take the adaptive stance and try to understand the adaptive function (or functions) which drove language evolution. The famous paper by Pinker and Bloom on *Behavioural and Brain Sciences* (Pinker and Bloom, 1990), which is largely accepted as the catalyst of the renewed scientific interest in language evolution (see Christiansen and Kirby, 2003), was indeed just an elaboration of this point: language is complex and only natural selection can explain complex design. And in fact, this is the only general reply one can make to the quoted assertions of Chomsky and Piattelli-Palmarini. Indeed, attributing the evolution of language to natural selection *is*, by itself, nothing more than believing that there is a naturalistic explanation of language. But the contrary is also true: denying the role of natural selection in language evolution is just denying (or refusing to cosider) any naturalistic explanation for it![16] For this reason, I will assume, as everybody in the field, that language is an adaptation. But the agreement that language evolved for natural selection is just the starting point in trying to understand language evolution. The follwiong step is to ask: what is the adaptive role of language?

The adaptive question is surely of the most importance, if one takes an evolutionary perspectiveon language. So it is somewhat surprising that in the literature on language evolution there is not much debate on this topic. The reason is that it is perhaps commonly assumed that the function of language is communication. According to what can be called the 'standard' or 'received' view, language is nothing but a very complex and powerfull communication system. What else could it be? It is interesting to note that the most famous opponent to this view has been, again, Noam Chomsky, who polemically argued that language might rather be a *representational* system. This view of language as a representational system rather than a communication system has

---

[15] As evolutionary biologist Christian Dobzhansky stated, “Nothing in biology makes sense except in the light of evolution” (Dobzhansky, 1973).

[16] Recently Chomsky seems to have lowered his skepticism on evolutionary explanations of language, as testified by the recent paper he has published in the journal *Science* together with the two eminent evolutionary biologists Marc Hauser and Tecumseh Fitch (Hauser et al., 2002).

been developed within a evolutionary framework by linguist Derek Bickerton (Bickerton, 1990, 1998). The basic idea is that language is, before than anything else, a representational system, serving the function of enabling complex individual reasoning. According to this account, the use of language for communication was just an exaptation from this primitive function. This exaptation would have happened when the mental (cerebral) apparatus dealing with what Bickerton calls 'theta analysis'[17] became connected to the brain regions governing the motor control of the vocal tract.

I will not discuss the (several) weaknesses of Bickerton's account in general (for such a discussion see, for example, Hurford, 2002). What is of interest here is the plausibility of the idea of language as primarily a representational system, as opposed to a communication system. The most obvious problem of this idea is that a great part of the features that characterize language seem to be clearly due to its communicative function. For example, most of the features of language which are related to syntax seem to be due to the need of communicating complex 'conceptual' structures over a sequential and noise channel. There seems to be no reason for the existence of things like the duality of patterning, the anaphor-antecedent relationships, the rules for word-order, case marking, passivization etc. in a system for internal representation. On the contrary, all those feature of language clearly play an important role in communicating complex situations over a serial channel (a similar argument has been made also by Pinker, 2003).

In other words, language today is a fundamental tool for communication and many of its aspects are clearly functional to this role. As a consequence, it seems undeniable that the communicative function of language has playied a fundamental role in language evolution, and cannott be considered just as a late exaptation. It is for this reason that a comparative approach to language, such as the one proposed by biologists Mark Hauser and Tecumseth Fitch (Hauser, 1996; Hauser et al., 2002; Hauser and Fitch, 2003; Fitch, 2005), is necessary. In other words, as language is no doubt primarily a communication system, it is important to consider both the properties of language which are shared with other animals' communication systems and those which are not. The simulations presented in Chapter 3 and Chapter 4 are devoted to the study of two problems raised by the evolution of communication in general, including language considered as a communication system.

On the other hand, language, in moder human, has not only a communicative function, but also an individual function: it is used for talking to oneself, as a cognitive aid. Indeed, during development, the use of language for oneself is manifest and impressive: it has been reported that children from 3 to 10 years of age use language for themselves 20-60% of the time (Berk, 1994). Furthermore, talking-to-oneself do not disappear in adulthood, but it is just internalized, becoming inner speech. Such an im-

---

[17] the representation of 'who did what to who with what'.

pressive individual use of language calls for an explanation. Back in the 1930s russian psychologist Lev Vygotsky provided strong arguments for the view that language is not only a communication system but also a cognitive tool, something that transforms the psychological functions of the pre-linguistic child in the higher-level cognitive functions of the adult (Vygotsky, 1978).

I will argue for the importance of language as a cognitive tool in Chapter 5. Here I just want to note that in the literature on language evolution this role is heavily underestimated, if not completely ignored. Indeed, the main contribution of this thesis to the evolution of language dabate is to point to the possible role of talking-to-oneself in language evolution. The basic idea is simple: the best way for thinking about the adaptive role of language during its evoluiton is to look at the functions that language plays in moder humans. And just as the importance of language as a system for communication cannot be ignored, so it cannot be ignored the role that language plays nowadays both in cognitive development and as a cognitive tool in adult life. Consequently, I suggest, if we want to understand language evolution we must try to think about the possible role that talking to oneself might have played in language evolution. In other words, beside classical questions like 'when, where and how did language evolve', we must also address questions like the following. When did the use of language as a cognitive tool begin? What role did it play in hominid evolution? Did this individual adaptive function of language have any consequence on the design features of language itself? Which? And so on and so forth.

Part II.

Simulations

3. Producer biases in the evolution of communication

3.1. Two problems in the evolution of communication

The emergence of communication (including the emergencec of language) requires the co-evolution of both good speakers and good hearers. Good speakers are individuals that produce signals that systematically co-vary with situations that are relevant for survival and reproduction. Good hearers are individuals that react to signals in ways that are appropriate to the situations with which signals systematically co-vary. If hearers do not respond appropriately to the signals produced by the speakers, there is no reason for speakers to emit appropriate signals. In other words, good speakers without good hearers are useless because signals are not understood and reacted to appropriately. Conversely, if speakers do not produce the appropriate signals in the appropriate circumstances, there are no useful signals for hearers to understand. In other words, good hearers without good speakers are meaningless because one cannot react appropriately to signals that do not co-vary with any relevant feature of the world.

This simple fact renders the emergence of communication tricky for two related but independent reasons. The first reason has to do with the historical paths that can lead to a communication system, that is to the phylogenetic question of Tinbergen (see 2.4). The second reason has to do with the adaptive advantage of communication itself, that is to the functional question. The phylogenetic problem can be conceived as a kind of chicken-egg problem: how might hearers appear if spekers are not already there sending useful signals? And, conversely, how might speakers appear if hearers are not already there understanding what they can say? More concretely: how can communication evolve if the traits necessary to its emergence - namely, good speaking and good hearing capabilities - taken in isolation are adaptively neutral in that they do not, by themselves, increase the reproductive chances of the individuals that possess them?

But besides the phylogenetic problem, the complementarity between hearers and speakers implies an adaptive problem, a problem of altruism. In fact, not all communication systems provide advantages to both actors of communication, that is, speakers and hearers. Generally speaking, with respect to adaptiveness, communicative systems can be broadly classified in three categories, according to who is advantaged by communicative interactions: only the speaker, only the hearer, or both. Communicative interactions that advantage both the actors of communication are typically those in which the speaker and the hearer have common interests and have to coordinate their behaviours (like in group hunting). Communicative interactions that advantage only the speaker are instances of manipulation (like in commands or in aggressive displays made for convincing the hearer to leave some contended resource to the speaker). Finally, communicative interactions that advantage only the hearer are typically those in which the speaker informs the hearer about some feature of the the environment which is relevant for the latter (like in alarm or food calls). If communication clearly advantages both the actors of communicative interactions, its evolution poses only the above-mentioned co-evolutionary problem. But in those communication systems in which only one actor of communication is advantaged the problem of co-evolution generates an adaptive problem due to the altruistic character of these kinds of communication. Why should a speaker produce good signals if only the hearer is benefited by communication? Conversely, why should a hearer respond appropriately to signals if this provides benefits only to the speaker? How could such communication systems evolve?

I have addressed both the phylogenetic and the adaptive problem within the same simulative scenario, but since the two problems are at least logically separated I have divided their treatment in two chapters. The present chapter is devoted to the phylogenetic problem, while the next deals with the adaptive one.

The reminder of this chapter is structured as follows. In Section 3.2 I first discuss the stance that evolutionary biologists typically take on the problem of co-evolution between hearers and speakers. Then, I introduce an idea which has been raised by the recent artificial life literature: the idea that the evolutionary emergence of communication can be favoured by a 'cognitive pressure' toward spontaneous good signal production. In Section 3.3 I describe my own artificial life model. In Section 3.4 I present the results of the basic simulation and discuss my hypotheses concerning the factors that can affect the emergence of communication in the model. The next two sections are devoted to the careful investigation of the validity of the proposed hypotheses. In Section 3.5 I study the evolutionary dynamics of a single run of the simulation by considering the population as a dynamical system moving in a multi-dimensional genotypic space. In Section 3.6 the results of the standard, base-line simulation are compared with those of two new simulations, which play the role of control-experiments. Finally, in Section 3.7 I discuss the relevance of the presented work with respect to both previous artificial life work and to the biological literature on the evolution of communication.

3.2. Relevant literature

3.2.1. The biological view

The recent biological literature on communication has focused mostly on manipulative communication, due to the controversial but deeply influential work of Dawkins and Krebs (Dawkins and Krebs, 1978).[1] Those authors gave a definition of communication which *identified* it with manipulative behavior. According to Dawkins and Krebs communication occurs "when an animal, the actor, does something which appears to be the result of selection to influence the sense organs of another animal, the reactor, so that the reactor's behaviour changes to the advantage of the actor". This theoretical framework was put forward as a reaction to the general view of communication of early ethologists (see, for example, Tinbergen, 1952) according to which the function of communicative signals was to insure the survival of the group or the species. This species-centered view of natural selection was challenged by the gene-centered view according to which the basic unit of selection must be the smallest unit of reproduction, that is the gene. Therefore, any advantage that a trait might produce for the species, the group, or even the individual, must be understood in terms of the selective advantage of 'egoistic genes' (Dawkins, 1976). Hence, the definition of Dawkins and Krebs: a view of communication as manipulation seemed to be more in line with the new gene-centered view than the old, cooperative view of early ethologists.

The historical importance of Dawkins and Krebs' framework is probably due to the fact that it has ever been the only comprehensive theoretical framework on the topic, and one which has produced a number of interesting empirical predictions (Hauser, 1996). Nonetheless, the identification of communication with manipulation seems to be unjustified both theoretically and empirically. The theoretical point is that the problem of the evolution of communication is symmetrical: pure adaptive reasoning implies that just as the signaller must get an advantage for sending information, the receiver must also get an advantage in order to respond to signals. Empirically, just as there are communication systems in which the result of the communicative interaction seems to benefit only the signaller, such as the aggressive displays produced when competing for

---

[1] The very brief historical sketch on the biological thought regarding the evolution of communication made in the present and next paragraph follows the discussion of Hauser (1996), especially section 2.2.

some resources or begging calls produced by new-borns for receiving extra feeding from parents, there are also communication systems which seem to provide benefit to both the actors of communication, such as mating calls or signals emitted during cooperative endeavours like group hunting, and communication systems which seem to benefit only the receiver, such as the food or alarm calls produced by many species.

While the communication-as-manipulation framework is well-suited for dealing with the first kind of communication systems, it seems less useful when dealing with cooperative communication and completely misleading when dealing with communication systems which seem to benefit the receiver. Indeed, the definition of communication of Dawkins and Krebs was criticized by several authors but the reply of Krebs and Dawkins (Krebs and Dawkins, 1984) was to accept some of the critics as just little amendments to the basic framework, and their way of framing the problem of the evolution of communication as the evolution of manipulation has been informing most of both the theoretical and empirical biological literature (again, see Hauser, 1996).

The influence of the idea of communication as manipulation is manifest in the treatment of the problem I am dealing with here, that is the problem of co-evolution between speakers and hearers. A weel known possible solution to this problem has in fact been given through the notion of the exploitation of the sensory (or receiver) bias (Ryan, 1990, 1993; Maynard Smith and Harper, 2003). The basic idea is that in order to manipulate the receiver, the speaker could exploit some behavioral bias. If some environmental feature triggers a specific behaviour in some organism that has an adaptive value for another one, the latter can produce signals that resemble that environmental feature in order to manipulate the behavior of the former for its own interests. This idea is certainly sound and can probably explain the evolution of several manipulative communication systems. But it seems much less adequate for explaining the emergence of communication systems which seem to advantage the hearer, like, for example, the alarm calls of vervet monkeys (Struhsaker, 1967; Cheney and Seyfarth, 1990) or the food calls of honey bees (Frisch, 1967; Seeley, 1992). It seems more reasonable that this kind of communication systems could emerge from an opposite mechanism, what we can call the *exploitation of a producer bias*. If organisms have some bias to produce behaviours that systematically co-vary with features of the environment which have an adaptive value, those behaviors can be used by other organisms as cues for their own purposes. In this way, the biased behaviors become signals, the organisms that exploit them become hearers and the individuals that produce them become speakers. Is the idea of a producer bias reasonable? Where could such producer biases come from? Some recent work in the artificial life field has provided a possible candidate: the signalling behavior of organisms could be biased in a positive way by the need for internally categorize experience in adaptive ways.

3.2.2. Computational modeling

Very little has been done in the artificial life field for studying the problem of co-evolution of signalling and receiving capacities (for a recent review of this literature, see Wagner et al., 2003). This is odd, since artificial life simulations could give an important contribution to the study of this kind of problems. The mathematical and game-theoretic approaches which are standardly adopted in theoretical biology are in fact not well suited for studying topics like behavioural biases, because they typically abstract away from details of behaviour which are not directly relevant to communication. Conversely, artificial life simulations are particularly well suited for studying this kind of problems since in an artificial life simulation communication is typically studied in a broader contest, which can include any kind of non-communicative behavior. Furthermore, artificial life simulations often include both genetic algorithms as models of evolution and neural networks as models of the nervous system of organisms. This makes it possible to study within the same simulation not only the adaptive function of communicative behaviours, but also its philogenesys and the neural mechanisms that underly it (that is three of the four Tinbergen's questions).[2]

In an artificial life model similar to the one presented here, Cangelosi and Parisi (1998) were able to evolve a communication system which benefits only hearers in a population of ecological neural networks.[3] Discussing their surprising result, those authors suggested that even though in their model there was no advantage for speakers to produce useful signals, communication could evolve thanks to the link between communicative behavior and other cognitive abilities. The argument runs as follows. In order to behave adaptively, organisms need to internally categorize experiences in ways that are relevant for their survival. As a consequence, if communicative behaviors depend on internal representations of experience, which is true in Cangelosi and Parisi's simulations and is assumed to be true also in real organisms, then produced signals will tend to be correlated to adaptively useful categories. As a result, the evolution of good signallers might be explained as a by-product of their independently evolving ability to categorize the environment.

The same idea has been re-stated by Marocco et al. (2003), who tried to evolve a communication system in a population of neural networks controlling a robot arm

---

[2] Some artificial life simulation have included also individual learning (see, for example, Kvasnicka and Pospichal, 1999 and Smith, 2002), in a way that would permit to study also the fourth Tinbergen's question, that is the ontogenesys of a trait. However, up to now this has been done by excluding the study of adaptation. In fact, in most of these studies the fitness of the organisms corresponds directly to communicative success (but see Munroe and Cangelosi, 2002 for an exception; the 'docility' simulations presented in the next chapter are the only other exception, to my knowledge, to this rule).

[3] With 'ecological neural networks' I mean simulations in which the neural networks are considered as the control system of artificial organisms living in an environment and interacting with it and between themselves (see Parisi et al., 1990; Parisi, 1999).

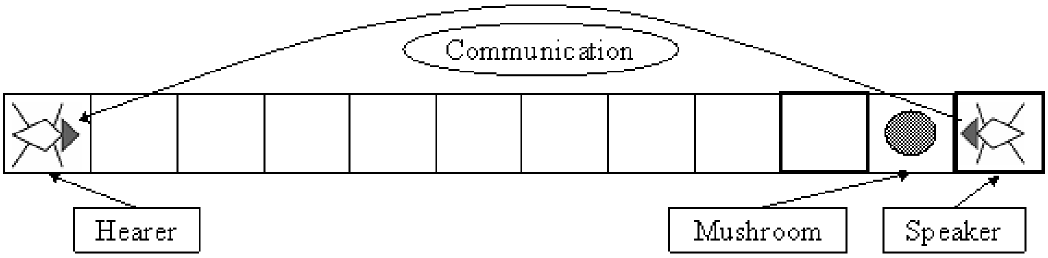


Figure 3.1: The environment

whose task was to categorize objects by physically interacting with them. But in that work the supposed cognitive pressure towards good communication was not sufficient for the evolution of communication. In fact, Marocco et al. had to introduce in their simulation also a very strong kin-selective pressure (see nex chapter) and had to make networks pre-adapt to solve the task before introducing inter-agent communication. Furthermore, notwithstanding those facilitations, a good communication system evolved only in 7 out of 10 replications of the simulation.

Indeed, I think that the idea of the 'cognitive pressure' on good signal production is sound, but it is not sufficient for explaining the results of Cangelosi and Parisi (1998) for two reasons: first, the cognitive pressure can produce only an *individual* tendency to produce good signals, but cannot explain why signals are *shared* in the population; second, the *emergence* of a good communication system does not explain its *evolutionary stability*, that is why the communicating population is not invaded by mutant individuals that cheat conspecifics by producing misleading signals. The simulations and analyses presented in this chapter are intended to make clarity on these topics. Do the need for categorizing experience in adaptive ways constitute a drive toward good individual signalling behavior? What else is needed for good communication to emerge at the level of the population? And do these factors, by themselves, render the communication system evolutionary stable?

3.3. The model

The model consists in a population of 100 organisms living in a one-dimensional environment (a corridor, shown in figure 3.1a). An individual's behavior is controlled by a neural network and its fitness is a function of the number of movements the individual makes and the number and type of mushrooms it eats during its life. There are 420 possible mushrooms, each different from all the others in its perceptual properties, encoded as strings of 10 bits in the visual input units of an organisms neural network. Half of the mushrooms are edible, in that they increase the fitness of the organism that eats them, while the other half are poisonous, in that they decrease fitness. The archi-

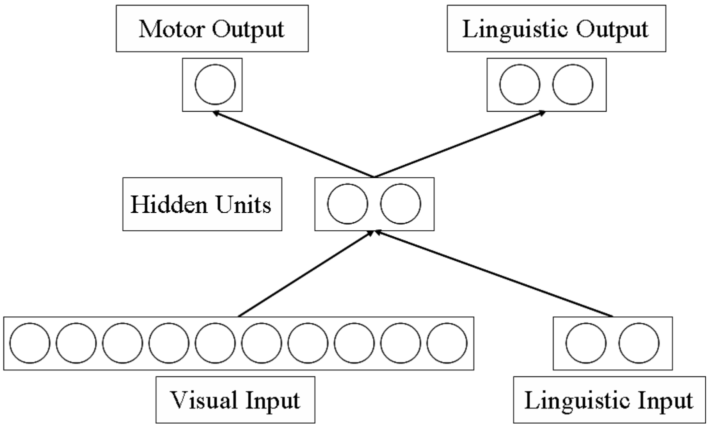


Figure 3.2: The neural network. Arrows represent connections from all the neurons of the sending group to all the neurons of the receiving group

tecture of the neural networks is shown in figure 3.2b: it includes 10 visual input units, 2 communicative input units, 2 hidden units, 1 motor output unit and 2 communicative output units. Networks connection weights are encoded as real values in the range [-4, 4] in the genome of the organisms. The visual range of our organisms is limited in that the visual input units encode a mushroom's properties only if the organism is in the cell immediately preceding the mushroom cell, otherwise the organism sees nothing, i.e., the activation of all visual input units is set to 0. The motor output unit of the network has a step activation function which determines whether the organism moves one step forward (if the activation is 1) or stays still (if the activation is 0). In our model signals consist in the two-dimensional vectors of the activations (in [-1, 1]) of the communicative output units of an organism (the speaker), which are copied in the communicative input units of another organism (the hearer).

Each individual of each generation lives for 420 trials, one for each mushroom. In each trial the individual is put in the start cell of the corridor, one of the mushrooms is put at the end of the corridor, and another individual is chosen randomly from the population to act as speaker. The speaker is placed near the mushroom and emits a signal through its communicative output units which is received by the tested organism (the hearer) through its communicative input units. The trial ends either if the hearer reaches the mushroom and eats it or after 11 input-output cycles, which guarantees that the hearer will never eat it. Fitness is calculated according to the following formula:

$$f(x) = \frac{an_e - bn_p - n_m}{norm} \tag{3.1}$$

where n_e and n_p are, respectively, the number of edible and poisonous mushrooms eaten by x, n_m is the number of movements made by x, a and b are two constants (set to 30 and 5, respectively), and $norm$ is the normalization factor (maximum possible fitness) which equals 4200.

After all the individuals of one generation have lived their lives, the next generation is created by selecting individuals with a probability proportional to their fitness (using the roulette wheel method with windowing), making them reproduce sexually (with double-point crossover), and mutating the genes (connection weights) of the offspring with a probability of 0,1% (for an explanation of the genetic algorithm, see Appendix B and Mitchell, 1996).

Simulations run for 2000 generations, during which I calculate various statistics, the most important ones being average fitness and what I call 'communication system quality'. For a communication system to be good, three conditions must apply: (a) objects belonging to the same category must be signalled in similar ways, (b) objects belonging to different categories must be signalled in different ways, and (c) all individuals must signal in the same way, meaning that the system is shared in the population. In the model, signals are two-dimensional vectors, which can be represented as points in a two-dimensional space, whose coordinates are the activations of the two communicative output units of speakers. Hence, the communication system of a population consists of two clouds of points: the points that represent the signals emitted by all the organisms in presence of all edible mushrooms and the points corresponding to the signals emitted in presence of all poisonous mushrooms. Condition (a) above is satisfied if the two clouds are small (meaning that mushrooms of the same category are signaled in similar ways). Condition (b) is satisfied if the two clouds are distant from one another (meaning that mushrooms belonging to different categories are signaled in different ways). Condition (c), which regards the sharedness of the communication system, is implicitly satisfied by the other two conditions, since the clouds represent the communication system of the entire population. Our communication system quality is calculated as the average between the two measures that make a communication system good: (1) the mean distance of the points of each cloud from the cloud's geometrical centre, and (2) the distance between the geometrical centres of the two clouds.[4]

---

[4] Basically, the way I calculate communication system quality is just the same way in which we can calculate the quality of a categorization process in a neural network, as disccussed in Appendix A, with only two differences. First, here we take the vectors of the communicative output units as the points which constitute the clouds, while in measuring categorization we take the vectors of the hidden units. Second, in the case of communication clouds are social, in that they include the signals (points) produced by all the organisms of one population, while the quality of categorization is (generally) an individual property. That is, the internal representations of objects which constitute the clouds belong all to the same individual.

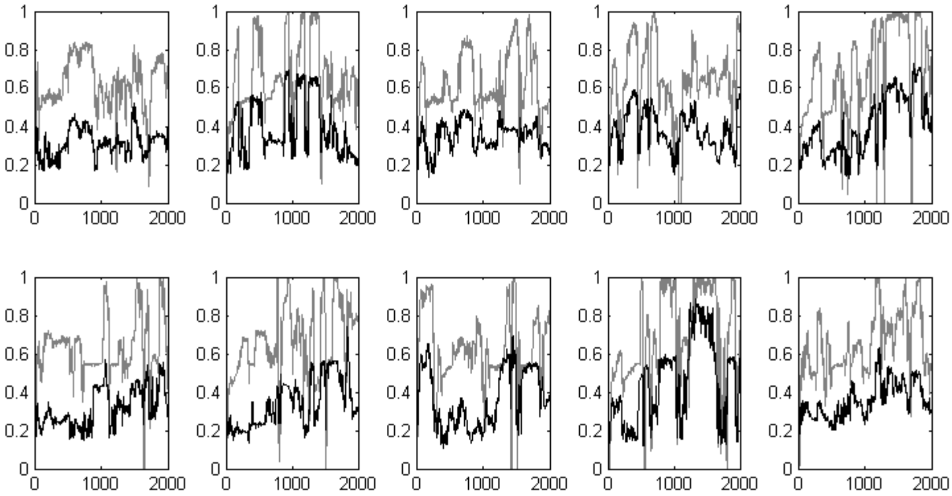


Figure 3.3: Average fitness (gray line) and communication system quality (black line) of 10 replications of the simulation.

3.4. Cognitive and Genetic Factors in the Evolution of Communication

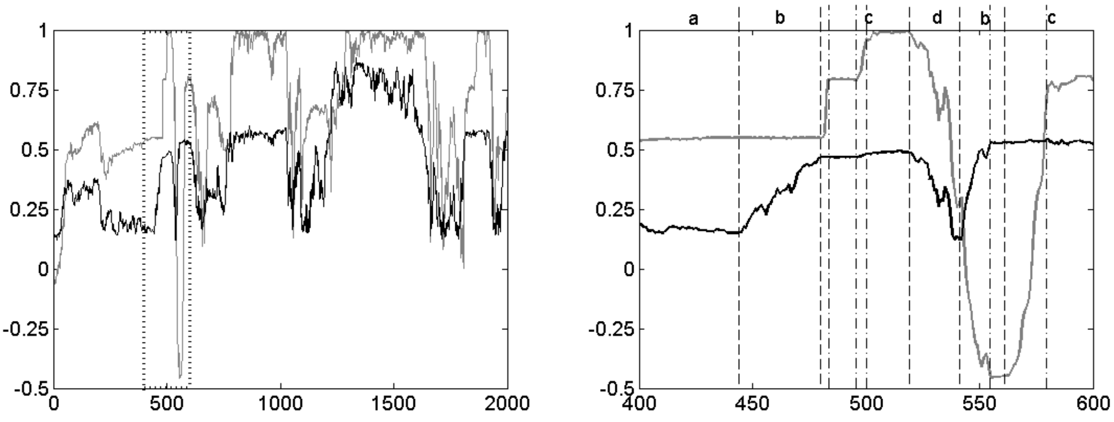
In my simulation's scenario the emergence of a good communication system would provide an advantage only to hearers. A hearer who receive good signals could in fact use the received information for deciding whether to approach mushrooms signaled as 'edible' or to stay still in the case the received signal stands for 'poisonous'. This would prevent the hearer to waste its energy by approaching poisonous mushrooms and thus would increase its fitness. For speakers, however, there is no benefit whatsoever in producing good signals. On the contrary, since individuals compete for reproduction, producing good signals is a purely altruistic trait in that, by giving an advantage to competitors, it indirectly decreases the reproductive chances of the good signaller.

The results of 10 out of 50 replications of the simulation are shown in figure 3.3. The results are quite surprising: average fitness and communication system quality fluctuate cyclically between very low and very high values. This means that communication emerges continually but it is also continually disrupted. As this very unstable pattern is present in all the replications of the simulation, the mechanisms that generate it must be very strong and reliable. What are these mechanisms?

The sudden drops in the communication system quality can be explained as consequences of the altruistic character of communication. Once a good communication system has emerged in the population bad speakers will be selected against good speakers because they will act as cheaters: they take advantage of the good signals they receive while misleading other organisms by producing bad signals. The result is that cheaters rapidly invade the population with the consequence of a sudden decrease in the quality of the communication system. But why a good communication system should emerge in the first place or re-emerge after being disrupted? The reason, I argue, lies in two factors that are constantly at work, pushing the communication system towards good quality.

The first factor is the 'cognitive pressure' suggested by Cangelosi and Parisi (1998). In order to survive and reproduce organisms must categorize perceived mushrooms according to their quality, moving forward in the corridor when they perceive edible mushrooms and staying still when they perceive poisonous mushrooms. For producing this kind of behavior, mushrooms must be appropriately categorized in the neural networks hidden units. In other words, in order for an organism to behave efficiently, its network's connection weights must be such that mushrooms belonging to the same category elicit similar activation patterns in the hidden units (similar internal representations), while mushrooms belonging to different categories elicit different internal representations. But the signal emitted by an organism when perceiving of a mushroom directly depends on the internal representation elicited in the hidden units by the mushroom. This means that the adaptive pressure to appropriately categorize mushrooms causes a pressure to produce good signals. Mushrooms belonging to the same category elicit similar internal representations, which, in turn, tend to elicit similar signals. Vice versa, mushrooms belonging to different categories elicit different internal representations which tend to elicit different signals.

However, this cognitive explanation of the spontaneous emergence of a good communication system is not enough. The cognitive pressure can in fact explain only the fact that each individual tends to produce good signals, that is, signals that systematically co-vary with the category of perceived mushrooms. It cannot explain why the communication system is shared in the population, that is, why different individuals tend to produce similar signals. But signals sharedness is necessary for good communication, and is indeed considered in our measure of the communication system quality. (Remember that the measure is taken at the populational level, considering the signals emitted by all individuals). In order to explain the sharedness of the communication system we need to introduce another factor, namely genetic convergence. In fact, if organisms have similar genes they will tend to have similar (categorical) internal representations and to produce similar signals. Genetic similarity can in turn be explained by genetic convergence due to selection. Only the fittest individuals reproduce in each generation. As a consequence, their genes will tend to spread in the population reducing genetic diversity.



(a)

(b)

Figure 3.4: (a) Average fitness (gray line) and communication system quality (black line) of one replication of the simulation. (b) Zooming in the 200 generations (from generation 400 to generation 600) enclosed in the dotted lines of graph (a). The dashed lines of graph (b) divide macro-evolutionary phase, while the dot-dashed lines subdivide this evolutionary period in the 11 sub-phases described in table 3.1. See text for details

3.5. Analyzing evolutionary dynamics

3.5.1. Understanding evolutionary dynamics by zooming in evolution

In the previous section I have given a possible explanation of the continuous rise and fall of communication in my model. This explanation is based on three factors: (1) the cognitive pressure toward spontaneous individual good signalling, (2) the genetic pressure towards the sharedness of the communication system due to the convergence of the population and (3) the pressure against good (shared) signalling produced by the altruistic character of communication. In order to test this analysis of the complex evolutionary dynamics demonstrated by the simulation, I took a single run of the simulation, selected a part of it (from generation 400 to generation 600), and tried to analyze this part in detail. Figure 3.4 shows the process of zooming in the 200 generation I selected. I choose this particular part of this particular run for my analysis because, looking at the dynamics of average fitness and communication system quality, it seemed to clearly exemplify the typical cycle of rise and fall of communication which repeats itself continually in all the runs.

As can be seen from the picture, I have divided those 200 generation in 11 sub-parts by the identification of evolutionary transitions. These transitions consist in significant changes in average fitness and/or communication system quality. For the sake of clarity, I have grouped these sub-parts in 4 macro-evolutionary phases which can be described as follows:

Phase a (generations 400-444): Communication system quality is low and communicative input is ignored, resulting in an average fitness of about 0.55, which is the maximum value that can be reached without the aid of communication;

Phase b (generations 445-480): Communication system quality improves because of the cognitive and genetic pressures towards good communication;

Phase c (generations 481-519): Good signalling starts to be exploited and fitness increases until it reaches almost maximum value;

Phase d (generations 520-541): Bad speakers are selected against good ones because they cheat others and invade the population: the result is a sudden decrease in the communication system quality and, consequently, fitness.

Once the communication system quality has reached its minimum it starts to grow again, thanks to the cognitive and genetic pressures (second occurrence of phase b: generations 542-561). This lasts until the quality of the communication system reaches a very high value, communication re-starts to be exploited, and fitness increases again (second occurrence of Phase d: generations 562-600). A more detailed analysis of the phenomenology of this 200 generations is the following:

1. The communication system is bad (signals do not convey information on the quality of mushrooms) and, therefore, it is ignored (generations 400-444).
2. The communication system improves gradually (generations 445-480). Since there is no selective pressure towards good communication and signals are ignored, this increase in communication system quality seems to be due to the neutral evolution produced by the two hypothesized factors, that is, the cognitive pressure and the genetic convergence of the population.
3. Once signals have become potentially informative, mutants appear that exploit them. The exploitation of communication results in higher fitness and the 'good-hearing' genes rapidly spread in the population (generations 481-485).
4. For a few generations (486-495) the population remains in a sub-optimal situation in which communication is good but is only partially exploited.
5. Other mutations result in a better capacity to exploit communication, which translates into another rapid evolutionary phase, leading to the full exploitation of communication by all individuals (generations 496-500).
6. For about 20 generations (501-519) communication stays good and is fully exploited, resulting in an optimal average fitness.

7. Individuals appear that produce ’bad’ signals, which means either signals that do not differentiate between poisonous and edible mushrooms or signals that do contain information about the quality of the mushrooms but encode it in a way which is different from that currently used in the population. Those mutating individuals are cheaters, in that they do not inform reliably the others about the quality of mushrooms. Cheaters tend to have higher fitness and invade the population, resulting in a decrease of both communication system quality and average fitness (generations 520-541).

8. After having reached its minimum, the quality of the communication system increases again because of neutral evolution due to the cognitive and genetic pressures. In the meantime, average fitness keeps on decreasing. This suggests that organisms keep on reacting to signals in the same way as they were used to, while the emerging communication system has changed. In other words, edible and poisonous mushrooms are signalled in a reliable way, but in a way that is different with respect to the communication system which was present before the invasion of the population by cheaters.

9. For a very few generations (556-561) the communication system is good but ’misinterpreted’, and average fitness is very low.

10. As mutating individuals appear that can interpret correctly the signals emitted by others, they obtain very high fitness and their genes spread rapidly in the population, resulting in a fast evolutionary phase (generations 562-580) during which average fitness improves dramatically.

11. Finally, during the last 20 generations (581-600) the communication system remains good and is partially exploited by the individuals.

This analysis of the 200 generations I am dealing with here is summarized in table 3.1, where I have classified the evolutionary phases just described as phases of equilibrium (E), of neutral evolution (NE) and of fast evolution (FE). The reason for doing so is that the theory of punctuated equilibria (Eldredge and Gould, 1972; Gould and Eldredge, 1993) seems to apply quite well to our simulations: there are equilibrium phases during which fitness does not substantially change, punctuated by phases of rapid evolution (in which fitness either rapidly increases due to the exploitation of communication or rapidly decreases because cheaters invade the population). Furthermore, as I will explain below, the rapid increases in fitness are made possible by the neutral evolution of good communication due to genetic drift, the linking between non-communicative and communicative behaviours in the organisms’ neural networks, and genetic convergence of the population.

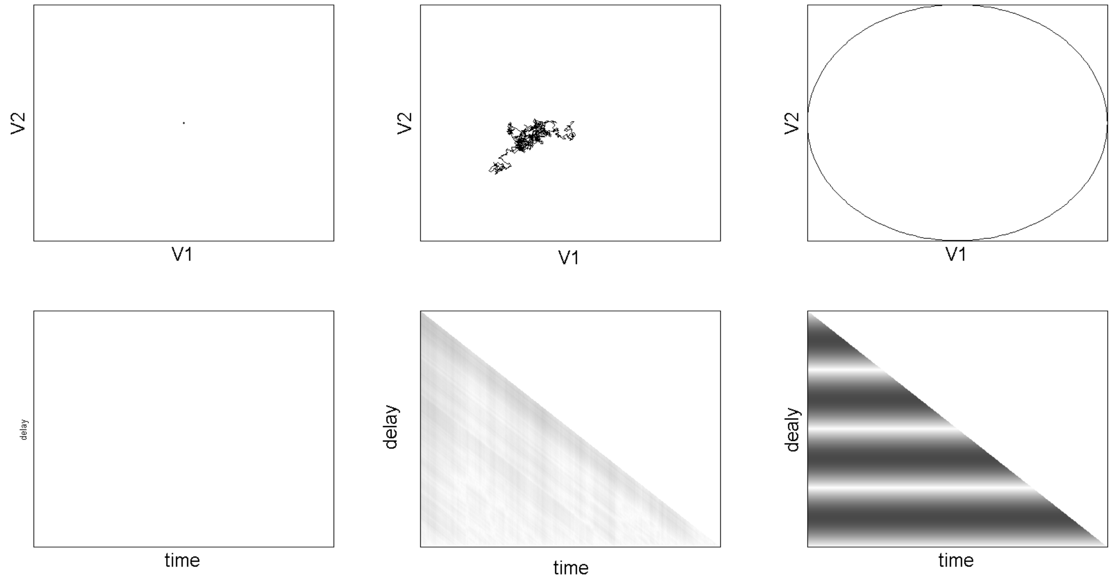
phase	generations	description	kind
1	400-444	communication is bad and ignored	E
2	445-480	communication improves gradually	NE
3	481-485	fitness improves rapidly	FE
4	486-495	communication is good and partially exploited	E
5	496-500	fitness improves rapidly	FE
6	501-519	communication is fully exploited	E
7	520-541	communication system quality and fitness decrease	FE
8	541-555	fitness decreases but communication improves	NE
9	556-561	communication is good but not understood	E
10	562-580	fitness improves rapidly	FE
11	581-600	communication is good and partially exploited	E

Table 3.1: Evolutionary stages of the analysed simulation (E = equilibrium, NE = neutral evolution, FE = fast evolution)

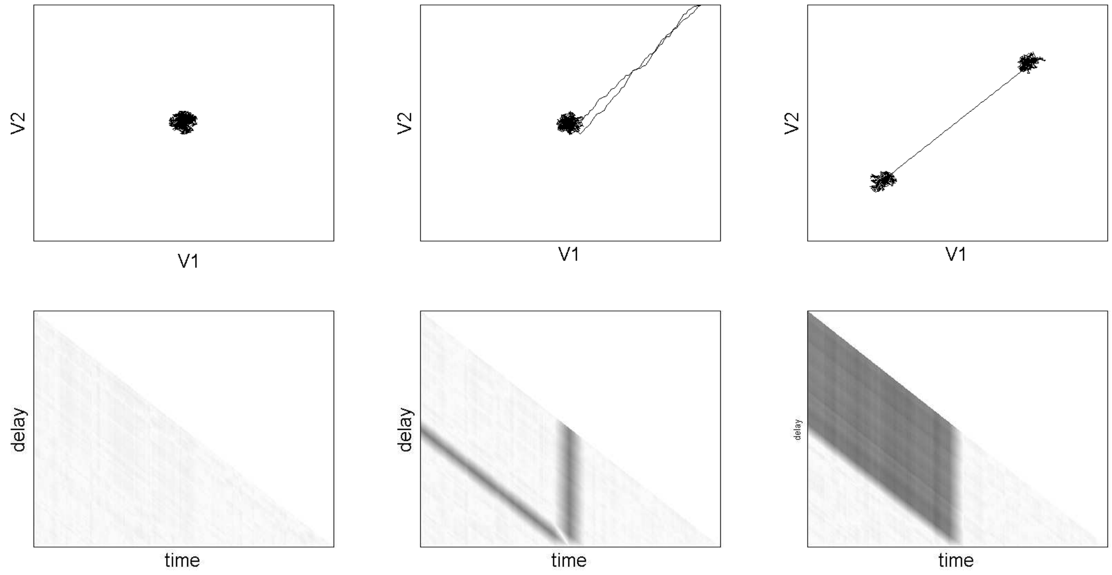
3.5.2. The population as a dynamical system

In order to understand which are the mechanisms that underlie the rich evolutionary dynamics described in the previous section, I have applied a (sofistication of the) method presented by Burtsev (2004). An evolving population can be represented as a cloud of points in a multi-dimensional genotype space. In our case, each dimension corresponds to one of the 32 neural networks' connection weights (encoded as real values in the genome). The points correspond to the individuals of the population and each point's coordinates are the values of their genes (weights). In order to understand how the population moves in its space we can consider the centroid of the population, that is the geometrical centre of the points representing the individuals of each generation. From the dynamical systems point of view, the centroid is a system moving in discrete time (time-steps correspond to generations), in a 32-dimensional continuous space (with each dimension bounded in [–4;4]), with a very complex evolution operator (which is implicitly defined by the rules which govern individual life and the selection, reproduction, and mutation operators) and highly stochastic: there is stochasticity in the individuals' interactions and in all the genetic algorithm's operators.

One way to start understanding the trajectory of this dynamical system is to create a bitmap in which the grey scale represents the displacement of the system for different time delays. The x axis represents generations, while the y axis represents time delays: for each point in the bitmap, the level of grey represents the distance between the centroid at generation x and the centroid at generation $x + y$. The rationale for doing so is that such a figure can indicate how the system moves in time. Figure 3.5 represents six basic kinds of movements that a two-dimensional system can perform together with the corresponding bitmaps.

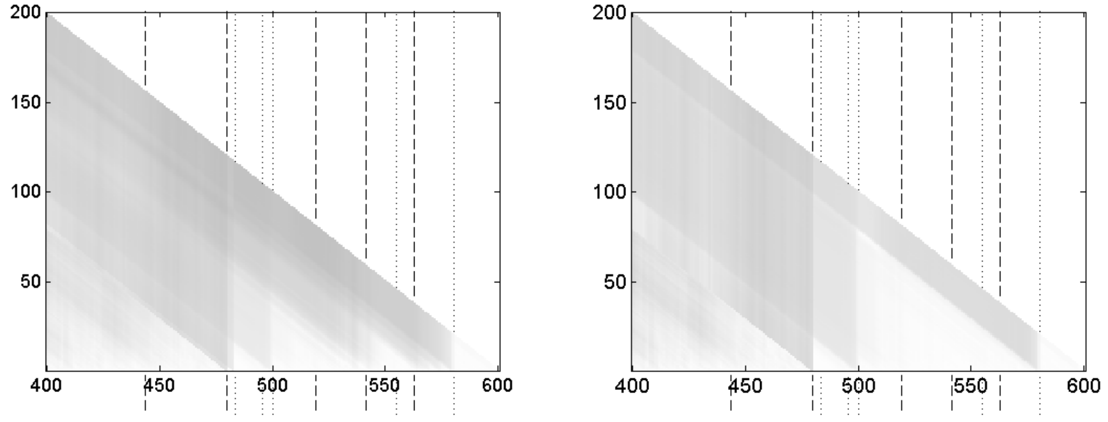


(a) fixed point (b) random walk (c) limit cycle



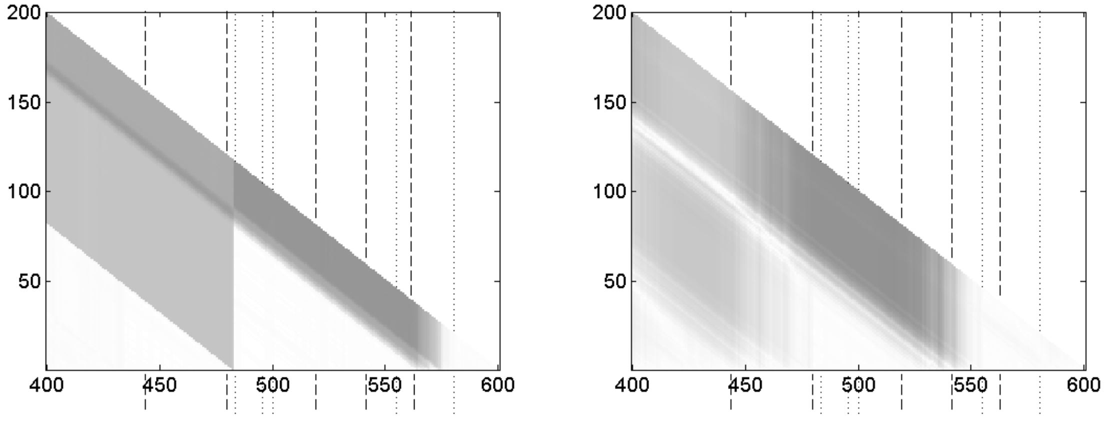
(d) around an attractor (e) attractor, jump and back (f) jump between attractors

Figure 3.5: 6 kinds of movement of a two-dimensional dynamical system (top), and the respective diagrams of the displacement after delays (bottom). (a) A fixed point generates a white diagram, because the displacement is always null. (b) A random walk generates a gray diagram, which tends to get darker in the upper part. (c) A limit cycle generates a dark diagram with horizontal white lines, with the distance between the lines corresponding to the period. (d) An erratic movement around an attractor generates a uniform light gray diagram. (e) A point which moves around an attractor, jumps away and then returns generates an oblique and a vertical dark line (corresponding to the jump) on a light gray diagram. (f) A jump between two attractors generates two light grey triangles, corresponding to the periods in which the point is around the two attractors, divided by a darker parallelogram, which corresponds to the distance between the two attractors



(a) all-weights

(b) non-communicative



(c) communicative-input

(d) communicative-output

Figure 3.6: Diagrams of the displacement of the centroid of the population after delays for the entire system (a) and the three functional sub-systems: non-communicative (b), communicative-input (c) and communicative-output weights (d). Lines divide the 200 generations under analysis in the evolutionary phases described in table 3.1. See text for details.

Figure 3.6a shows the bitmap for the centroid of the population in the 200 generations we are analyzing. Since even in this very limited evolutionary period the system has a quite complicated phenomenology, summarized in table 3.1, it is no surprise that the bitmap representing the trajectory of the population is much more confused than any of the bitmaps in figure 3.5. Nonetheless, the fact that small discontinuities between grey regions seem to appear in correspondence with the evolutionary stages that we have recognized (see table 3.1) is promising.

As different transitions between the stages would presumably depend on different sets of genes, I modified the method of Burstev (2004) in the following way. I divided the system in three functional sub-systems: non-communicative, communicative-output, and communicative-input. In order to do that, I grouped genes according to their roles in communication (see figure 3.8), and re-analyzed the population as represented by three centroids, one for each functional group of connection weights (genes). and created a bitmap for each of them. The bitmaps for the non-communicative, communicative-input, and communicative-output sub-systems are shown in figure 3.6b, c, and d, respectively. The division of the system in functional sub-systems appears to be rewarding: the new graphs are much clearer than the preceding one, and give us useful insight into the dynamics of the system.

If we look at the bitmap of the non-communicative centroid we see that:

- Before generation 480 (that is, during phases 1 and 2 described in table 3.1) the centroid is slightly moving around an attractor.
- At generation 480 there is a sudden jump to another attractor, where the system stays for about 20 generations. The fast evolution of phase 3 is due to this transition of the non-communicative centroid from an attractor which does not permit the (partial) exploitation of communication to one which does.
- Between generation 495 and generation 500 the system makes another (smaller) jump to a third attractor, in which the population can fully exploit communication (this explains the fast evolution of phase 5).
- Finally, still another jump to a different location in the non-communicative genetic space appears just before generation 580, that is, during the fast evolution of phase 10 which permits the population to exploit communication again.

Even clearer is the reading of the bitmap regarding the communicative-input populational centroid (figure 3.6c). During the 200 generations under analysis, this centroid occupies mainly three different locations in its space (the first during generations 400-480, the second during generations 481-560, the third during generations 580-600), with two clear transitions:

- The first transition is the sudden jump happening at generation 480, which permits (together with the jump of the non-communicative centroid discussed above) the fast evolution of phase 3.
- The second transition is more gradual and happens between generation 560 and generation 580. During this transition, which determines the fast evolution of phase 10, the population re-adapts its communicative-input weights (together with the non-communicative weights) so that individuals can exploit communication again.

Finally, figure 3.6d shows the bitmap regarding the communicative-output connection weights. The reading of this figure also is quite straightforward. The centroid of the communicative-output genes clearly occupies three distinct locations in between generations 400-445, 480-520, and 555-600, while passing through two gradual transition phases:

- The first transition, between generation 445 and 480, corresponds to the neutral evolution of phase 2, during which communication system quality gradually increases.
- The second transition, between generation 520 and 555, corresponds to phases 7 and 8 during which communication is disrupted by the invasion of cheaters (phase 7) and then improves again due to neutral evolution (phase 8).

Overall, the bitmaps of the dynamics of the three functional subsystems in which we have divided the genome provided the following insights:

- during phase 1 the population is moving (apparently randomly) in a region of the non-communicative genes space which is neutral with respect to fitness;
- during phase 2 the population moves (for a neutral evolution in which a role is played by both the cognitive and genetic pressures discussed above) towards a location in the communicative-output genes space in which communication is good;
- the rapid evolution of phase 3 is due to selective pressures in both the non-communicative and communicative-input genes which result in the almost sudden jump of the population in regions of the space which permit the exploitation of communication;
- the other rapid increase in average fitness happening in phase 5 is due to another jump of the population which takes place only in the non-communicative genetic sub-space;
- the rapid decrease of fitness during phases 7 and 8 is due to the displacement of the communicative-output centroid from a point which determines good communication to another one, passing through regions in which communication system quality is low, which determines the inability of individuals to correctly interpret received signals;
- finally, the rapid evolution of phase 10 is due to the displacement of both the non-communicative and the communicative-input centroids which allows the re-adaptation of the population to the new communication system.

3.5.3. Analyses

The interpretation of the population as a dynamical system and the visualization of the displacements of three functional subsystems in their respective genetic spaces has given us various insights about the evolutionary dynamics of the system we are analysing. But we can go even further. In order to do that, I have collected a number of statistics and plotted them together in figure 3.7.

Figure 3.7a shows the variances of non-communicative, communicative-input and communicative-output genes as they change during the 200 generations under analysis. Genetic variance for each weight x (V_x) is calculated, standardly, according to the following formula:

$$V_x = \frac{\sum_{i=1}^{N}(\bar{w}_x - w_x^i)^2}{N} \tag{3.2}$$

where N is the number of individuals in the population (i.e. 100), $\bar{w}_x$is the mean value for weight x and w_x^i is the weight x of the i_{th} individual. Consequently, the variances (V^k) for the three categories ($k \in K$) of weights (non-communicative, communicative-input, and communicative-output weights) are calculated as follows:

$$V^k = \frac{\sum_{x=1}^{N^k} V_x}{N^k norm} \tag{3.3}$$

where N^k is the number of weights of the k_{th} category (i.e. 28, 4 and 4 for non-communicative, communicative-input and communicative-output weights, respectively), V_x is the variance on weight $x \in k$ and norm is the normalization factor, so that V^k is in [0;1].

Figure 3.7b shows the displacement of the three functional centroids (non-communicative, communicative-input and communicative-output) between generation x and generation $x-1$. Formally, I plotted the following function:

$$y(k) = dist(C_x^k, C_{x-1}^k) \tag{3.4}$$

where k is the plotted category (non-communicative, communicative-input and communicative-output), C_i^k is the centroid of category k at generation i, and $dist(a, b)$ is the standard euclidean distance between points a and b.

Figure 3.7c represents the evolution of the populational communication system. As discussed above, as signals are represented by points in the two-dimensional space of the communicative-output units of organisms, the communication system can be represented as the two populational centroids of the points that represents signals emitted in response to edible mushrooms and signals emitted in response to poisonous ones, respectively. In figure 3.7c I have plotted the two coordinates of the two centroids of the signals used for edible and poisonous mushrooms.

Figure 3.7d shows the comparison between four values: the quality of the populational communication system, the average quality of individuals' communication systems, the quality of the populational representation system and the average quality of individuals' representation systems. As discussed above, the communication system quality that I have discussed so far is calculated on the centroids of signals emitted by all the individuals of one generation for all the edible and poisonous mushrooms. But we can also calculate the quality of the individuals' communication systems: for each individual we calculate its communication system quality by considering the centroids of the signals emitted by that individual. By averaging those quantities over all individuals in a population we can look at the quality of individual communication systems regardless of the sharedness of communication in the population.

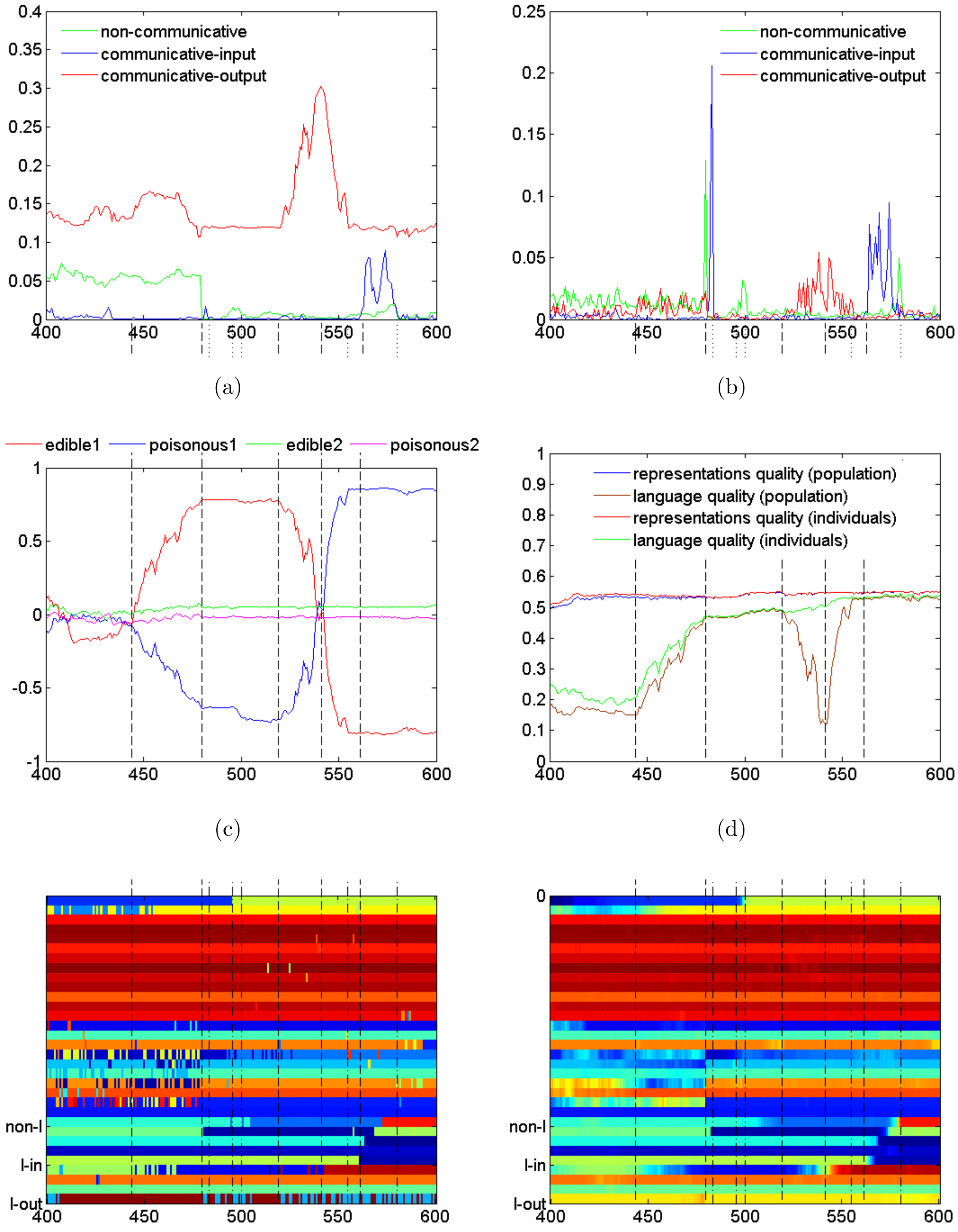
Just as the signal emitted by an individual for a given mushroom is the two-dimensional point whose coordinates are the activations of the two communicative output units of that individual for that mushroom. Consequently, the internal representation of that individual for that mushroom is the point whose coordinates are the activations of the individual's hidden units when it sees the mushroom. We can calculate both populational and average individual representation qualities just in the same way as we calculate communication system quality, but considering the points represented by the individuals' internal representations of mushrooms in their hidden units instead of those represented by the signals emitted through the communicative output units.

The last two images of figure 3.7 show the evolution of the genes of the best individual of each generation (figure 3.7e) and of the populational centroid (figure 3.7f). The two images are bitmaps, where each point represent in a gray-scale the value of one of the 32 genes (connection weights) of a given generation. As usual, genes are grouped according to their functional role (non-communicative, communicative-input, and communicative-output).

3.5.4. Results

In this section, we will try to understand in detail the mechanisms that underlie the 11 evolutionary phases we have singled out (see table 3.1) by looking at the various statistics described in the previous section and plotted in figure 3.7.

Even though phase 1 is an evolutionary stable phase, in which neither fitness nor communication system quality change significantly, during this phase the populational centroid moves slightly in the non-communicative genetic sub-space, while it does not move significantly in the other two sub-spaces (figure 3.7b). This continuous displacement of the non-communicative centroid during phase 1, already visible in the bitmap of figure 3.6b, is due to the competition between a few genotypes which have different



(a) (b)

(c) (d)

(e) (f)

Figure 3.7: Evolutionary dynamics during generations 400-600. (a) Average variances of the non-communicative, communicative-input and communicative-output genes. (b) Displacement of the non-communicative, communicative-input and communicative-output populational centroids. (c) Coordinates of the two centroids of the signals used for edible and poisonous mushrooms. (d) Quality of the communication and representation systems of the population and average quality of individuals' systems. (e) Gray-scale bitmap of the genes of the best individual of each generation and (f) of the populational centroid (genes are grouped as non-communicative (non-C), communicative-input (C-in) and communicative-output (C-out)). The lines on or under the graphs correspond to the evolutionary phases described in table 3.1.

non-communicative genes, but are phenotypically quite similar in that they provide comparable fitness. This is shown by the fact that during this period some of the non-communicative genes of the best individuals change continually (specifically genes number 2, 14, 17, 19, 20, 21 and 22, see figure 3.7e), resulting in corresponding changes in the populational centroid (figure 3.7f). The simultaneous presence of competing genotypes during this phase is demonstrated also by the existence of some genetic variance in the non-communicative genes (figure 3.7a).

Phase 2 is characterized by the slow increase in the communication system quality of the population. How and why does this happen? During this phase, the activity in the non-communicative genetic space due to the competition between genotypes with equivalent fitness goes on, as demonstrated by the graphs of figure 3.7a, b, e and f. However, during this competition the best individual of the population starts to have a different communicative-output gene (the first one, see figure 3.7e). As a consequence, the communicative-output populational centroid starts to slowly displace itself (figure 3.7b) until the population has converged on the new value of the mutated gene (figure 3.7f). Remember that during this phase signals are ignored because they are uninformative; as a result, there is no selective pressure whatsoever regarding communicative-output genes. Hence, the new communicative-output gene spreads in the population just for genetic drift. Notwithstanding the variability in the non-communicative genes, the population has already converged at the level of internal representation: the populational representation quality is in fact very good, almost as good as individual categorization quality (figure 3.7d). This convergence on categorized and shared representations of mushrooms makes it possible for the spreading in the population of the new communicative-output gene to bring about a simultaneous increase in both individual and populational communication system quality. In fact, the new communicative-output gene just makes it manifest in the communicative-output units (in particular in the first unit, see figure 3.7c) the categorization of mushrooms which is already present in the hidden units.

Phase 3 is characterized by the sudden increase in the average fitness of the population. This is achieved by two sudden jumps of the populational centroid which permit to the individuals to exploit, even only partially, the good communication system which now exists. The first jump happens in the non-communicative sub-space and is visible in the high peak in the displacement of the non-communicative centroid at generation 480 (figure 3.7b). Evidently, in this generation one individual was born with a mutation in the non-communicative units that allows it to correctly interpret the signals emitted by other. This individual obtains a much higher fitness than all the other individuals of its generation and its genes immediately spread in the next generation. As a result, the competition between non-communicative genotypes with comparable fitness stops immediately (this is visible in the immediate stabilization of both the best and centroid's non-communicative weights, figure 3.7e and f), and the population suddenly converges to the new non-communicative genotype (see the sudden drop in the non-communicative genetic variance in figure 3.7a). The second jump happens in the communicative-input space after a few generations, which is visible in the peak of the displacement of the communicative-input centroid (figure 3.7b) and in the sudden change of the first communicative-input genes of the populational centroid (figure 3.7f) which follows, after a few generations, the same change in the genotype of the best individual of the population (figure 3.7e).

Phase 4 is an equilibrium phase, during which nothing important happens. The only interesting thing to be noted is the continuous alternation in the best individual's last (communicative-output) gene (figure 3.7e). This alternation clearly explains the relatively high genetic variance in the communicative-output genes during all the 200 generations we are discussing here (figure 3.7a). Evidently, two alleles of the last gene co-exist in the population throughout all this period, due to the fact that during this period this gene has no effect on fitness nor on communication system quality. This co-existence is demonstrated by the fact that during all the period we are discussing the populational centroid has not one of either 'alleles' for that gene, but rather a mean between the two (figure 3.7f). This means that about half of the population possess one of the two alleles while the other half possesses the other.

During phase 5 there is a second sudden increase in average fitness. This is due to the rapid displacement of the non-communicative populational centroid (figure 3.7b). In particular, this phase starts with a mutation in the first (non-communicative) gene which allows the mutant individual to fully exploit communication and hence have higher fitness (figure 3.7e), and ends when the population has converged on that gene (figure 3.7f) and all organisms can thus obtain almost maximum fitness.

Phase 6 is extremely stable, and nothing interesting happens until a new mutation occurs in the first communicative-output genes which makes the mutant individual produce signals which are different from those of all the other organisms in the population. At the very beginning of phase 7, we see that the best individual happens to be one with a new allele for that gene (figure 3.7e). This new allele starts to spread in the population, even though slowly, because individuals that possess it cheat the others. When they act as speakers, they produce the signal that is interpreted as meaning 'edible' when there is a poisonous mushroom and the signal interpreted as meaning 'poisonous' when there is an edible mushroom. As a consequence, hearers are mislead and obtain lower fitness. The fact that the new allele makes individuals invert their signals, and does not simply make individuals produce signals that do not distinguish between edible and poisonous mushroom is demonstrated by the following fact. While during phase 7 the quality of the populational communication system decreases, the quality of individual communication systems slightly increases (figure 3.7d). This means that the organisms with the new alleles in the first communicative-output gene produce, individually, even better signals than those with the old alleles, but the new signals have opposite meaning with respect to the old ones.

When the new allele has spread in half of the population, the communication system quality of the population has reached its minimum. In fact, half of the organisms signal 'edible' with a high value in the first communicative-output unit and 'poisonous' with a low value in the same unit, as in the preceding generations, while the other half, with the new allele, does the opposite, signalling 'edible' with a low value in the first communicative-output unit and 'poisonous' with a low value. This explains why at generation 540 the mean value of the first communicative-output unit is 0 for signals emitted for both edible and poisonous mushrooms (figure 3.7c), while individual communication system quality is quite high (figure 3.7d). The presence of two opposite alleles in the population of the first communicative-output gene explains also the peak in the genetic variance in the communicative-output genes (figure 3.7a). Now there is high variance not only in the last non-communicative gene but also in the first one, which produces a variance that is about twice as great as that of preceding generations. Finally, the simultaneous presence of the two alleles is also demonstrated by the following fact. While the first communicative-output gene of the populational centroid changes from a very low value (blue) through the mean value (0, represented by green) to a very high value (red, see figure 3.7f), the best individuals have always either the blue allele or the red one (figure 3.7e). This means that the change in the populational centroid is due to a change in the proportion of the presence of the two competing alleles in the population.

Phase 8 is just a continuation of phase 7. The population keeps on moving in the communicative-output space (figure 3.7b) towards convergence on the new allele for the first communicative-output gene (figure 3.7a) because this gene keeps on providing a relative selective advantage due to cheating. As a result, populational communication system quality starts to increase again (figure 3.7d), as the population converge to the new signalling behavior: low values in the first communicative-output unit for 'edible' and high values for 'poisonous' (figure 3.7c).

The reason why average fitness keeps on decreasing during phase 8 while communication system quality increases is just the same that explains the convergence of the population to the new allele for the first communicative-output gene. Even though the new communicative-output gene makes speakers produce good signals, those signals are misinterpreted by hearers, which have communicative-input genes adapted to the previous (opposite) communication system. As a result, average fitness maintains itself in its minimum value for a few generations (phase 9).

During phase 10 a series of mutations occur that allow the population to re-adapt to the new communication system. Chronologically, changes in the fourth, second, and first communicative-input genes and in the last non-communicative gene of the best individuals (figure 3.7e) are rapidly followed by analogous changes in the populational centroid (figure 3.7f). This makes the populational centroid rapidly displace itself in both the communicative-input and non-communicative sub-spaces (figure 3.7b). At generation 580 the population has converged in both the communicative-input and non-communicative genetic spaces (figure 3.7a), and phase 11 starts, during which the system is stable and nothing relevant happens.

3.6. The no-cognitive-pressure and no-communication simulations

In the previous section I have analyzed in detail the evolutionary dynamics of 200 generations of a single run of our simulation in order to understand the evolutionary pressures that underlay the complex behavior it manifested. This analysis clearly confirmed my hypotheses on the causes underlying the continuous rise and fall of communication (and, consequently, of fitness) in the model. But even if the analysis has been proved to be correct for the evolutionary period I have studied, we cannot still be sure that the same is true for all the periods of all the replications of the same simulation. In other words, the analysis of a single run is not enough for the extrapolation of general principles. In order to test whether the tendency towards the emergence of communication due to cognitive and genetic pressures is general, I compare the average results of 50 replications of the simulation I have discussed so far (which from now on I will call the base-line simulation) with those of two new simulations, which I call the no-cognitive-pressure and the no-communication simulations.

3.6.1. Simulations

The no-cognitive-pressure simulation runs exactly as the base-line one except for the architecture of organisms' neural networks (figure 3.8b). The difference is that the neural networks of the no-cognitive-pressure simulation have two groups of hidden units (composed by 2 neurons each). The first group receives connections from both the visual and the communicative input units and send connections to the motor output units. The second group receives connections from the visual input units and sends connections to the communicative output units. In this way I artificially avoid any impact of categorization on communicative behaviour: since the communicative output does not depend on the same neurons on which the motor output does, with the new architecture there cannot be any cognitive pressure towards good signal production.

Organisms of the no-communication simulation have the same neural network as the one used in the base-line simulation (figure 3.8a) but in this case communication is

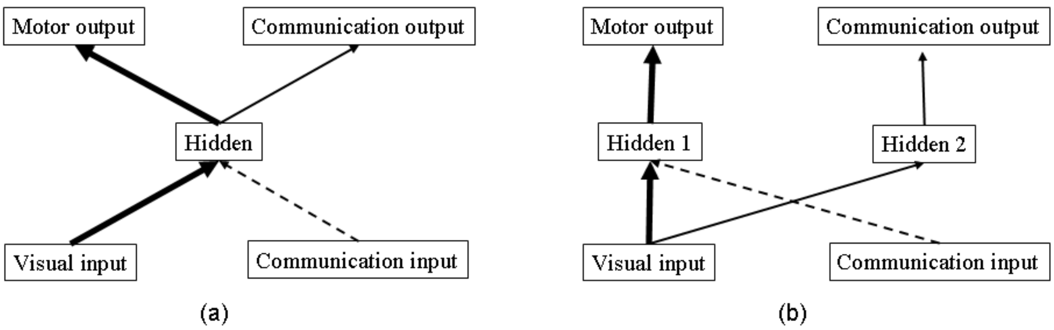
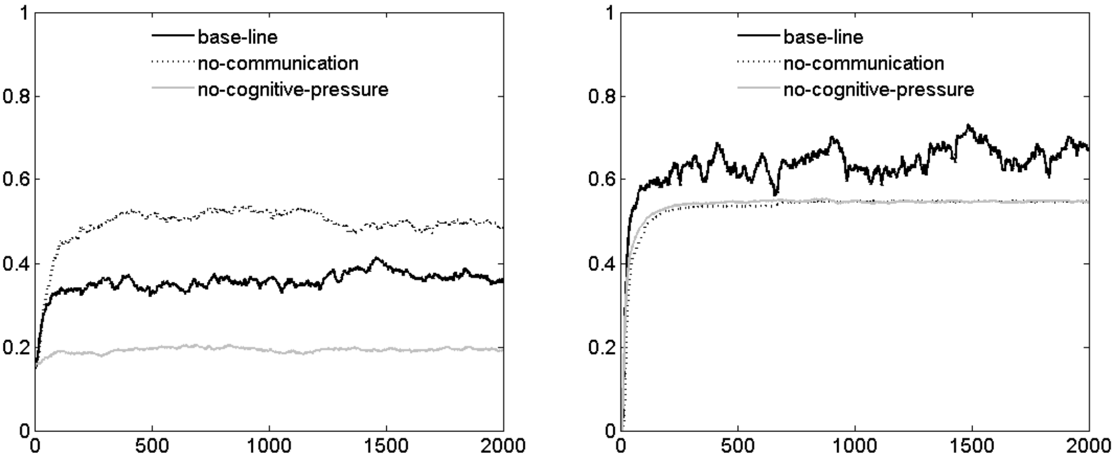


Figure 3.8: Comparison between the neural network used in the base-line and no-communication simulations (a) with that used in the no-cognitive pressure simulation (b). Connection weights are grouped according to their roles in communication: non-communicative (thick arrows), communicative-output (thin arrows) and communicative-input (broken arrows).

not allowed. In other words, in this simulation the organisms never act as speakers. Consequently, tested individuals never receive any communicative input. Under such conditions the only way for an individual to behave correctly is to move towards the mushroom until the mushroom is close enough to be clearly perceived. At that point, the organism can decide whether to eat the mushroom or refrain from doing so according to the perceived mushroom's quality. As a consequence, the maximum fitness that can be obtained by the individuals of the no-communication simulation is 0.55 because these individuals must always pay the cost of moving towards mushrooms. This cost could be avoided, when mushrooms happen to be poisonous, by organisms that receive reliable signals (in the simulations with communication). Note that in the no-communication simulation the communicative weights (those that link the communicative input and output units with the hidden units) are never used. Hence, they are adaptively neutral and subject to genetic drift. Nonetheless, we can still test what signals are produced by the networks located near to a mushroom and, consequently, measure the communication system quality of the no-communication simulation even if the communication system is never used.

3.6.2. Results

Figure 3.9 shows the results of the 3 simulations (base-line, no-communication, and no-cognitive-pressure) in terms of communication system quality (a) and average fitness (b). The results, averaged over 50 replications of each simulation, clearly support the hypothesis regarding the cognitive pressure towards the emergence of a good communication system. The communication system quality in the no-communication simulation is quite high (about 0.5, see figure 3.9a), in particular, significantly higher then in the base-line simulation. This means that the need to categorize mushrooms



(a)

(b)

Figure 3.9: Communication system quality (a) and average fitness (b) of the base-line, no-communication and no-cognitive-pressure simulations. Average results of 50 replications.

and the genetic convergence of the population are sufficient to produce a good communication system even in a population in which communication is not allowed and the communicative-output weights are subject to genetic drift. Furthermore, the fact that the quality of the communication system in the no-communication condition is considerably higher than in the base-line condition (about 0.35) means that, if signals are used, there is a direct selective pressure *against* good speaking, due to the altruistic character of communication.

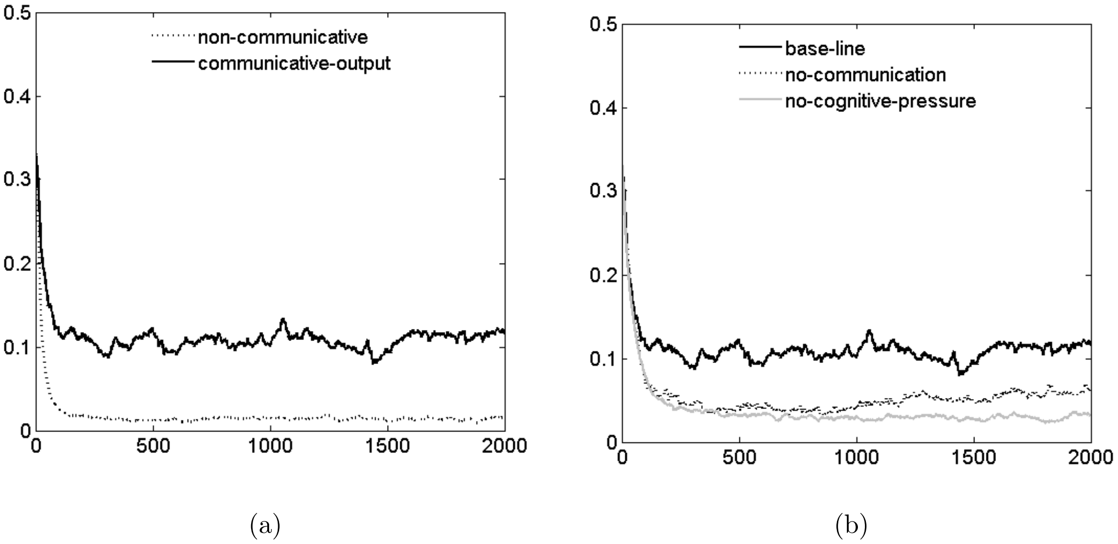
The results of the no-cognitive-pressure simulation confirm our analysis. If we prevent any influence of cognition on signal production by manipulating the architecture of the neural network, a good communication system never emerges. This results in an average communication system quality of about 0.2, which is considerably lower than that of the base-line condition.

The results in terms of average fitness (figure 3.9b) are quite clear. In both the no-communication and the no-cognitive-pressure simulations fitness reaches the maximum value reachable by organisms which cannot get any benefit from communication: in the first case because communication is not allowed, in the second case because signals are useless and cannot be exploited. Since the organisms of the base-line simulation can sometimes take advantage of communication, their average fitness is a little (but not much) higher.

In order to directly measure the genetic similarity of the population I have also analyzed the genetic variance of the connection weights during evolution, dividing them into the three functional categories used in the analysis of the populational dynamics of section 3.5.2: non-communicative, communicative-input, and communicative-output weights (see figure 3.8). The reason for doing this is twofold. First, I wanted to determine the genetic convergence of the population, which is necessary, according to my hypothesis, for explaining the sharedness of the communication system which spontaneously emerges in the base-line simulation. Second, since my explanation postulates different genetic pressures on different sets of genes, I wanted to determine whether those different genetic pressures would result in different genetic variances (variances are calculated according to formula 3.3).

In figure 3.10a I have plotted the variances of non-communicative and communicative-output weights of the base-line simulation (as usual, I show the average results of 50 replications). There is indeed a considerable genetic convergence since both variances are very low, which explains the sharedness of the spontaneously emerging communication system. But the communicative-output weights have a much higher variance (about 0.1) than the non-communicative weights (about 0.01). The explanation is to be found in the selective pressure against good signal production due to the altruistic character of communication. The cheaters which invade the population when the communication system is good are those organisms that, due to mutations, have communicative-output weights that are different from those on which the population has converged and which, consequently, let them produce misleading signals. In other words, the variance of the communicative-output weights is higher than the variance of the non-communicative weights because each time the population has genetically converged and the quality of the communication is high, there is a strong and reliable pressure to possess communicative-output genes that differ from those of the population in order to cheat competitors.

This analysis is further confirmed by the comparison between the variances on the communicative-output genes in the three conditions: base-line, no-communication, and no-cognitive-pressures (figure 3.10b). The three different values reflect the different pressures on those genes in the three conditions. In the no-cognitive-pressure simulation there is no spontaneous tendency to produce a good communication system, and the low variance of communicative-output weights (about 0.03) is explained by the genetic convergence of the population. The fact that the variance of communicative-output weights in the no-communication condition is higher (about 0.06) and tends to grow after reaching its minimum is due to the fact that in that simulation there is no pressure whatsoever on the communicative-output genes, so they are subject to genetic drift. On the contrary, in the no-cognitive-pressure simulation there is some mild pressure to generate bad signals. Nonetheless, the variance of the communicative-output genes is still higher in the base-line simulation than in the no-communication simulation. This confirms the hypothesis that when communication is used there is a reliable selective pressure against good speakers, resulting in a pressure against the genetic convergence



(a)

(b)

Figure 3.10: Average variances on the non-communicative and communicative-output weights of the base-line simulations (a) and average variances on the communicative-output weights of the base-line, no-communication and no-cognitive-pressure simulations (b). Average results of 50 replications of the simulations.

on the communicative-output weights: mutations on those genes that generate cheaters are rewarded by selection and spread in the population.

3.7. Discussion

In this chapter I have presented a simple artificial life model of the evolution of communication which demonstrates a quite unusual behavior: a sort of limit cycle, present in all the replications of the simulation, in which both average fitness and the quality of the communication system continually oscillate between very low and very high values. In order to understand the mechanisms that underlay this striking result, I used two methods. First, I have isolated a typical cycle of rise and fall of communication happening in one replication of the simulation and I have analyzed the evolutionary dynamics during this period. Second, I have compared the average results of the base-line simulation with other two simulations in which I have disabled communication (no-communication simulation) or I have changed the neural networks which control the behavior of artificial organisms (no-cognitive-pressure simulation). The results of both analyses clearly demonstrate the proposed hypothesis. The selective pressure against good signalling due to the altruistic character of the simulated communication system (a food call) is contrasted by the continuous emergence of spontaneous good signalling in the population due to two factors: (a) the need of categorizing stimuli in adaptively relevant ways and (b) the genetic convergence of the population. In this section, I discuss the relevance of the presented work with respect to previous artificial life literature (3.7.1) and to the biological literature on the evolution of communication (3.7.2).

3.7.1. Artificial life simulations

The work I have just presented clearly demonstrates the soundness of the hypothesis proposed by Cangelosi and Parisi (1998) that the emergence of communication can be favoured by a cognitive pressure towards spontaneous good signalling. Indeed, I have shown that the cognitive pressure due to the need to categorize stimuli in adaptive ways acts only as the individual level. The sharedness of the communication system between different individuals of the same population is due to another factor, namely the genetic convergence of the population itself. But while in the simulations by Cangelosi and Parisi the cognitive and genetic pressures were enough for communication to evolve and remain stable, in my simulations the emerged communication system is continually disrupted by the invasion of the population by cheaters. What is the explanation of this important difference? And, more importantly, is the cognitive pressure towards good communication sufficient for the evolutionary stability of communication?

The cognitive pressure cannot be sufficient for the evolutionary stability of a communication system which gives benefits only to one actor of communicative interactions (in this case the hearer). As demonstrated by my simulations, the reason is that a population in which there is a good communication system of this altruistic kind is always subject to the invasion of cheating individuals which benefit from the received signals but do not reciprocate (or *lie*, thus misleading competing individuals), *unless some other factor does not prevent this to happen.* In the next chapter I will present other simulations showing three of the several possible factors which can prevent the invasion of cheaters in a communicative population: kin selection (Hamilton, 1964), the social learning of the communication system through docility (Simon, 1990) and the use of signals not only for communication but also for individuals purposes, for example as memory aids (see also Chapter 5 and Chapter 6). Other possible factors are sexual selection through the handicap principle (Zahavi, 1975; Welder and Graham, 2001; Grafen, 1990; Bullock, 1998) or reciprocal altruism (Trivers, 1971; Axelrod, 1981). Why, then, does the communication system in Cangelosi and Parisi (1998) seem to be evolutionary stable even though in those simulations none of the factors discussed above were present?

A possible answer lies in the genetic algorithm used by Cangelosi and Parisi, in particular in their peculiar mutation operator. As in the present work, in Cangelosi and Parisi (1998) neural networks' connection weights are codified in the genome of organisms as real values. But in that work weights are initialized in the range [–1; 1] and each weight has a probability of 10% of being mutated by adding to the current value a random number in [–1; 1]. As a result, *weights have no limits.* This peculiar way of implementing the mutation operator might prevent the invasion of cheaters for the following reason. As demonstrated by the analysis of evolutionary dynamics, during the periods in which the communication system quality is low, received signals are ignored. Consequently, communication-output connection weights are under genetic drift in that they do not affect the fitness of any organism: neither that of the speakers, which is never influenced by its own produced signals, nor that of the hearers, since they ignore incoming signals. But a genetic drift without any limits on the connection weights is likely to produce weights of high absolute values. And a mutation which changes a weight's value by adding a random number in the interval [–1; 1] will tend to be behaviorally uninfluential for weights with high absolute values. This is especially true for the signalling behavior of the organisms simulated in Cangelosi and Parisi (1998), since the transmission function of 'linguistic' neurons in that simulation is the step function, whose value can be changed only by changing the sign of the input received by the neurons. The consequence is that once a good communication system has emerged as a result of cognitive and genetic pressures, cheaters cannot emerge because mutations on the communication-output connection weights are completely neutral in that they cannot affect signaling behavior.

On the contrary, in our simulations mutations can change radically the value of connection weights (the mutated value is substituted by a random value always in the same range), so cheaters can emerge when a mutation to one of the communicative-output connection weights changes the signalling behavior of an organism. The same is true for the robotic experiment of Marocco et al. (2003), in which connection weights are coded as strings of 8 bits and mutations act on single bits. In this way, all the bits coding for a single weight can change simultaneously, and, in general, a mutation in one single bit (for example in the bit coding for the weight's sign) can have a radical effect on the value of the weight and, consequently, on communicative behavior. And in fact, in the work of Marocco et al. the cognitive pressure towards good signalling is not sufficient for the emergence of communication.

The mutation operator used in Cangelosi and Parisi (1998) has two fundamental problems, one theoretical and one practical. Theoretically, it seems to be not biologically plausible. In real organisms, mutations can happen in various different ways (see Maki, 2002) and they can always have any kind of effect on the mutating gene (from no effect to the substitution of the codified protein). Hence, there seems to be no biological analogous to the mechanism, present in the simulations of Cangelosi and Parisi, which prevents a gene which has been subject to genetic drift from having non-neutral mutations. Pragmatically, the reason for not adopting the kind of mutation operator used in Cangelosi and Parisi (1998) is that it is too easily subject to falling in local maxima. In general, during a genetic search, some parts of the genome tend to evolve early and other lately. This happens either because some part of the genome code for a trait which is a precondition for the evolution of the trait coded from other parts or simply because of chance. With a mutation system like the one we are discussing, the genetic drift of the genes which are not functional during a given evolutionary phase can prevent the subsequent adaptive evolution of those genes just because they have reached values whose absolute value is too high with respect to the range on which mutations act. Consequently, this kind of mutation operator can compromise the evolutionary process.

From this analysis I can conclude that the cognitive pressure towards good individual signalling and the genetic pressure towards sharedness of signals can constitute a bias towards the emergence of communication. On the other hand, they cannot guarantee, by themselves, the evolutionary stability of the communication system in the cases in which communicative interactions favour only hearers.

3.7.2. Evolution of communication

The fact that communication requires two independent capabilities, speaking (producing useful signals) and hearing (understanding perceived signals), poses two related problems: an adaptive problem and a phylogenetic problem. The adaptive problem is a problem of altruism: why should speakers produce good signals if they gain no benefit in doing so? And, conversely, why should hearers respond adequately to signals if they don't receive any advantage from that? Notwithstanding the emphasis of biological literature on manipulative communication, the situation seems to be completely symmetrical, both theoretically and empirically. And the presence of natural communication systems that seem not to benefit both speakers and hearers needs to be explained either by kin selection or by the handicap principle or by some other means (see next chapter).

But besides the adaptive problem there is also the phylogenetic problem: what evolutionary paths can lead to the emergence of communication systems, since speaking and hearing abilities are complementary and hence adaptively neutral if they are not both present at the same time? The biological literature has provided a partial answer to this question with the concept of a sensory (or producer) bias: if some stimuli trigger in some organism a behavioural response that advantages another organism, the latter can produce signals that resemble those stimuli in order to manipulate the behavior of the former. In this way, the problem of co-evolution is bypassed through the exploitation of behavioral biases in the receivers which are assumed to be present for reasons other than communication.

In this chapter I have proposed the complementary idea of the exploitation of a producer bias. This idea has been neglected so far just because the biological literature has been biased itself by the conception of communication as manipulation. The idea of a producer bias is the following. If organisms have some bias to produce behaviours that systematically co-vary with relevant features of the environment, those behaviors can be used by other organisms as cues for their own purposes. In this way, the biased behaviors become signals, the organisms that exploit them become hearers and the individuals that produce them become speakers.

Just as the exploitation of receiver biases is probably the mechanism by which most manipulative communication systems have evolved, the exploitation of producer biases can have played an analogous role in the historical emergence of most communication systems which benefit (principally) hearers, like alarm and food calls. Furthermore, the simulations presented here suggest also one possible mechanism which can lead to producer biases: namely, the cognitive pressure due to the need for organisms to internally represent stimuli according to adaptively relevant categories. In fact, as my simulations have shown, good categorization of stimula can result in the spontaneous production of good signals which can be exploited by hearers.

But is the hypothesis of a 'cognitive pressure' towards good signalling behavior plausible? Does it hold only for our simulations or is it a possible candidate for explaining the evolutionary emergence of real communication systems? I think that the hypothesis is indeed plausible and that it should be taken in consideration when trying to explain the evolutionary emergence of animal communication systems, in particular those which seem to benefit principally receivers. After all, the only assumption which needs to be made for my hypothesis to work is the link, in organisms' brains, between communicative and non-communicative behaviors. In fact, if such a link exists, then the need for individuals to categorize the environment in adaptive ways will generate a bias towards the production of useful signals. And this assumption seems quite reasonable: what should signals reflect if not what an organism has in mind?

This hypothesis is also empirically testable. One way to test it is the following. Take an animal capable of sending signals. Train it to send a signal in at least two situations which differ from the point of view of the animal's survival (like in presence of edible vs. poisonous items, as in our simulation, or for asking two significantly different kinds of things, like food vs. a partner). The hypothesis on the cognitive pressure on good signalling predicts that the animal will tend to *spontaneously differentiate* between the signals emitted in the two different situations.

It is of the most importance that a first confirmation of this prediction can indeed already be found in recent empirical experiments on Japanese monkeys performed by Hihara et al. (Hihara et al., 2003). In these experiments a monkey was trained to produce a generic call in two different conditions: (a) for receiving food and (b) for receiving a tool with which it could reach a distant food. Surprisingly, without any reward for doing so, the trained monkey started to spontaneously differentiate the sounds emitted in the two conditions. The authors of the experiments interpret their striking finding as the result of a change, in the monkey, from emotional to intentional vocalizations. But the spontaneuos differentiation of monkey's calls can be explained, less speculatively, by the simple fact that in the two conditions the monkey's brain was in two different states: one corresponding to the need for a piece of food and the other corresponding to the need of a tool with which to reach the food. As predicted by my hypothesis, different internal states, be they 'conscious' (as Hihara et al. suggest), 'intentional' or 'emotional', tend to produce different (communicative) behaviors. To put it shortly: what you have in mind will influence what you say. In other words, the signals an organism produces will tend to reflect the way in which the organism categorizes its experience.

4. Adaptive factors for the emergence of communication

4.1. Altruism in the evolution of communication and language

In the previous chapter I have dealt with one of the two fundamental problems that the emergence of communication poses from an evolutionary point of view, namely the phylogenetic problem of the co-evolution between speakers and hearers. But besides the phylogenetic problem the complementarity between hearers and speakers implies also an adaptive problem, a problem of altruism. In particular, for any communication system which does not benefit both the actors of communication, it must be explained how communication could emerge and remain stable notwithstanding its altruistic character. Why should speakers send useful (informative) signals if they do not increase their reproductive chances by doing so? And, conversely, why should hearers react appropriataly to received (manipulative) signals if this does not increase their reproductive success? In fact, the simulations presented in the previous chapter have demonstrated that if communicative interactions benefit only hearers, then a communication system does not emerge under normal conditions, or, when it does, it is always disrupted because of the invasion of the population by cheaters.

The problem of altruism is of the major importance in the literature on the evolution of communication since a great number of animals' communication systems seem to be essentially altruistic, be they manipulative (like several aggressive or mating displays) or informative (like alarm or food calls). This problem is also relevant for the evolution of language (see, for example, Dessalles, 2000; Fitch, 2004) because language seems to be particularly sophisticated for informing conspecifics about relevant features of the environment and, most importantly, human beings seem to be particularly prone to provide each other useful information through language.

By developing the model presented in the previous chapter I will show that a communication system which confers advantages only to hearers but not to speakers can evolve in at least three conditions. The first condition is that the speaker and the hearer share the same or similar genes. In these circumstances communication emerges because the altruistic genes of the speaker are maintained in the population through the advantages conferred by the communicative behavior of speakers to hearers that possess the same (or similar) genes. In other words, altruistic communication can emerge in kin-related groups as a result of kin selection (Hamilton, 1964).

A second condition in which communication emerges even if it is advantageous only to the hearer is a condition in which the communication system is culturally rather than biologically transmitted. If what the organisms inherit genetically is not the communication system itself but only the propensity to learn from others, then this propensity – which, following Herbert Simon (1990), I call 'docility' – can be the basis for learning the communication system from individuals that already know it. Most importantly, this can happen irrespectively of whether in any particular context of use communication confers an advantage to the speaker or to the hearer.

A third condition for the emergence of a communication system which is useful only for the hearer exploits another peculiar characteristic of human language. Unlike animal communication systems, human language is used not only for social communication, i.e., when the speaker and the hearer are two different individuals, but also for talking to oneself, that is, when the speaker and the hearer are the same individual (I will develop this topic in more detail in Chapter 5). This use of language for oneself may have been present in the very early evolutionary stages of language and it may have represented an evolutionary pressure for its emergence. In the last simulation I will show that a useful shared communication system can biologically evolve in groups of genetically unrelated individuals if signals are used as a memory aids, i.e., to allow an individual to keep in memory useful pieces of information which have been communicated by other individuals.

4.2. Relevant literature

4.2.1. The biological view

As discussed in the previous chapter, the recent biological literature on communication has been deeply influenced by the controversial definition of communication by Dawkins and Krebs according to which communication is a form of manipulation. But any form of manipulation implies that there is an individual which gets manipulated, and which, therefore, provides advantages to another individual without receiving any benefit from doing so. As a result, the two main questions that biologists have been asking about the evolution of communication concern adaptive factors: (a) why should hearers let themselves be manipulated by speakers? And, (b) is there anything that can guarantee the reliability of manipulative signals, so that they can be informative and hence useful also for the hearers? Theoretically, the answer to question (a) has been kin selection theory (Hamilton, 1964), while the answer to question (b) has been the handicap principle (Zahavi, 1975, see also Grafen 1990).

According to kin selection theory, an altruistic behaviour can evolve if it is preferentially directed towards kin related individuals. In fact, in this case the individuals that are advantaged by the altruistic behavior will tend to share the same genes of the altruist, including the genes that determine the altruistic behavior itself. This kind of reasoning can be formalized: an altruistic trait can spread in a population as far as

$$c < br \tag{4.1}$$

where c is the cost of the altruistic behavior for the individual who performs it, b is the benefit for the receiver and r is the coefficient of kin-relatedness between the performer and the receiver of the behavior. Kin selection theory can explain, for example, the evolution of begging calls produced by new-borns for receiving extra feeding from parents: as far as the cost for the parent to provide un-necessary food is not greater than the benefit for the new-borns multiplied by the relatedness coefficient, which in sexual species is 0.5 between parents and siblings, the manipulative signalling behavior of the new-borns can spread in the population.

On the other hand, some forms of manipulative signalling behaviors, namely those in which the signaller tries to convince the receiver of its strength (like in mating or aggressive displays), can evolve even among non-kin because their reliability can be guaranteed by the handicap principle. According to the handicap principle, honest signalling can evolve if signals are costly. Typically, in both mating and aggressive displays the message that has to be passed regards the strength and value of the signaller. The function of these signals is to convince the receiver to either mate with the signaller or to flee without fighting. If the signal are costly, in particular more costly for weak individuals than for strong ones (Grafen, 1990), then they will be honest, because only really valuable individuals can afford to pay the costs of signalling.

Both kin selection and the handicap principle have been suggested to have played an important role in the evolution of language. For example, making an explicit reference to the handicap principle, Dessalles has argued that human language evolved as an advertizing display for finding good allies in an a social context in which the capacity to form strong coalitions was of the major importance (Dessalles, 2000). Fitch has criticized explanations of the altruistic uses of language which make reference to sexual selection through the handicap principle on the ground that they tend to make wrong predictions. In fact, the standard outcomes of sexual selection are (a) strong differences between the genders in the expression of the trait and (b) a relationship between the expression of the trait and puberty. In other words, a sexually selected trait tends to be expressed principally (if not exclusively) by males and tends to appear when the male reaches sexual maturity. On the contrary, language is learned by children very soon and much earlier then puberty, and, most importantly, there are no differences in linguistic performance between the sexes (indeed, females seem to be even slightly better than males). Hence, Fitch argues that sexual selection cannot have played a major role in language evolution. Rather, according to Fitch kin selection might have played an important role in that it can explain the evolution of altruistic communication without making the wrong predictions which seem to be implied by the theory of sexual selection through the handicap principle.

4.2.2. Computational modeling

Most of the artificial life literature which has dealt with adaptive problems in the evolution of communication has mainly disccussed one of the two hypotheses proposed by biologists: namely, that altruistic communication can evolve due to kin selection or to the handicap principle. In particular, models by Ackley and Littman (1994) and Oliphant (1996) have shown that a communication system which is advantageous only for hearers does not emerge under 'normal conditions', that is with random assortment between communicating individuals. The reason for this, they argued, lies just in the altruistic character of this kind of communication. On the other hand, if assortment is not random, then communication can emerge. In particular, in both models a good communication system could emerge if the population was spatialized, in the sense that individuals occupied specific places in space and both communicative interactions and reproduction were local. In order to explain their results, both Ackley and Littman and Oliphant appelled to kin-selection. In particular, they argued that in spatialized populations communicating organisms will tend to be kin-related. Hence, under these conditions altruistic genes can emerge because altruistic agents will tend to benefit other altruistic agents.

Di Paolo (1999) has criticized the interpretation given by those authors of their own results. By formally analysing kin relatedness in a model similar to that used by Ackley and Littman and Oliphant, he has shown that kin selection cannot fully explain the effects that spatialization has on the emergence of a communication system. Instead, spatial organization by itself, together with other factors such as discreteness and stochasticity, can play a major role in the emergence of cooperative communication. The role of spatial factors in the emergence of simple, altruistic signalling systems is also discussed in Noble et al. (2001), together with other ecological factors such as the presence of noise. Noble et al. discuss also previous work by Bullok (1998) which demonstrated, through a simple artificial life model, how the handicap principle could lead to the evolution of honest signalling in a manipulative communication system.

As discussed in the previous chapter, Cangelosi and Parisi (1998) were able to evolve an altruistic communication system in a population of ecological neural networks. They suggested that the evolution of good signallers in their model had to be explained as a by-product of the independently evolving ability to categorize the environment. But as demonstrated both by my simulations and by those of Marocco et al. (2003), the cognitive pressure towards good signalling cannot be the whole story. In fact, as I have discussed at lenght in the previous chapter, the cognitive pressure towards good signalling can constitute a produced bias, which might solve the phylogenetic problem of co-evolution of speakers and hearers. But it is not sufficient for solving the adaptive problem. In fact, it cannot prevent the population to be invaded by cheaters which profit of the benefits of communication but do not reciprocate.

4.3. Simulations

If, in the base-line simulation of the previous chapter, the quality of the communication system decreases because of the altruistic character of speaking, then language quality should not decreased if we add to the base-line simulation some adaptive factor that makes it possible for altruistic behaviours to evolve. In this section I present three variations of the same simulation which test three different adaptive factors: kin selection, docility, and talking-to-oneself.

4.3.1. Kin selection

As discussed above, kin selection has already been claimed to play a role in other models of the evolution of communication (Ackley and Littman, 1994; Oliphant, 1996), but this hypothesis has never been tested directly. In fact previous works assumed that kin selection was the at work due to the spatialization of the organisms interactions and reproduction, an assumption which has been criticized by Di Paolo (1999). So, I decided test the power of kin selection directly, through the manipulation of individuals' interactions. In particular, I varied the probability that communicative interactions happen among close kin. The rationale for doing so is the following. Close kin tend to share the same genes. So, the higher the probability to speak to a close kin, the higher the chances that the altruistic behavior of producing good signals is preserved because it tends to confer an advantage to organisms with the same genes and hence with the same altruistic behavior.

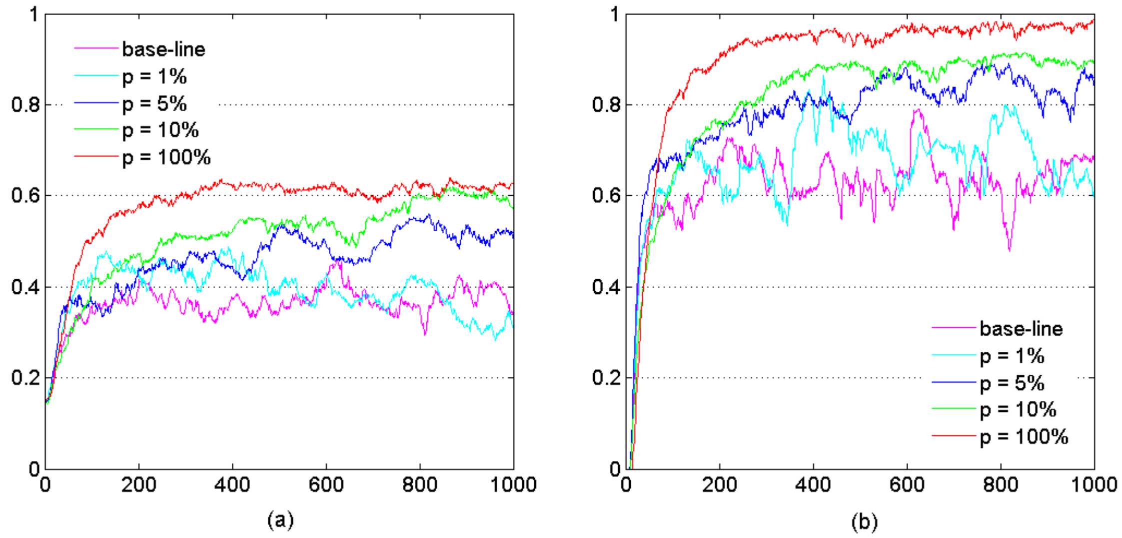


Figure 4.1: Comparison between the base-line and of 4 kin selection simulations with different values of p (1%, 5%, 10% and 100%): (a) communication system quality and (b) average fitness. Average results of 10 replications of the simulations

The kin selection simulation runs as the base-line one, but for the following difference. In this simulation 50 out of the 100 individuals in each generation are males and 50 are females. For each breeding cycle, one male and one female are selected for reproduction with the roulette selection algorithm (see Appendix B) and the pair generates two offspring, one male and one female, with different cross-over points and different mutations on connection weights. This breeding method makes sure that each organism of the population has at least one full sibling. So, I implemented kin selection by varying the probability p with which in any trial the speaker is one of the full siblings of the hearer (with probability $1 - p$ the speaker is not one of the hearers full siblings, but it can be a half sibling). I run several different simulations with different values of p. Figure 4.1 compares the results of the base-line simulation with those of the kin-simulations with $p = 1\%$, 5%, 10% and 100%.

The results of the simulations are very clear: the higher the probability p of speaking to a close kin, the higher both the quality of the communication system and average fitness. Vice-versa, the lower this probability, the lower the communication system quality and the average fitness of the population. With $p = 0.01$, the results are practically the same as those of the base-line simulation. The fact that the results of the simulation with $p = 0.1$ are not very different from those with $p = 1$ should cause no surprise. After all, p determines only the probability that the speaker is a full sibling of the hearer, but even when this is not the case, the speaker can share the same genes of the hearer either because it is a half sibling or simply because of the genetic convergence of the population discussed (and demonstrated) in the previous chapter. In fact, augmenting the kin selection factor has just the effect of adding an additional selective pressure for the stabilization of the communication-output genes: in other words, kin selection simply strengthens a genetic pressure towards the emergence of communication which was already present also in the base-line simulation.

4.3.2. Docility

The preceding simulations have shown that the kind of communication system we are dealing with here is altruistic and, as such, it fails to emerge through genetic evolution unless under kin selection conditions. One of the specific properties of human language with respect to animal communication systems is that human language is culturally, not genetically, transmitted. Could this property have an influence on the emergence of the kind of (altruistic) use of communication we are dealing with here? In 1990 Herbert Simon proposed an original hypothesis on the evolution of altruistic behaviour in humans which was based on the notion of 'docility' (Simon, 1990). Docility refers to the human propensity to socially learn, either through imitation or thanks to explicit teaching, how to behave. This propensity is biologically inherited in our species. Furthermore, it is so strong that we can make the hypothesis that there has been a strong adaptive pressure during hominid evolution for the emergence of such a trait. (Consider the advantages that could derive from an innate predisposition to learn to imitate the behaviour of others in a context of highly social primates that are evolving the capacity of tool making and tool use, for instance).

But for an individual to determine the contribution to its fitness of each behaviour which the individual is learning from others would be extremely hard, if not completely impossible. As a consequence, a docile organism will tend to learn whichever behaviour it will be taught, be it egoistic or altruistic, provided that the overall advantages of learning from others are higher than the disadvantages. According to Simon's docility theory, in a population of docile individuals an altruistic behaviour can evolve if the following three conditions are satisfied: 1) there is some advantage d in being disposed to learn from others, i.e., in being docile; 2) organisms are not able to evaluate the contribution of each particular behavior that they learn from others to their own fitness; 3) the advantage d of being docile is greater than the cost c of the altruistic behaviour.

As the use of communication that we are dealing with here is altruistic and as human language is culturally transmitted (although on a genetic species-specific basis), it is interesting to test whether Simon's explanation of altruism seems to be applicable to language evolution. In order to do so, I run a new simulation, the docility simulation, in which the connection weights of all individuals are always random at birth and are not inherited from parents. Instead, the genome of these organisms is constituted by one only gene, encoded as an integer number, which specifies an individual's docility, implemented as the number of learning trials for that particular individual. In the first generation each individual is assigned a random value in the interval [0; 200] for this gene and this value is genetically transmitted with a 2% probability of being changed by adding or subtracting a random number in the range [–100; 100]. In any case, docility is forced to stay in the interval [0; 500].

The life of organisms in this simulation is divided into two periods: infancy and adulthood. During infancy, the organism is supposed to follow its parent and learn from its parent how to behave in different situations. Its inherited docility gene determines the number of back-propagation learning cycles to which the infant exposes itself. Learning is imitative in that the teaching input of the back-propagation algorithm is the output of the infant's parent (Hutchins and Hazlehurst, 1995; Denaro and Parisi, 1997). Since there are three kinds of situations which organisms are exposed to during their life, there are three different learning conditions: (1) comprehension learning, (2) decision learning, and (3) naming learning. Comprehension learning takes place when the organism is distant from a mushroom and has to decide whether to move or not to move according only to the signal it receives from another organism. Decision learning takes place when the organism is near the mushroom and its decision whether to move or not to move into the mushroom cell and eat the mushroom depends on both the visual input from the mushroom and the linguistic input. Finally, naming learning takes place when the organism acts as a speaker: the organism receives the perceptual properties of a mushroom as input and it has to produce a signal. In short, the number of learning cycles for each organism is determined by the organisms inherited docility and for each learning cycle this is what happens:

1. one of the three learning situations is randomly chosen together with one of the 420 mushrooms;
2. the appropriate input is given both to the learner and to its parent;
3. both the organisms output and its parents output are calculated;
4. the output of the parent is given to the child as the teaching input with some added random noise (a random value chosen in the interval [–0.25; 0.25] is added to the teaching input);
5. finally, the child's connection weights are changed according to the back-propagation algorithm (with a learning rate of 0.3 and a momentum of 0.8).

At the end of infancy an individual starts its adult life, which is identical to that of the genetic (base-line) simulation.

Figure 4.2 shows language quality, average fitness, and average value of the docility gene in this simulation. Since in this simulation organisms are born with random connection weights, they are bound to behave randomly unless they learn from their

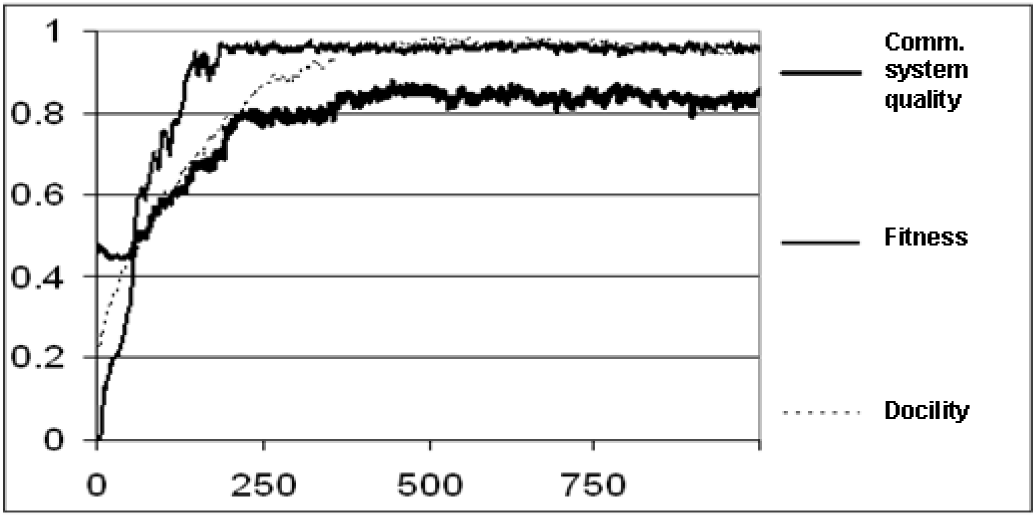


Figure 4.2: Communication system quality, average fitness and average value of the docility gene of the docility simulation. Average results of 10 replications of the simulation

parents. The fact that the teachers of any given generation are the selected organisms of the previous generation guarantees that what is learnt is a good behavior. On the other hand, the fact that some noise is always present in the cultural transmission of behavior guarantees that there is some added variability which is necessary for cultural evolution to take place.

The results show that the average value of the docility gene rises very quickly until it reaches almost its maximum value (figure 4.2, dotted line). Since organisms with a good foraging behavior tend to be also good speakers because of the cognitive pressure towards language emergence, young organisms will learn not only to discriminate edible from poisonous mushrooms, but also to produce good signals. Most importantly, because of the fact that the behavior of these organisms is culturally learnt, the invasion of the population by cheaters is prevented. Only individuals which are not very docile can produce bad signals, but those individuals cannot learn how to behave efficiently and therefore they will tend have a lower fitness than more docile (and altruistic) individual. As a result, to the increase in the average value of the docility gene corresponds a parallel increase in the quality of the communication system produced by the organisms, which reaches after about 200 generations the very high value of about 0.8 (figure 4.2, thick line). As it turns out, the correlation between docility and language quality is very high: 0,967. Finally, since the organisms of this simulation can exploit all the advantages provided by a good communication system, their average fitness reaches almost the maximum possible value (figure 4.2, thin line).

4.3.3. Talking to oneself

In all the simulations described so far we have assumed that signals are used for social communication purposes, that is, in situations in which the emitter and the receiver of signals are two different individuals. But for human beings language can have useful functions even in situations in which the emitter and the receiver of a signal are the same individual, i.e., when an individual talks to herself. The particular function that I will consider here is language as an aid for memory. The initial situation is identical to that of the preceding simulations. An individual is placed in the start cell of the corridor, another randomly chosen individual is placed near the mushroom and it generates a signal which is heard by the first individual. On the basis of this signal the first individual can decide whether to approach the mushroom or refrain from doing so. However, since it takes 10 cycles to reach the mushroom, the signal must be available to the first individual in all these cycles in order for the individual to know what to do in each successive cycle. In the preceding simulations this problem was solved by assuming that the second individual, the speaker, continued to emit the signal until the end of the trial. In the present simulation the situation is different. The emitter of the signal emits the signal only in the first cycle and then it goes away. All the first individual can do in this situation is to try to remember the signal by repeating the signal to itself until the mushroom is reached. In the first cycle, when the signal arrives from the conspecific, the individual responds to the signal not only by either moving one cell forward toward the mushroom or avoiding doing so, but also by producing a signal using its own linguistic output units. In the next cycles, the individual hears this self-produced signal and responds to it.

Will a a good communication system evolve in these conditions? Notice that in the base-line simulation the communication system was very unstable because there were both cognitive and genetic pressure for its emergence and a strong selective pressure against good communication due to its altruistic character. In the present simulation, in which individuals talk to themselves, they have an interest in producing good signals because sometime they are the receivers of their own produced signals. Therefore emitters that produce good signals will tend to have more chances to reproduce than emitters of bad signals because the emitters of good signals can remember correctly the information received about the quality of the mushroom present at the end of the corridor. This effect is independent from kin-relatedness (it might be interpreted as kin-relatedness in a single individual). Hence, talking to oneself may constitute a selective pressure for the emergence of good communication even in populations in which language is exchanged between pairs of non-kin-related individuals. This prediction is confirmed by the results of the talking-to-oneself simulation. Figure 4.3 compares average fitness and communication system quality in the baseline and the

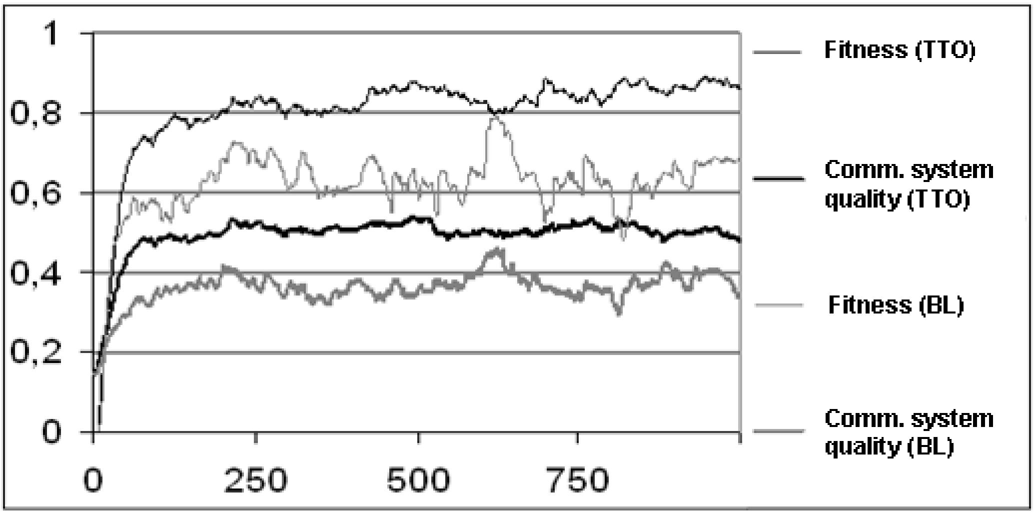


Figure 4.3: Comparison between the base-line (BL) and the talking-to-oneself (TTO) simulations with respect to average fitness and quality of the communication system. Average results of 10 replications of both simulations

talking-to-oneself simulations. The communication system quality in the talking-to-oneself simulation is more stable and significantly higher than that of the baseline simulation: the range of fluctuation of the quality of the communication system is in fact [0.45; 0.55] and [0.25; 0.45], respectively. As a result, the average fitness of the population is higher and more stable in the talking-to-oneself condition than in the baseline condition: fluctuation range is [0.8; 0.95] and [0.5; 0.8], respectively.

4.4. Discussion

For communication to emerge biologically it must involve some selective advantage for both the emitters and the receivers of signals. Communication requires both good speakers, i.e., individuals that emit the appropriate signals in the appropriate circumstances, and good hearers, i.e., individuals that respond appropriately to the signals that are produced by the speakers. But good speakers emerge only if it is advantageous for them to produce the appropriate signals in the appropriate circumstances, and good hearers emerge only if it is advantageous for them to respond appropriately to these signals.

If signals provide the hearer with useful information, then the production of signals by the speaker is an altruistic behaviour provided that the behaviour with which the hearer responds to the signals have no useful consequences for the speaker and do not increase the speaker's reproductive chances. In fact, our simulations show that if communication benefits the hearer but not the speaker it fails to emerge. Individuals that happen to be good speakers may increase the reproductive chances of hearers which benefit from the linguistic signals produced by the speakers but which, when it is their turn to function as speakers, may turn out not be good speakers. In this manner, although it would be generally useful, a stable shared communication does not evolve.

Indeed, the results of our base-line simulation support this analysis and replicate the results of other simulations of the emergence of altruistic communication systems (Ackley and Littman, 1994; Oliphant, 1996; Marocco et al., 2003). The simulations analyses of the previous chapters have shown that the structural coupling between categorization abilities and communicative behavior can indeed favour the emergence of communication, as suggested by Cangelosi and Parisi (1998), but cannot account for the stability of the communication system. In fact, an altruistic communication system is bound to be disrupted by the invasion of cheaters which benefit from received signals but do not reciprocate, if no other mechanism prevents this to happen. In the simulations presented in this chapter I have tested three possible mechanisms.

The first mechanism is kin selection. The kin selection simulations show that, in accordance with kin selection theory, the probability that the speaker and the hearer share the same genes directly affects the stability of a communication system which benefits the hearer but not the speaker. Animal signals mostly communicate information about the sender of the signal – its location, species, identity, emotional state, intentions and attitudes –, not information about the external environment. Clearly, there are exceptions, like food and alarm calls, but human language clearly is the most sophisticated communication system for communicating information about the external environment. While kin selection is likely one of the major factors in the evolution of social insects communication systems (like the famous honey-bees dance: see Frisch, 1967), it is not clear what could have been the role of kin selection in the evolution of primate (including human) communication. Surely, hominid evolution took place in small, kin-related groups, and the long juvenile period of humans may have contributed to a substantial increase in the probability that communicative interactions took place between kin related individuals (see Fitch, 2004 for similar lines of reasoning). As our kin selection simulations demonstrate, if this is the case, then the increase in genetic relatedness between the speaker and the hearer might have been one factor, among many others, which favoured the evolution of (the altruistic use of) language.

However, if language is restricted to kin-related groups its usefulness appears to be limited. Language becomes much more useful if it represents a vehicle of communication and interaction among larger groups of genetically unrelated individuals. Language that benefits the hearer but not the speaker may emerge in larger groups of unrelated individuals if it is culturally rather than biologically transmitted. Cultural transmission is learning from others. Human beings appear to have a genetically inherited tendency to learn from others whatever behavior others may care to teach them. This docility evolves because it confers an advantage to the individual that possesses it: a docile individual can learn from others behaviours which would be more difficult to learn by directly interacting with the non-social world. These behaviours generally tend to increase the reproductive chances of the learning individual but in some circumstances and in some of their uses they may benefit others. But for docile organisms, learning from others tends to be blind, in the sense that the learning individual cannot easily distinguish between what is beneficial for himself or herself and what is beneficial for others, especially because this kind of learning takes place mostly when the individual is young. Thus, altruistic behaviours, including the altruistic use of language we are dealing with here, can emerge as a cultural by-product of the genetically evolving docility of a population. Studying the influence of cultural learning on language emergence is one of the major themes in the literature on computational models of language evolution (see, for example, Hutchins and Hazlehurst, 1995; Batali, 1998; Hurford, 1999; Kirby, 2000; Steels and Kaplan, 2002). But none of the previous computational models of language evolution tested the hypothesis that the evolution of cultural transmission could have favoured the emergence of an altruistic use of language. The docility simulation shows exactly this. It shows that linguistic docility, that is, a tendency to learn language from others, can emerge biologically and be biologically transmitted. If this is the case, the homogenization of linguistic behaviour induced by cultural transmission can favour the (cultural) emergence of an altruistic use of language.

Finally, a third mechanism may explain the emergence of a language that benefits the hearer and not the speaker in groups of genetically unrelated individuals even if the language is biologically and not culturally transmitted and evolved. This third mechanism is using the language not only to communicate with others but also to communicate with oneself. Generally, there is a tendency to think that language was used by humans to communicate with oneself only when language was already well developed and was sophisticated and complex; hence, quite recently compared with the first appearance of a proto-language. However, this is not necessarily the case. Even a very simple proto-language, for example, a language made up of single words (or holophrases), may be used to talk to oneself, for example as an aid for memory, with advantages for the individual that uses the language in this way. Based on this hypothesis, my talking-to-oneself simulation shows that signals that benefit the hearer but not the speaker can emerge biologically among genetically unrelated individuals if the hearer has to repeat the signals to himself or herself in order to keep them in memory. When one speaks to another individual and the signals benefit the hearer but not the speaker, the speaker acts altruistically and may benefit a hearer who is a poor speaker. This, as we have seen, is an obstacle for the biological emergence of language. But if the hearer has to repeat the signals to himself or herself in order to keep them in memory, then there is a positive selective pressure towards good speaking abilities. In fact, in talking-to-oneself the same individual is both speaker and hearer. Hence, the advantage of understanding linguistic signals can only be exploited if it is accompanied with good productive capacities.

Using language as an aid to memory can be advantageous for at least two reasons: (a) delegating the memory function to the linguistic system can leave the sensory-motor system free to process other information useful for acting in the environment while linguistically remembering previous information, and (b) linguistic signals may occupy less space in memory than the raw information they refer to. It might also be the case that another peculiar characteristics of human language, namely the fact that it uses displaced signals (Hockett, 1960), has become possible only after hominids had improved their memory by talking to themselves (for the possible role of linguistic memory in the evolutionary origin of language, see Aboitiz and Garcia, 1997). Furthermore, as I will discuss in the next chapters, the use of the linguistic system as an aid to memory is only an example of the many possible individual uses of the social communication system (Vygotsky, 1978; Jackendoff, 1996; Clark, 1998). In general, the use of language as an aid to cognition might have played an important role in language evolution. The talking-to-oneself simulation is just a very simple demonstration that this could be the case. In the next chapter, I will discuss the role of language in human cognition in more detail, and I will start studying this role by providing a computational model of the effects of language on categorization.

5. Language as an aid to categorization

> The natural laws of perception most clearly observed in the receptive processes of the higher animals undergo basic changes due to the inclusion of speech in human perception, and human perception thus acquires an entirely new character.
> Vygotsky and Luria: Tool and symbol in child development, p. 126 (Vygotsky and Luria, 1994)

5.1. Language and cognition

5.1.1. The classical view

What is the role of language in human cognition? This is one of the most important questions we have to answer if we want to understand the human mind in general. The standard view of classical cognitive sciences on this topic can be reassumed in two statements: (a) cognition is, generally speaking, 'linguistic' in itself, in that it consists in the manipulation of linguage-like structures (propositions) according to formal rules; (b) the essential function of natural language is just to express the contents of thought, hence natural language does not affect cognition in any substantial way.

The view of cognition as symbol manipulation is at the very heart of classical cognitive science, constituting the common assumption on which at least three of the subdisciplines that gave birth to cognitive science were based: artificial intelligence (i.e. the symbol system hypothesis: Newell and Simon, 1976), cognitive psychology (i.e. the language of thought hypothesis: Fodor, 1975) and cognitive-science-related philosophy of mind (i.e. the computational theory of mind: Putnam, 1963). If one considers cognition as fundamentally linguistic, there is no reason for considering language anything more than a very complex and powerfull communication system. And in fact, this is something that has been seldom if ever questioned within classical cognitive science.

5.1.2. Challanges to the classical view

In the last twenty years the basic assumption of classical cognitive science has been questioned from several perspectives in such ways that now we are substantially rethinking our view of cognition. A number of philosophical arguments have been put forward against the view of cognition as symbol manipulation, some of which I found compelling (see, for example, Dennett, 1978; Churchland, 1981), while others just misleading (see, for example, Searle, 1980; Penrose, 1989). But untill concrete alternative views had not been proposed, advocators of the view of cognition as symbol manipulation could still claim that their hypothesis was "the only game in town" (Fodor, 1975).

The publication, in 1986, of the two volume *Parallel Distributed Processing* by Rumehlart, McClelland and the PDP group (Rumelhart et al., 1986) is to be considered a milestone in cognitive science history. In fact, it provided for the first time a concrete and rather detailed account of cognition which was proposed as an alternative to the symbol manipulation paradigm. According to this alternative view cognition is not the manipulation of symbols according to formal rules, but rather the parallel and distribute processing of sub-symbolic information: that is the transformation of purely quantitative values (the vectors of the activations of groups of units) through other quantitative values (the connection weights linking the groups of units).[1]

The connectionist framework, with respect to the symbol-manipulation one, had at least two advantages. First, it was more biologically plausible, being directly inspired by the knowledge of our own control system, that is the brain, in contrast to being inspired by the computer, which is not the product of natural selection but an human-made artifact. Second, connectionist networks have several properties which are similar to those demonstrated by natural cognitive systems (humans) but which are not possessed by classical symbolic systems. For example, in contrast to symbolic systems, neural networks are robust both to damage and to noisy input: if you present a network noisy input the performance of the network degrades gracefully, and the same is true if you destroy (lesion) some of the units or the connections of the network. In contrast, a catastrophic failure of the functioning of symbolic systems is the typical result if you damage some part of the system or provide noise. Furthermore, neural

---

[1] For a very brief introduction to neural networks, see Appendix A. For more general introductions, the best place to start is probably still the PDP book, which is generally considered as the 'bible' of connectionism. A good technical introduction is Floreano and Mattiussi (2002), while for discussions on the theoretical relevance of connectionsism for cognitive science see, for example, Smolensky (1988); Clark (1989); Bechtel and Abrahamsen (1991).

networks, as natural cognitive systems and unlike symbolic ones, typically generalize quite well to novel, unexperienced stimuli. Neural networks are also particularly well adapted to problems that require the resolution of a number of (possible conflicting) constrains in parallel. And most if not all of natural cognitive tasks such as object recognition, planning, and sensory-motor coordination seem to be problems of this kind. Furthermore, the kind of distributed representations neural networks have seems to be much more realistic than the classical symbolic representations. It is in fact quite clear that our categories are not completely clear-cut as those of symbolic systems; rather, categories seem to be delimited in somewhat vague and flexible ways (consider the notion of family resemblance (Wittgenstein, 1953) or the prototype theory of concepts (Rosch, 1973)). As I explain in Appendix A, connectionist models are especially well suited to accommodating graded notions of category membership of this kind. Furthermore, since representations are coded as patterns of activations of vectors of neural units, relationships between representations are coded in the similarities and differences between these patterns. This results in another fundamental property. Since the internal properties of a representation in a neural network carry information on what the representation is about, a neural networks memory retrieval can be (and in fact is) content-based. In contrast, local, symbolic representations are completely conventional and hence they are not content-addressable.

Another fundamental attack to the classical framework of cognition as symbol manipulation came in the early 1990s by the field known as behavior-based robotics, particularly from the work of Rodney Brooks (Brooks, 1990, 1991). Brooks' arguments against symbolic processing approaches to artificial intelligence derived from a kind of biological considerations slightly different from the ones which prompted connectionist research. While classical artificial intelligence focused on high-level cognitive capacities like playing chess, rational problem-solving or logical reasoning, natural intelligence has evolved for solving more basic, low-level tasks, like real-world perception, sensory-motor coordination or navigation. And these are clearly the kind of tasks we are particularly good at, in contrast with high-level cognitive capacities which are both philogenetically and ontogenetically older, and quite difficult to acquire during development. But the symbolic systems constructed for solving high-level problems resulted to be completely un-adapted for dealing with more basic cognitive capacities, and for fundamental, rather than accidental reasons. For example, they suffered what has been called the frame problem, which can be stated as follows (for a more accurate discussion, see Dennett, 1987). Since the number of possible logical inferences that can be made at any time when dealing with the real world are infinite, trying to accomplish real-world task through symbolic systems requires distinguishing what are the inferences that are relevant, and therefore must be made, from those which are irrelevant, and therefore must be ignored. This has proven to be an un-solvable problem for traditional symbolic artificial intelligence. In order to decide which are the inferences that must be made, a symbolic system must first decided which are the relevant things which must be taken into account for taking that decision, and so on and so forth. In other words, faced with real world problems, a symbolic system gets easily stuck because there is no clear way for avoiding the evaluation of an infinite number of true but completely irrelevant statements. The problem is that symbolic systems are particularly good at truth-preserving, logical reasoning, but very bad at practical reasoning, which is exactly what is needed when dealing with a complex and constantly changing environment.

In order to construct robots able to interact effectively with the physical world, Brooks argued, we need to get rid of the symbolic paradigm, which is based on complex reasoning on complex representations. On the contrary, we must make the best use of simple control mechanisms which exploit the continuous interactions between a robot and its environment, because this is exactly what real organisms do all the time. In other words, organisms do not have to work on complex representation for solving their real-world problems just because their interactions with the environment are constant and on-line. As a consequence, they can "use the environment as its own best model" (Brooks, 1991).

Hence, if connectionism pointed to the sub-symbolic character of cognitive processes, the new robotics pointed to the fact that cognitive processes are always 'embodied', 'situated' and (partially) 'distributed' in an organism's environment. They are embodied in that the body and its physical properties are determinant for the way a given task is solved. They are situated because the constrains provided by the environment can act also as opportunities for the task's solution. And they are partially distributed because they do not happen only inside an organism's head, but they substantially depend on the organisms environment, which, especially in the human case, includes also artefacts and other agents.

Finally, another challenge to the classical, symbolic approach to cognition came from dynamical systems theory. Proponents of the dynamical hypothesis argue that cognition should not be considered in computational terms, but rather through differential equations, and using dynamical systems concepts like equilibrium points, cyclic behaviour, attractors, bifurcations and so on. In other words, cognition must be understood by interpreting a cognitive system as a point moving in a multi-dimensional space, and by sorting out the kind of trajectory that the system takes and which are the laws that govern that trajectory (see Smith and Thelen, 1993; Port and van Gelder, 1995; van Gelder, 1998; Beer, 2000).

The concepts and tools of connectionism, robotics and dynamical system theory, together with original studies in developmental psychology (i.e. Thelen et al., 2001), neuroscience (i.e. Churchland et al., 1994; Kelso, 1995) and anthropological and sociocultural studies (i.e., Hutchins, 1995), opened up several lively areas of research, especially of the synthetic kind: artificial life, the simulation of adaptive behaviour, evolutionary and developmental robotics and so on. The overall result is that we are now rethinking substantially our view of cognition. In particular, the fundamental assumption of classical cognitive science that cognition is the manipulation of symbols according to formal rules is being replaced by a view according to which the mechanisms that explain cognitive behaviours are non-symbolic or sub-symbolic, and cognition consists in the adaptation of the agent to its environment, which in turn depends heavily on the dynamics of the interactions between the agent, including its body, and the environment it lives in, including artefacts and other agents (Bechtel et al., 1998; Clark, 2001).

5.1.3. Language as a cognitive tool

Apart from the works of connectionsists, which address high-level classical topics but without taking into account 'embodiment' and 'situatedness', the new cognitive science has been so far concerned mostly – if not *exclusively* – on low level behavior, such as perception, learning, sensory-motor coordination, navigation, and so on. The question remains whether the same broad framework can scale up to explain also the higher forms of cognition demonstrated by humans such as problem solving, complex reasoning, and planning, or if in order to explain characteristic human cognition we must go back to the symbol manipulation paradigm. From the point of view of the new cognitive science, the most promising way of addressing the question, I argue, is to consider language not only as a communication system, but also as a kind of cognitive tool. In fact, as I will argue, the learning and use of language substantially transforms the adaptive pattern of human beings, and is at the roots of the development of higher cognitive functions.

The view of language as something that transforms all human psychological processes dates back as early as the 1930s, with the work of Russian scholar Lev Vygotsky (Vygotsky, 1962, 1978). According to Vygotsky, the most important moment in child development is that in which the child starts to use language not only as a social communication systems, but also as a tool for controlling her own actions and congnitive processes. The idea is the following. When the child is challanged by a task which she cannot solve but through the help of an adult or a more skilled peer, she asks for help, which typically takes the form of linguistic help. Later on, when the child is facing the same or a similar task all alone, she can reharse the social linguistic aid which helped her succeeding in the problem. This is what is called 'private speech', which, according to Vygotsky, plays a fundamental role in the development of all human psychological processes.

The linguistic social aid coming from adults takes several different forms. Social language helps a child in learning how to categorize experiences, in focusing her attention to important aspects of the environment, in remebering useful information, in inhibiting un-useful, spontaneous behavior, in dividing problems in easier sub-problems and constructing plans for solving complex tasks, and so on and so forth. When the child is talking to herself she is just making to herself what others used to do to her, that is, providing all sorts of cognitive aids through linguistic utterances. Once the child has mastered this linguistic self-aid, private speech tends to disappear, but only apparently. In fact, it is just abbreviated and *internalized*, becoming inner speech. Hence, most of adult human cognitive processes are linguistically mediated, in that they depend on the use of language for oneself.

This Vygotskyan view of language as a cognitive tool has been largely ignored in mainstream cognitive science. The principal reason is perhaps that the Vygotskyian works began to be translated in english only in the 1960s, when developmental psychology was already dominated by the more individualistic theory of Jean Piaget, which denied any substantial role to socio-cultural factors in cognitive develppment. Recently, the idea of language as a cognitive tool has beeing raising increasing interest in the cognitive-science-oriented philosophy of mind (Carruthers and Boucher, 1998). For example, Daniel Dennett (Dennett, 1991, 1993, 1995) has argued that the human mind, included its most striking and difficult to explaing property, consciousness, depends principally not on innate cognitive abilities, but on the way human plastic brains are 're-programmed' by the effects of cultural input coming through language. Andy Clark (Clark, 1997, 1998, 2005) developed further those Dennettian ideas by providing several arguments about how animal-like, embodied, situated and sub-symbolic cognitive processes can be augmented by the learning and use of linguistic signs. According to Clark, language is not only a communication system, but also a kind of "external artifact whose current adaptive value is partially constituted by its role in re-shaping the kinds of computational space that our biological brains must negotiate in order to solve certain types of problems, or to carry out certain complex problems." (Clark, 1998, pag. 163)

Apart from the philosophical ideas of Dennett and Clark, the Vygotskyan view of language as a cognitive tool has recently been raising increasing interest in empirical cognitive sciences (see, for example, Diaz and Berk, 1992; Gentner and Goldin-Meadow, 2003a). In the next section I will review some of the empirical evidence that has recently been gained in favour of the Vygotskyan theory, and the very few computational models related to the topic. In the rest of the chapter, I will present an original model with which I start to study the vygotskyian idea of inner speech as the internalization of linguistic social aid, in particular with respect to one of the most basic cognitive function, namely, categorization.

5.2. State of the art

Before reviewing the empirical eviedince in support of the view of language as a cognitive tool, a caveat must be made. In fact, this view is sometimes confounded with a related but different view. This is the so called 'Sapir-Whorf' hypothesis, according to which the particular language one speaks affects the way one thinks.[2] This Whorfian hypothesis is compatible with the position I am endorsing but is more limited. Indeed, the influence of a particular natural language on the way one conceptualizes her world is a possible (but not necessary) consequence of the Vygotskyian hypothesis that the possession and use of language in general, as a culturally acquired symbol system, deeply transforms human cognitive functions. Even though there is indeed a considerable amount of empirical literature on the whorfian hypothesis, I will not discuss it here. The reason is twofold: first, the debate on this issue is still open, with some empirical evidence indicating differences in cognitive performance realted to linguistic diversity (see, for example, Hunt and Agnoli, 1991; Levinson et al., 2002; Majid et al., 2004), while other denying this effect (see, for Pinker, 1994; Li and Gleitman, 2002; for general reviews on the the whorfian hypothesis, see Bowerman and Levinson, 2001 and Gleitman and Papafragou, 2005). Second, as I have said above the claim that language in general (whatever the particular natural language one has) improves cognition can and must be separated by the Sapir-Whorf hypothesis.[3]

5.2.1. Empirical evidence

What about the empirical evidence on the view of language as a cognitive tool? Notwithstanding the relative dis-interest for this idea in mainstream cognitive sciences, the idea has not completely disappeared and several different and un-related lines of research are providing more and more empirical support for it.

The first line of evidence comes from developmental studies on private speech (for a review, see Diaz and Berk, 1992 and Berk, 2001). First of all, it has been shown that private speech does indeed represents a quantitatively very important phenomenon. Depending on the situation, it accounts for 20 to 60% of linguistic production of children from 3 to 10 years of age (Berk, 1994). Also the process of internalization – that is, the path from private to inner speech – has gained empirical support. For example, it has been shown that the decline of audible task-relevant private speech corresponds to the increase in the visible signs of inner speech, like un-understandable muttering

---

[2] For recent discussions of the whorfian hypothesis see the recent book edited by Dendre Gentner and Susan Goldwin-Meadow (Gentner and Goldin-Meadow, 2003a), especially the editors' introduction (Gentner and Goldin-Meadow, 2003b).

[3] Indeed, if I had to, I would argue in favour of the whorfian stance, for both theoretical and empirical reasons (I find the evidence for a particular language influencing cognitive processes more compelling than that against the hypothesis), but I will not discuss the matter any longer here.

and lip and tongue movements (Frauenglass and Diaz, 1985; Berk and Landau, 1993; Duncan and Pratt, 1997). Furthermore, the Vygotskyian idea that private speech has a social origin has also been supported by several investigations: a good correlation between children's social and private remarks has in fact been found (Berk and Garvin, 1984); private speech seems to be favoured by rich social contexts (Ramirez, 1992), and its frequency decreses when children are isolated (Berk and Landau, 1993). Most importantly, there is growing empirical evidence confirming the Vygotskyan hypothesis that private speech plays a fundamental role in the development of children's cognitive abilities. It has been shown that the amount of private speech a child is engaged in when trying to solve a challanging task can predict future performance on that kind of task or, more generally, future gains in problme-solving competence (Bivens and Berk, 1990; Azmitia, 1992; Behrend et al., 1992). Finally, there is evidence that private and inner speech play an important role in the development of self-regulation: a good correlation has been found between the ability to focus one's attention to a given task and (a) the amount of task-relevant utterances (Berk and Landau, 1993), and (b) the speed with which those audible utterances are transformed in silent ones, that is, in inner speech (Bivens and Berk, 1990).

Another important line of research has repetedly and consistently demonstrated that language can facilitate category learning. Several studies with subjects of different ages (from 9 months old children to adults) have in fact demonstrated that providing linguistic input to somebody who is learning to categorize objects can substantially ease and speed up the learning process (see, for example, Waxman and Markow, 1995; Nazzi and Gopnik, 2001; Lupyan, 2006). These findings strongly suggest that labels have the function of 'inviting' category formation by guiding our attention onto 'meaningful' aspects of our environment and by providing important cues about how to categorize them.

Language clearly plays a major role in our highly abstract, logico-mathematical abilities. This has been demonstrated by several studies, especially by the work of Dehaene (Dehaene, 1997). A vast amount of empirical evidence (of various kinds) supports the idea that our mathematical skills are based on two functionally and anathomiocally distinct systems: (a) an innate 'number sense', shared with other animals and pre-linguistic infants, which allw us to appriciate changes in quantity, relative quantities, and a very few absolute quantities (one, two and three, probably); (b) a culturally acquired system which depends on language, and allow us to consider exact quantities other than 1, 2 and 3, and consequently to develop all other complex mathematical skills. Here is some of the empirical evidence supporting this view. A first line of evidence comes from a study of bilinguals which are asked to make either exact or approximate sums of two-digits numbers. In the approximate condition (select which of two candidates is closer to the result of the given sum), switching the language between the training and the testing session did not affect performance, while in the exact condition (select which of two close candidates is equal to the given sum) it did. In particular, subjects responded slower if the language in which they were tested was different from the one in which they were trained compared to the condition in which the test language was the same as the training one (Dehaene et al., 1999). Another line of evidence comes from comparining the memory capacities for numers of speakers of different languages. In particular, Chinese speakers, whose names for number are very short, can memorize perfectly well 7 single digits, while English speakers have a 50% chance of failure in the same task (Dehaene, 1997). Another line of evidence comes from neuropsychological studies: for example, patients with lesions to their left-hemisphere are not able to do even very simple exact calculations (like 2 + 2) while they can do approximate ones (for example, if asked which between 3 and 9 is closer to 2 + 2 they would reliably choose the former). Finally, neuroimaging studies seem to provide other support to the same idea (Dehaene et al., 1999): exact arithmetical tasks activate language related areas of the left frontal lobe, while approximate tasks activate bilateral ares of the parietal lobes related to visuo-spatial cognition.

Another important way in which language seems to improve our cognitive abilities is by providing the resources for combining different, otherwise encapsulated, kinds of information in order to solve a given task. This seems to be the best explanation of a set of experiments on spatial reasoning. Rats and up to 2 years old children have proven to have interesting similar limitations in their spatial reasoning. If they are shown the location of food (or of an interesting object) in a rectangular room and then they are disoriented, they are able to re-orient themselves only with respect to the shape of the room, but they do not take into account the colour of the walls, notwithstanding the fact that both rats and young children are able, in other tasks, to respond selectively to colour information (Cheng, 1986; Hermer and Spelke, 1994). On the other hand, human adults tested under similar conditions are able to integrate spatial and colour information quite easily in order to solve the task (Hermer and Spelke, 1994). Interestingly, developmental studies have shown that the acquisition of the capacity to perform correctly in this task is closely related to the acquisition of spatial terms: children who are able to produce utterances conjoining spatial and color terms (i.e. who can describe something as, say, 'to the right of the long green wall') are also able to solve the problem. While those who do not display this kind of word use do not outperform rats (Hermer-Vazquez et al., 2001). Most importantly, adult's good performance has proven to be due to language use (for oneself). In another set of experiments, adults have been asked to solve the same task while performing one of two other tasks. The first one involved the repetition of speech played over headphones, while the second one involved the repetition, with the hands, of a rhythm played over headphones. While rhythm shadowing caused only a general degradation in performance but did not prevent subjects to integrate both geometrical and nongeometrical information, speech shadowing caused adults' performance to degradate at the level of young children and rats (Hermer-Vazquez et al., 1999). Taken all together, these and other[4] experiments clearly suggest that using language for oneself is what enables humans to put together different kinds of otherwise encapsulated information for solving complex tasks.

Another fundamental role that language has in the development of human cognition is related to working memory. Both psychological and neuroscientific evidence demonstrates that humans have at least two distinct working memory systems. the first one is a multi-modal system which we share with non-human primates. The second one is linguistic, and hence species-specific (see, for example, Baddeley, 1992; Petrides et al., 1993; Becker and Morris, 1999). Recent neuro-imaging studies have also been discovering the neural basis of these memory systems, suggesting that the linguistic memory system is subserved by moslty left-hemispheric areas which underlay normal (audible) speech (see, for example, Gruber, 2002; for a detailed review see Gruber and Goschke, 2004). Furthermore, verbal memory seems to be more efficient and flexible than the older multi-modal system, and therefore it seems to function as the predominant reharsal mechanism. For example, a number of studies using different experimental paradigms have consistently shown that articulatory suppression significantly increases the difficulty of a task by making the retrieval of the taks goal difficult (see, for example, Baddeley et al., 2001; Emerson and Miyake, 2003; Miyake et al., 2004). This suggests that it is inner speech that is normally used for retrieving and activating relevant information for solving a given task. More generally, these kind of evidences support the Vygotskyan idea that inner speech, in the form of self-instructions, plays a major role in self-control and the voluntary control of action.

Besides providing an efficient means for remembering task information (the goal you are pursuing), language can help the development of voluntary control also as a powerful means of *abstraction*. Recent studies on language-trained chimanzee seem to provide this kind of evidence (Boysen et al., 1996). Boysen and her collegues presented chimps with two bowls containing different numbers of candies. The animal is given the bowl which it doesn't point to. Hence, to get the most rewarding result the chimp has to point to the bowl containing the smaller number of candies. Surprisingly, chimps never learned to do this. But if the same chimps, who had previously been taught symbols referring to numerals, were presented with numerical symbols instead of candies, they quickly learned to point to the smaller numerical symbol in order to get the larger quantity of candies. This result seems to demonstrate that the use of (numerical) symbols can enable chimps to master their otherwise overwhelming food-related responses. And this is possible just because of the abstractness of the symbols. When presented

---

[4] For a more comprehensive review, see Spelke (2003), who discuss in detail also the role of language in human-like numerical competence.

with symbols the chimps can just focus on the information which is relevant for solving the task. On the contrary, when presented with real food the richness of the sensory stimulus prevents them to inhibit their responses in order to reason on the strategy to adopt.

The role of language as a fundamental tool for engaging in abstract problem-solving has been even more clearly demonstrated in the contest of relational matching, with studies on both humans and chimps (for a review, see Gentner, 2003). In two series of experiments with children, Gentner and her collegues (Rattermann and Gentner, 1998; Loewenstein and Gentner, 2005) clearly demonstrated that the use of relational language helps children to solve analogical (relational) mapping tasks across a wide range of ages and task difficulty. Indeed, Gentner et al. provided also evidence for the Vygotskyan hypothesis that the linguistic aid undergoes a process of internalization: while younger children need to be provided with relational language even for solving simple tasks, older children do not, but they need linguistic help if confronted with more difficult tasks. These studies clearly point to the importance of language for reasoning on abstract (relational) properties of the world. In fact, acquiring and using a *name* for describing a relational pattern helps the child to *abstract* that pattern from the concrete context in which it has been experienced and thus it increases the probability that the same abstract pattern is recognized the next time it is encountered. In other words, labeling an abstract[5] property changes the perceptual apparatus of a child, in that it can render that property *perceivable* instead of discoverable.

But the most impressive evidence for this view comes, again, from empirical works on chimanzees (Thompson et al., 1997, see also Oden et al., 2001). Chimps (as several other animals) can learn quite easily to succeed in a match-to-sample task: that is, to choose, among two different objects, the one of the same kind of the given sample (given an A as the sample, the chimp has to choose an A gainst a B). But normal (non-linguistically-enculturated) chimps were not able to learn and solve a relational-matching task: that is, to choose, among two *pairs* of objects, the one whose objects are in the same relation as the ones of the sample pair (given an AA as the sample, the chimp has to choose a BB pair against a CD one). Most strikingly, chimps which had been previously trained to use two different symbols for the two relations 'sameness' and 'difference' were able to solve the relational-matching task. Note that in order to solve the relational-matching task chimps must apply the same/different distinction at the relational level, that is, at the level of the relation between objects. So, in order to solve this second-order problem (judging relations between relations) all it is needed is to reduce it to a first order one (judging relations between objects), which we know chimps are able to solve. And this is exactly what language training seems to do:

---

[5] In this case, the abstract property is relational, but the point can be generalized to any kind of abstract labels.

practicing with relational symbols seems to change chimps' perception. Linguistically trained chimps are able to 'see' the relation holding between two objects so that they have just to decide whether two perceived relations are same or different (beside the original paper of Thompson and collegues, see Clark, 2005 for a similar analysis of the same experiments). This, I think, is just what Vygotsky and Luria meant in the passage quoted at the beginning of this chapter: language learning transforms perception in that it makes those categories which are 'labeled' by language directly visible to the language user. The result is that the language user can apply his cognitive processes not only on simple perceptual experiences, but also upon the 'concepts' themselves, which are made perceivable through the labels.

I consider the interpretations I just gave of the exeriments cited above quite interesting, but they are surely also somehow vague and metaphorical. Science needs more rigour and hence more concrete models of the mechianisms which are supposed to explain the effects of language on cognition which have been discovered experimentally. The work I will present in this chapter represents just the first attempt to provide a computational model of the effects of language on categorization.

5.2.2. Computational modeling

In the last years there have been a very few attempts to studying the relationships between language and cognition through computational modeling. As we have seen above, classical artificial intelligence used to consider language either at the core of thought or as simply a means of communication (or both), with the consequence of denying any effect of language on other cognitive capacities. The advent of connectionism brought substantial new ideas about cognition, but most of the effort of connectionist researchers was put in trying to demonstrate how neural networks could perform 'classical' tasks as well as (or better than) classical symbolic, rule-based systems. As a result, the great majority of neural network modeling on language has focused on classical topics such as learning and understanding of grammatical forms (see, for example, Elman (1990); Plunkett et al. (1992); MacWhinney (1998); Christiansen and Chater (1999)).[6]

On the other hand, the new approaches of situated and embodied cognition replaced the study of high-level cognitive functions with that of low-level behaviors, with the result of almost completely abandoning the study of language, let alone the study of the relationships between language and cognition. In the last ten years there has been a considerable amount of computational work on language evolution inside the artificial life community, but most of these works have been focusing on the auto-

---

[6] But see chapter 14 of Rumelhart et al. (1986), the 'bible' of connectionism, and O'Brien and Opie (2002) for two interesting discussions, from the connectionist perspective, of the relationships between language and cognition very similar to the one which I present here.

organization of communication systems in populations of individuals and have almost completely neglected the possible roles that these communication systems could play in the cognitive capacities of individual agents. There are just a very few exceptions, all of which regard, in one way or the other, the role that language can play in the learning of categories.

For example, Schyns (1991) and Lupyan (2005) have shown with neural network simulations how linguistic label can simplify category learning. In fact, providing neural netowrks with labels accompanying perceptually presented objects during learning has been shown to speed up category learning (Schyns, 1991) or to improve internal representation of obejects, specifically of those categories of objects which are more difficult to learn (Lupyan, 2005).

Steels and Belpaeme (2005) have shown how the self-organization of a linguistic system between a collection of agents can co-evolve with the process of categorization of perceptual experiences because of the structual coupling between the 'conceptual' and the 'linguistic' systems. In other words, while a populational linguistic system self-organize, agents' conceptual systems adapt themselves in order to maximize communicative success.

Using an artificial life framework similar to the one presented in the previus chapters, Cangelosi and collegues (Cangelosi and Harnad, 2000; Cangelosi et al., 2000) have shown how organisms with language can learn to categorize their experience in adaptive ways not only through genetic evolution or individual learning by trial and error, but also through social learning, with what they have called ' symbolic theft'. In the symbolic theft condition learning happens thanks to (a) the prehexisting ability to categorize some stimula and (b) the exposition to others' language, which incorporates the information on how to categorize new experiences. The results of those simulations have shown that symbolic theft can give an adaptive advantage with respect to standard phylogenetic or individual learning in that it is both significantly faster and less dangerous (you don't risk to suffer the cost of errors).

Other simulations by Steels (Steels, 2003) have shown that using language for oneself can also be used to boostrap the developent of language itself towards higher levels of complexity. Discussing one of his computatoinal experiments in which agents had to evolve a compositional language, Steels reports that posing attention to the self-produced linguistic utterances proved to be necessary for a population of agents to develop a linguistic system with case grammar.

Finally, the simulations discussed in the previous chapter have shown that the use of a simple signalling system not only for communicative purposes but also for talking-to-oneself can facilitate the evolutionary emergence of the communication system itself. In particular, the reported simulations demonstrated that linguistic signals which benefit the hearer but not the speaker do not evolve if they are only used for communication while they do evolve if the hearer has to repeat the signals to himself or herself as an aid to memory.

In the remaining part of this chapter I will present an original neural network model which I built up for clarifying and developing the interesting speculations on the idea of language as a cognitive tool which I have reviewed so far. In particular, I have started with the study of the effects on language on one of the most fundamental cognitive functions, namely categorization. More specifically, the model I am going to present in the next section is going to assess the effect on the internal representations of perceived objects produced by the acquisition of the mapping between learned categories of objects with the labels that name those categories.

5.3. A neural network model of early language acquisition

5.3.1. The model

During her first year of life, the child learns to control her movements, to make appropriate sensory-motor mappings, to categorize perceptual experiences, and to reproduce her own sounds and the linguistic sounds which are present in her environment. Notwithstanding all these progresses, this phase of child development is called pre-linguistic because in her first 10-12 months the child does not show any strictly linguistic competence, that is, she is able neither to understand nor to meaningfully produce words. It is only around the end of her first year that the child learns to connect the linguistic sounds that have become familiar to her with their meanings as indicated by the fact that she reacts correctly to linguistic stimuli and she produces words in the appropriate circumstances.

Early language acquisition can be considered to involve three sub-tasks (Kit, 2002): the acquisition of linguistic forms, the acquisition of non-linguistic sensory-motor mappings, and the association between linguistic forms and specific sensory-motor mappings, which become the meanings of the linguistic forms. Behavioral evidence (Waxman, 2004) suggests that the acquisition of linguistic forms and the acquisition of sensory-motor mappings run quite independently until the end of the first year. Only after the child has acquired a certain ability to map sensory inputs into motor outputs and to categorize experiences, on one side, and to recognize and produce linguistic forms, on the other side, the third task, the association of linguistic forms with specific sensory-motor mappings, can begin. The model of early language learning presented here is based on this kind of behavioral evidence.

The neural network used in this simulations is modular. It is constituted by two sub-networks with three layers each, which I call the sensory-motor sub-network and

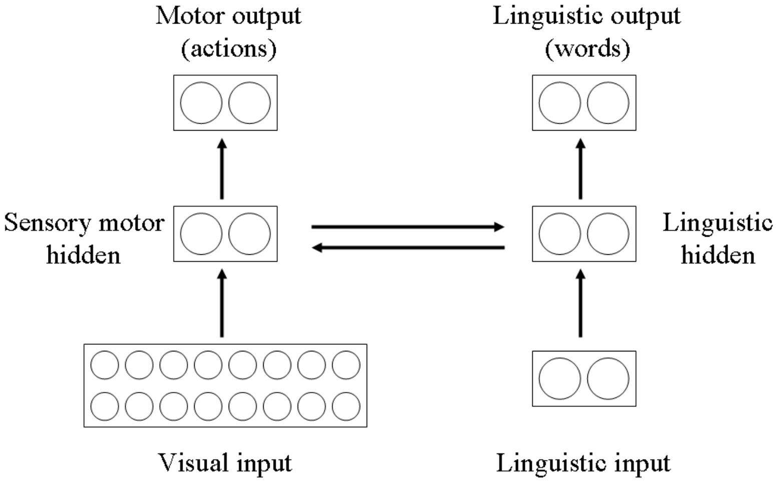


Figure 5.1: The modular neural network

the linguistic sub-network. The hidden layers of the two sub-networks are reciprocally connected by two matrices of connection weights so that the linguistic and the sensory-motor systems can interact with each other (figure 5.1). The sensory-motor sub-network has 16 bipolar input units (each unit's activation can be either –1 or 1) which encode the properties of perceived objects, 2 hidden units (with continuous activation in the interval [–1; 1]), and 2 output units which encode the action performed by the network in response to an object. The activation of the two output units is thresholded to be either –1 or 1, so that there are only four possible actions: <–1; –1>, <–1; 1>, <1; –1>, <1; 1>. The network's environment consists of 480 objects, belonging to 4 categories of 120 exemplars each. There are four prototype vectors, one for each category, and the perceptual properties of objects are generated by flipping 4 bites of the prototype which the object belongs to.[7]

The linguistic sub-network has 2 input units encoding incoming linguistic signals, 2 hidden units, and 2 output units which represent emitted sounds. All the units of the linguistic sub-network have continuous activations in the interval [–1; 1]. The linguistic environment is constituted by 4 words, which can be interpreted as the names of the four kinds of objects or of the appropriate actions to be performed upon them. Since words are pronounced in different ways by different persons and by the same person at different times, the acoustic inputs are created by changing slightly 4 prototype vectors, one for each word (<–0.5; –0.5>, <–0.5; 0.5>, <0.5; –0.5>, <0.5; 0.5>). For reasons of symmetry, there are 120 instances of each word. Each instance is produced by changing both values of the corresponding prototype vector by an amount randomly chosen in the range [–0.25; 0.25] .

---

[7] The four prototype vectors are: <–1; –1; –1; –1; –1; –1; –1; –1; 1; 1; 1; 1; 1; 1; 1; 1>; <–1; –1; –1; –1; 1; 1; 1; 1; –1; –1; –1; –1; 1; 1; 1; 1>; <–1; –1; 1; 1; –1; –1; 1; 1; –1; –1; 1; 1; –1; –1; 1; 1> and <–1; 1; –1; 1; –1; 1; –1; 1; –1; 1; –1; 1; –1; 1; –1; 1>

The entire network passes through two successive stages of learning which are meant to correspond approximately to child's learning from birth to 1 year and to her learning from 1 year on, respectively. In the first stage of learning the two sub-networks are trained independently to accomplish two different tasks. The sensory-motor network learns to categorize objects while the linguistic network learns to repeat or imitate words. At the end of this first stage the connections between the hidden layers of the two sub-networks become functional and in the second stage of learning the network learns to associate words with the categories of objects which they refer to.

At the beginning the network is initialised with random connection weights in the range [–0.5; 0.5]. Then, 4500 cycles of backpropagation learning (with a learning rate of 0.2) are run in the following way. For the sensory-motor sub-network, one of the 480 different objects is randomly chosen, its perceptual properties are encoded in the input units of the sensory-motor network, the response of the network is calculated, and the correct action is provided to the network as teaching input (correct responses are <–1; –1>, <–1; 1>, <1; –1> and <1; 1> for objects belonging to category A, B, C and D, respectively). For the linguistic sub-network, one of the 480 possible instances of words is randomly chosen and encoded in the linguistic input units, the sound produced by the linguistic network in response to this heard sound is calculated, and the same sound heard as input is given as teaching input (in other words, the linguistic network has to accomplish an autoassociative task).

During this first stage of learning, the connections that link the hidden layers of the two sub-networks are non-functional. After 1000 cycles of back-propagation learning, the inter-network connections become functional and in the second stage of learning their weights are modified so that the network learns to associate the internal representations of objects (the vectors of activation of the sensory-motor hidden layer) with the internal representations of the appropriate words (the vectors of activation of the linguistic hidden layer), and vice versa. This second phase of learning runs as follow. One of the 480 objects is chosen randomly, together with one of the 120 instantiations of the word that designates the category which the object belongs to. The object and the word are given as input to the sensory-motor network and to the linguistic network, respectively. Then, the vectors of activation of the two hidden layers are calculated and two cycles of the delta rule learning algorithm are applied by considering the two hidden layers with the connection weights in between them as two distinct perceptrons: one perceptron has the sensory-motor hidden layer as its input layer and the linguistic hidden layer as its output layer; the other perceptron has the linguistic hidden layer as its input layer and the sensory-motor hidden layer as its output layer. The delta rule is applied by using the vector of activation observed in the linguistic hidden layer as teaching input for the first perceptron and the vector of activation observed in the sensory-motor hidden layer as teaching input for the second perceptron. This procedure is applied 500 times with a learning rate of 0.2.

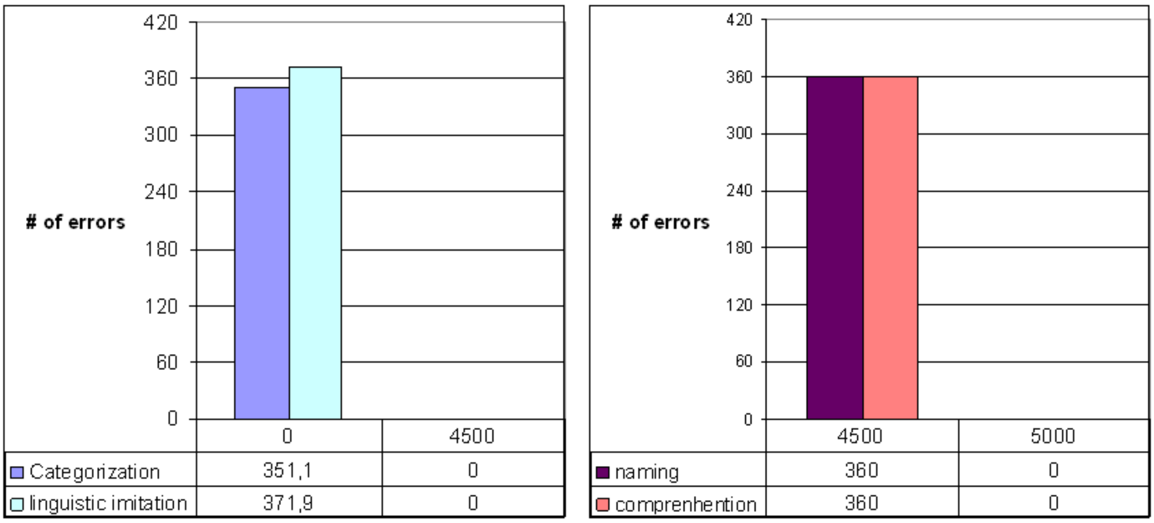
5.3.2. Results

Before studying the effects of labels on categorization we had to validate the model by assessing performance, that is by checking whether our neural network had learned the language it was exposed to. In order to do that, we run two pre-linguistic tests – categorization and linguistic imitation – and two linguistic tests – naming and comprehension.

In the categorization test the sensory-motor network is given an object as input and its performance is calculated as the difference between its action response and the right one (according to the objects category). In the linguistic imitation test the linguistic network is given the instantiation of a word as input and performance is how much the network is able to reproduce the heard sound in its linguistic output. In the naming test we give the network one of the 480 objects as input, activation spreads to the sensory-motor hidden units, then to the linguistic hidden units, and finally to the linguistic output units. We calculate the network's performance with respect to the vector prototype of the word that designates the category of the perceived object. The comprehension test is symmetrical to the naming one. In this test, we give the network an instantiation of a word as input and let the activation spread from the linguistic input units to the linguistic hidden units, then to the sensory-motor hidden units and finally to the motor output units of the sensory-motor networks. The network's performance is calculated with respect to the response that is appropriate to the meaning of the word, that is, to the action to be produced in response to objects of the category which the heard word refers to.

For each of the four behavioral test I plot the number of errors produced by the network (figure 5.2). The two non-linguistic tests (categorization and linguistic imitation, figure 5.2a) are run at the beginning of the simulation (cycle number 0) and after the first stage of learning (cycle number 4500), while the two linguistic tests (naming and comprehension, figure 5.2b) are run before the second stage of learning (cycle number 4500) and after all learning has taken place (cycle number 5000).

The network learns both to perform the correct action in response to a perceived object and to imitate the linguistic sounds it hears: for both tasks the number of errors decreases from about 360 (which represent about the 75% of the 480 total objects) at the beginning of the simulation, to 0, that is no error at all, at the end of the first stage of learning. Even the second stage of learning is very successful: while at the end of the first stage of learning the network gives wrong responses about the 75% of the time in both the linguistic tests, at the end of the second stage the network has acquired a perfect linguistic competence in that it always names objects correctly and responds



(a) (b)

Figure 5.2: (a) Number of errors in the categorization and linguistic imitation tasks before (0) and after (4500) the first stage of learning. (b) Number of errors in the naming and comprehension tasks before (4500) and after (5000) the second stage of learning.

appropriately to all the linguistic signals which are present in its environment.

5.4. How language affects categorization

5.4.1. Private and inner speech

An important characteristic of human language is that it is used not only for communicating with others but also for communicating with oneself. Indeed, the use of language for oneself starts as soon as language is acquired, and represents a great amount of linguistic production. This fact is well-known at least from the 1920s, thanks to the work of developmental psychologist Jean Piaget, who called children's talking-to-themselves 'egocentric speech'. In trying to explain this very puzzling evidence, Piaget appelled to the notion of an egocentric cognitive phase. According to Piaget, talking to oneself was just a by-product of the immaturity of the social competence of the child which played no significan role in child development itself.

Russian psychologist Lev Vygotsky gave a rather different, and much more convincing, explanation of the same phenomenon (Vygotsky, 1962). According to Vygotsky private speech is not a pre-social phenomenon, but a post-social one. It is not meaningless for child's development. On the contrary, it is of the most importance. And it does not completely disappear, but is gets transformed by being internalized. Vygotsky's theory can be summarized in the following way. When the child is talked to by her parents and other adults (or even elder children), linguistic stimuli helps her in a number of ways: they help her learn new concepts, they help her remember relevant experiences or facts, they guide her attention to important aspects of the environment, they help her solving difficult tasks and so on. Consequently, when the child starts talking to herself, she is just trying to do to herself, through imitation, what others used to do to her, that is, helping her through linguistic stimuli. In other words, children's private speech is an effect of rendering intra-personal the inter-personal linguistic social aid. Furthermore, as the child grows up and seems to talk to herself less frequently, the disappearence of private speech is only apparent: in fact, talking to oneself is just being internalized, thus becoming inner speech. Finally, this process of internalization of the linguistic social aid is of the most importance, in that it marks the transition from elementary, animal-like, cognition, to the high-level cognitive capacities characteristic of humans (Vygotsky, 1978).

5.4.2. Analyzing internal representations

As I wanted to test whether those Vygotskyian ideas could hold even with respect to a basic cognitive function like categorization, I analyzed the internal representations of perceived objects under 5 different conditions, which I call (a) no-learning, (b) no-language, (c) social language, (d) private speech and (e) inner speech. In my neural network, the internal representation of a perceived object is the activation pattern that is evoked be the sight of that object in the sensory-motor hidden units. We can represent internal activation patterns as points in an abstract space with the same number of dimensions as the number of units in the corresponding internal layer of units and with each dimension representing the level of activation of the corresponding unit (in the range [–1; 1]). A particular internal activation pattern will be represented by a specific point in the abstract space located in the appropriate position with respect of each dimension, reflecting the activation level of each unit in the pattern. All the different members of a specific category of objects, say, all the different apples, will be represented by a cloud of points.[8] At the very beginning of learning the connection weights of both modules are random. Consequently, the cloud of points representing the internal activation patterns evoked by the different, say, apples will be very large and it will largely overlap with the clouds of points representing other categories of objects. This is what is actually observed (see figure 5.3a).

After learning, the internal representations of object belonging to the same category form much better clouds, that is, clouds which are much smaller, do not overlap with other clouds, and have a considerable distance between their respective centers (figure 5.3b). The clouds of both the no-learning and no-language conditions are calculated by giving to the network only the perceptual properties of the object (figure 5.4a).

---

[8] For a more detailed explanation of the concept of a 'category cloud' and of how we can measure it, see Appendix A.

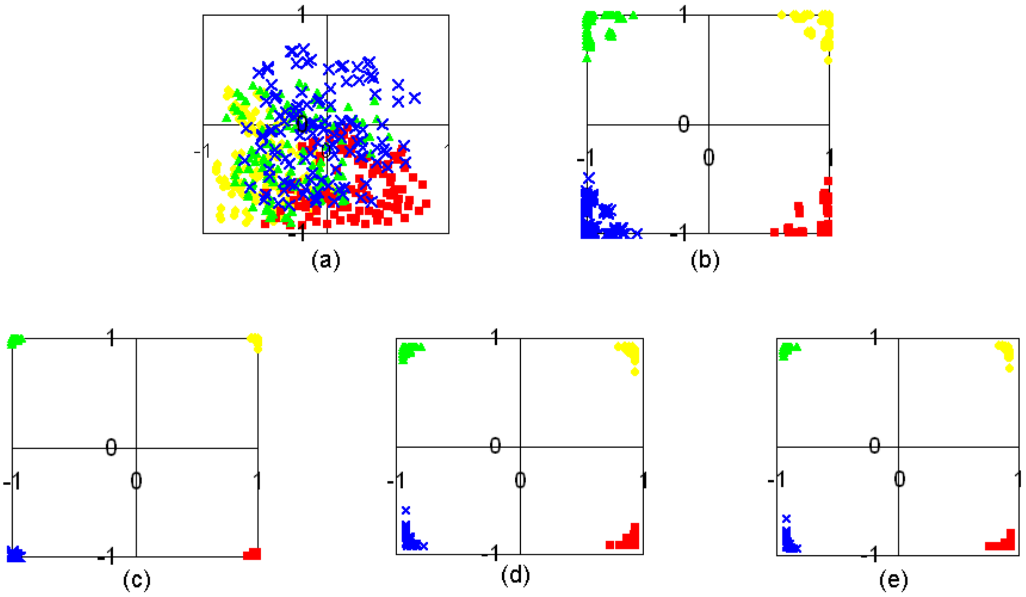


Figure 5.3: Internal representations of objects (category clouds) in the conditions of no-learning (a), no-language (b), social language (c), private speech (d), and inner speech (e). See text for details.

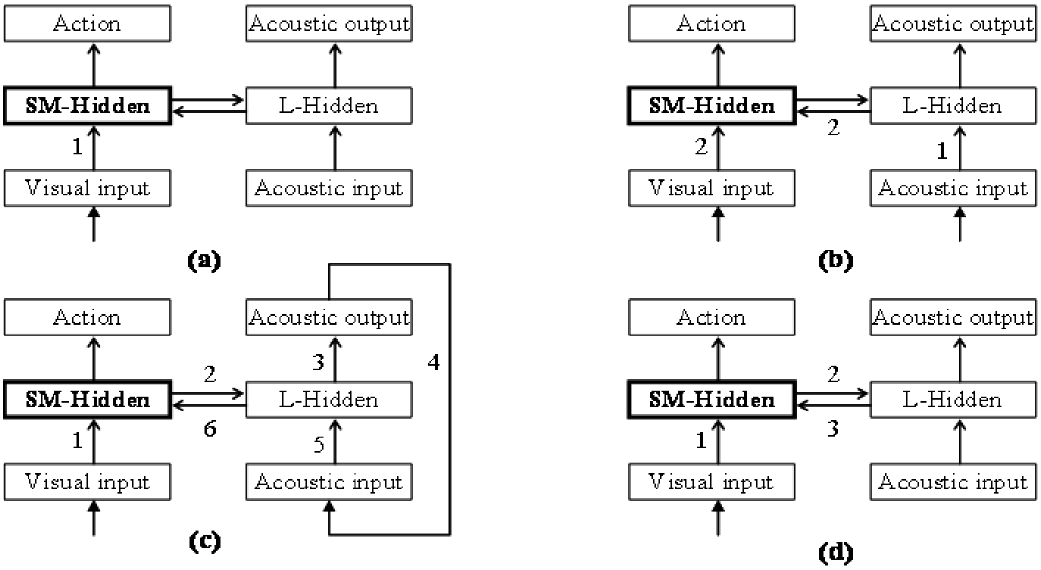


Figure 5.4: Experimental conditions for testing internal representations: (a) no-learning and no-language; (b) social language, (c) private speech, and (d) and inner speech. The process of activation is divided up into a sequence of discrete time steps. Numbers indicate connections that are involved in each time step. See text for details.

In order to test whether social language affects internal representations of object I calculated categories clouds by making the network perceive objects accompanied by instances of the words that name them (figure 5.4b). The results, shown in figure 5.3c, demonstrate that language improves categorization in that the clouds of different categories are both smaller and more distant the one to the other than the clouds of objects unaccompanied by language.

In order to study whether language can improve categorization not only if it is used socially, but also individually, I simulated the two ways in which humans talk-to-themselves: externally, as in private speech, and internally, as in inner speech. In the private speech condition (figure 5.4c) the individual encounters an object and prior to responding to it the individual produces the word that designates the object. Then the individual hears the sound that he or she has produced so that he or she can respond to an internal representation of the object which is influenced by the heard, self-produced sound.

In the inner speech condition (figure 5.4d) when the individual sees the object he or she does not produce any externally audible sound. However, the sight of the object evokes in the individual not only the internal representation of the object but also, through the connections leading from the sensory-motor module to the linguistic module, the internal representation of the word that designates the object. Through the connections leading from the linguistic module to the sensory-motor module, this internal representation of the sound that designates the object can influence the internal representation of the object in the sensory-motor module.

The results of the both the private language and the inner language conditions (figure 5.3d and e, respectively) confirm the prediction that talking to oneself improves categorization as much as the perception of the commenting of the world provided by other individuals. In fact, in both cases we see that there is about as much improvement in the internal representations of perceived objects as in the condition in which the sound that designates the object is provided by another individual. Indeed, the clouds of the two talking-to-oneself conditions are just a little bigger and less distant from each other than in the social language condition. The reason for this is just that in the social language condition internal representations on the sensory-motor hidden units are calculated by summing up the information coming from two suorces: that coming from the visual input units plus that coming from the linguistic hidden units. On the other hand, in both the two talking-to-oneself conditions, clouds are calculated by providing to the sensory-motor hidden units only the signal coming from the linguistic hidden units after talking-to-oneself, either externally or internally, has happened. In other words, after the visual stimulus has triggered the network response, it is removed so that the network responds only to the self-produced labels. This corresponds to a child seeing an object, producing its label (either loudly or just internally), and reacting to the self-produced labels without attending to the visual stimulus any more! In fact, if we test internal representation in a condition in which clouds depend not only on self-produced labels, but also on the visual stimulus, even the slight advantage of the social language condition with respect to the two talking-to-oneself condition completely disappears (results not shown).

5.5. Discussion

5.5.1. Language and categorization

In this chapter I have described a neural network model of early language acquisition. The model assumes that during her first year of life the child separately learns various sensory-motor mappings (reaching, manipulating, categorizing objects) and various sound-related abilities (recognizing sounds, repeating her own sounds and the sounds produced by others). It is only at the end of the first year that language learning begins. The two separate networks that are responsible, respectively, for the sensory-motor mappings and for the sound-related abilities become functionally connected, and language learning consists in learning the connection weights linking the two networks. These weights allow the child to produce linguistic sounds in the appropriate circumstances and to react appropriately to the linguistic sounds produced by other individuals.

But language changes the way in which the child categorizes reality. Input-output mappings require categorization. Categorization is to make the internal representations (patterns of activation in the network's internal units) of different inputs more similar if the different inputs must be responded to with the same action, and more different if the inputs must be responded to with different actions (Harnad et al., 1995; Di Ferdinando and Parisi, 2004). All organisms categorize reality. And infants categorize reality during their first year in order to generate the appropriate input/output mappings. My model shows that language can influence cognition by inducing better categorization of non-linguistic inputs.

This model can also be considered as a general model of the integration of multiple sensory-motor mappings in the brain. Much activity in the brain consists in exploiting the co-variations observed in experience to establish correlations between different sensory-motor mappings. For example, visual-motor mappings are correlated with proprioceptive-motor and tactile-motor mappings in such a way that given one type of input the brain can predict the other one. Language is but one example of this type of integration between different sensory-motor mappings. In particular, it is the integration between various non-linguistic sensory-motor mappings and the sensory-motor mapping from heard sound to pronounced sounds. Given a visual input from an object, the brain generates the internal representation of the word (sound) that designates the object. Or, given a word (sound), the brain generates the internal representation of the visually or haptically perceived object designated by the word.

However, language is somewhat different from other types of integration between different sensory-motor mappings. While other correlations between different sensory-motor mappings are not arbitrary, language is arbitrary, and this may require a species-specific genetically inherited basis for language learning. Given a visual input from a new object, the brain can predict how the visually perceived object would feel if touched. In contrast, given a new object the brain cannot predict what is the word that designates the object or, given a new word, the brain cannot predict which object is designated by the word. Furthermore, variations in how the same word is pronounced do not predict variations in the object designated by the word. Another important difference between language and other types of integration between different sensory-motor mappings is that the overall space of variation of heard or pronounced linguistic signals appears to be much smaller than the space of variation of the objects or actions that are designated by the linguistic signals. This is particularly important for categorization because it might contribute to explaining why language makes our sensory-motor categories more distinctive and more compact.

But the most important point to be made here is that language can be conceived as a system which is specifically devoted to the kind of mapping that can give the kind of advantage on categorization that our model has demonstrated. In fact, we know that acoustic input is very early distinguished by an human brain in linguistic and non-linguistic, and the two different kind of input are furtherly processed by different parts of the brain. And the mapping between, for example, visual and non-linguistic acoustic input is not systematic, at least not as systematic as the mapping between the visual and the linguistic acoustic input. It is certainly true that I sometime perceive, for example, the visual apparence of a dog together with its barking. But only a very small fraction of the visual categories that I have do have specific non-linguistic sounds systematically related to them, and, vice-versa, only a small fraction of the non-linguistic acoustic categories that I have do have specific visual odbjects systematically related to them. On the contrary, the linguistic system is devoted to the systematic mapping of words to the representations of categories of stimuli in other parts of the brain. So, generally speaking, each category we can have in mind has a specific word – or a collection of words, like a sentence or a discourse – for naming it. Conversely, words refer systematically to specific aspects of the sensory-motor experience of an individual – or to abstractions of that experience. This systematicity in the mapping between sensory-motor experiences and arbitrary labels is exactly what produces the kind of effect on categorization that we have seen in our simulation.

In short, I do think that perceiving *any* two distinct kinds of information which relate to the same category of objects do facilitate the categorization process. This will presumibly produce a better internal representation than in the case in which the information on the presence of the object comes from one single sensory modality. But this effect on categories is particularly important for the linguistic system because the mapping between the linguistic system and the other sensory-motor sytems is arbitrary and systematic, while in other cases it its either non-arbitrary or non-systematic or both.

Finally, in a vocal acoustic language the sensory-motor mapping from heard to pronounced words can easily become a circular or closed circuit. When one sees an object, the visual input from the object does not only generate the internal representation of the word that designates the object but it can generate the actual sound of the word through private or inner speech. In this chapter I have also provided a first neural netowork model of private and inner speech. The results of the analysis of internal representations show clearly that the advantage given by heard labels for categorization is mantained even when the individual talk-to-herself, both in private and in inner speech. These results support the Vygotskyian idea that talking-to-oneself constitutes a fundamental process of child development, in which the child internalizes the linguistic social aid. This work coonstitutes the first running computational model which supports this general Vygotskyian view for one of the most basic cognitive function, that is categorization. As I will discuss in the next chapter, the goal of this line of research is to develop further the model in order to understand the effect that language can have in the development of most – if not all – human cognitive functions.

5.5.2. Implications for language evolution

Human language has evolved. If we go back sufficiently in time we find ancestors of present-day humans who did not have language. What is less clear is how and why language has evolved. If any capacity has evolved one can ask what adaptive advantages the possession of that capacity conferred to the individuals that possessed the capacity over other individuals that lacked it. This should not imply a pan-adaptivist view of evolution. Novelties can appear evolutionarily not because they are adaptive but because they are inherited together with characteristics that are adaptive, or they can be almost necessary by-products of already existing adaptive characteristics, or for chance reasons. And even adaptive characteristics may not evolve if the concerned organism is not pre-adapted to them, i.e., if the organism does not already possess other characteristics that make the evolutionary emergence of the new characteristics possible. However, although these various limitations to a pan-adaptivist view of evolution are all appropriate and correct, still it is at least heuristically useful to ask what the possible advantages of an evolved characteristic are because this may allow us to explore various possible evolutionary scenarios that can be compared with what we actually know about the past.

In the case of language it seems clear that language may have evolved because it conferred social advantages by dramatically improving communicative capabilities of human beings which in turn improved coordination among different individuals. The idea that language is just a very complex and powerful communication system can indeed be considered as the standard view in the debate on language evolution (see Section 2.8).

However, another less often explored possibility is that human language may have evolved (also) because it made the cognitive functioning of single individuals more effective. Unlike animal communication systems human language can be used not only socially but also individually. It can be used to communicate with other individuals, by asking them information or providing them with information or by asking them to do one thing or another, but it can also be used to talk to oneself, to comment on what one sees, to put ones predictions, explanatory hypotheses, and plans into words. These individual uses of language can result in more effective behaviours on the part of the individual even when the individual is acting alone, and this may have represented a reproductive advantage and a selective pressure for the emergence of language. In the literature on language evolution the individual functions of language tend either to be ignored or to be thought to have appeared much more recently than it would be required if these uses were to represent a selective pressure for the very beginning of human language. It is assumed that human beings have learned to talk to themselves when their language was already completely developed and indistinguishable from the language spoken by present-day humans. The initial selective pressures for the emergence of language were social. When language was already evolved and fully modern, humans found that it could be usefully used to talk to oneself and not only to talk to others. But it is not clear that this is necessarily so. In fact, it is possible that even a very simple form of proto-language, consisting of words (or holophrases) that correlate with relevant experiences, can give important individual advantages once it is used not only for communication but also for talking-to-oneself.

The simulations presented in this chapter have shown that using language for oneself can improve an individuals categorization of the world. I argue that human language may have emerged not only due to social pressures, i.e., because language made it possible to have more sophisticated forms of social communication and coordination, but also due the advantages that language conferred on the cognitive functioning of the single individual when it is used for talking to oneself, either aloud or internally. Since these advantages can be demonstrated even with a very simple language lacking any syntax, I conclude that it is not necessary for language to be as sophisticated and complete as present-day language to provide individual cognitive advantages but that these advantages were probably already present in the very early stages of the evolutionary emergence of human language.

Part III.

Beyond

6. The influences of language on human cognition

> The internalization of socially rooted and historically developed activities is the distinguishing feature of human psychology, the basis of the qualitative leap from animal to human psychology.
> Vygotsky: Mind in Society, pag. 57 (Vygotsky, 1978)

6.1. Introduction

A crucial, but often neglected, characteristic of human language is that language is used not only for communicating with others but also for communicating with oneself, i.e., for thinking, whereas we don't have evidence for this type of use in animal communication systems. Inputs to an individual's linguistic network can come from another individual's linguistic network but they can also come from the individuals own linguistic network: the individual talks to itself. If the sounds are actually (physically) produced by the individuals phono-articulatory movements and actually heard by the sensory units of the individuals linguistic network, we call it private speech. If the loop does not include the organisms peripheral motor and sensory organs but is more internal, we call it inner speech (Chapter 5). Inner speech is faster than private speech, and has also some structural differences which can have important consequences for certain purposes, but in both cases a number of interesting effects on the individuals cognitive activity can be observed.

As I have discussed in the previous chapter, one could think that using language to talk to oneself is a late development in human evolution and it presupposes an already complex language. However, some of the simulations presented in this thesis have demonstrated that a full blow syntactic and compositional language is not necessary for benefiting from talking to oneself. In fact, the simulations of Chapter 4 have shown that talking to oneself can be a selective pressure for the evolutionary emergence of a very simple communicative system if the linguistic signals are used by the individual to keep in memory some information which has been received from another individual. Furthermore, the simulations of Chapter 5 have demonstrated that learning labels for already acquired 'concepts' can improve an individual's categorization ability. Hence, both these simulations suggest that using language for oneself can have advantages for the individual even if the language is very simple, and that these individual advantages, once discovered by evolving hominids, might have played a role in the subsequent evolution of language itself.

Using language to talk to oneself has a number of important consequences for human cognition. Indeed, it can be argued that it constitutes a (the) fundamental step in the development of most, if not all, human high-level cognitive functions. In this chapter I will provide some ideas on the effects that the acquisition and use of language can have on the following cognitive abilities: learning, abstraction, memory, attention, problem-solving, voluntary control, and mental life.

6.2. Learning

The importance of labels on individual learning is surely the most studied effect of language on cognition. In particular, as we have seen in Section 5.2, the few attempts of studying the relationships between language and cognition by computational modeling have been almost exclusively focused on the role of language in category learning (Schyns, 1991; Lupyan, 2005; Cangelosi and Harnad, 2000; Steels and Belpaeme, 2005). Nonetheless, the field is only at its very beginnings and much more work has still to be done in order to provide a full account of the mechanisms that underlie the role played by language in learning. I argue that the facilitatory effect of labels on category learning derives from the following two mechanisms: (a) linguistic inputs constitute additional stimula that focus the learner's attention to the specific aspects of perception that are relevant for categorization itself; (b) language itself can sometime represent the only ground in which the learner can develop the discriminative capacities that *constitute* categorization. Let's consider these two points in order.

Hearing the same linguistic stimulus, let's say the word 'red', when perceiving red cars, red apples and red flowers facilitates – or may even *induce* – the acknowledgement that all those different stimula have something in common, namely the red color. In neural network terms, this means that the occurrence of the same pattern in the acoustic input group of neurons – namely the pattern that correspond to the word 'red' – increases the similarity of the internal representations of all red stimula, and this in turn can help – or induce – the network itself to learning that all those stimula belong, in some respect, to the same category, namely that of red things. A similar point has already been made in the emirical literature on the topic (see, for example, Waxman, 2004). But, in my view, this is not the whole story. The point is that perception depends in a fundamental way to action: to categorize means to produce a given *behavior* A when perceiving a certain class of stimula and another behavior B when perceiving another class of stimula. It is the need to respond appropriately (and discriminatively) that makes an organism perceive the first class of stimula as *different* from the second one (similar action-based views of cognition have been developed, among others, by Di Ferdinando and Parisi, 2004; Gallese and Lakoff, 2005). So, when dealing with a categorization process, we must always ask the following question: which is the differential behavior exibited by an agent that makes us say that the agent is categorizing some experiences as belonging to the same category and other experiences as belonging to another category? I argue that, for many of human's categories, the answer is just: the production of different *words*! In other words, I am claiming that in many cases the human neural network learns to represents some patterns of inputs (for example those produced by red cars, red apples and red flowers) as similar to each other and different to other patterns (those produced by white cars, apples and flowers) principally because it is learning to produce, through its phono-articuilatory output units, the same (or different) *action*, which consists in the production of the same (or different) word: 'red' (or 'white'). This is not to deny that there is some internal 'appreciation' of different colours, nor that there is no genetical tendency to discriminate colours: we all know, from direct experience, that we have different appreciation of *redness* and *whiteness*, and we know from scientific investigations that there is some genetically based capacity to discriminate colours in certain ways. The point is that part of the specific way a particular human being categorizes experiences (in this case, related to colour) is in fact due to the way s/he learns to name them. In other words, according to this view, the amazing amount of categories that humans can have is in great part due to the role played by language in providing a behavioral ground for categorization.

Furthermore, the fact that human language is culturally learnt – at least the part of language which principally incorporates a category system, that is lexicon – does not only facilitate children's learning of categories, but has also another important consequence: it guarantees that learnt categories are more and more useful in the individual's adaptation to the environment. In fact, being culturally transmitted, the categorical distinctions embedded in language undergo a process of cultural evolution, which rises dramatically the likelihood for those distinctions to be adaptive. This could be tested with a simulation. My prediction is that a population of individuals interacting with their environment and possessing a cummunication system which is culturally transmitted and evolved would develop better categories than a population of individuals whose adaptation depends exclusively on genetic evolution and individual learning. The quality of the category system could be indirectly measured as the level of adaptation to the environment that the category system permits. The better adaptation of a population with a culturally trasmitted language would be explained by two factors: (a) the fact that the learning of the communication system influences positively the learning of the category system and (b) the fact that the category system itself can undergo a process of cultural evolution through which new adaptive ways of categorizing experience discovered by single individuals can spread in the population as a result of their embedding in the linguistic system. In other words, the acquisition of a linguistic system makes it possible for an individual to exploit the experience of other individuals, both living and long dead, and acquire more easily more useful categories.

6.3. Abstraction

The model presented in Chapter 5 has shown that language can have an important effect on the internal representations of sensory stimuli which is independent on learning. In fact, my simulations have shown that learning the mapping between pre-linguistically learned concepts and linguistic labels changes the internal representations of objects. The reason is that the internal representations of the non-linguistic network, which prepare the motor outputs with which the non-linguistic network will respond to its sensory inputs, tend to be influenced not only by the sensory inputs to the non-linguistic network but also by the linguistic network. This has the effect that the non-linguistic networks categories tend to become better categories, in the sense that the clouds of internal representations of objects belonging to different categories become smaller and more distant from each other. And since an organism's categories influence the organisms behaviour by making it easier for the organism to select the appropriate action in response to sensory inputs, an organism endowed with language will have a more effective behaviour.

This effect of labels on internal categorization has further important consequences, as can be seen if one considers the effects of language on categorization with respect to the process of *abstraction*. In fact, categorization requires abstraction. In order to respond in the same way to different stimuli which belong to the same category you need to abstract from their differences. And, viceversa, in order to respond in different ways to similar stimuli which belong to different categories you have to abstract from their similarities. Reducing the size of a category's cloud is just improving the first kind of abstraction (ignoring differences bewteen intra-category stimuli), while increasing the distance between clouds is just improving the second kind of abstraction (ignoring the similarities between inter-category stimuli).

The model presented in the previous chapter had one important simplification: in that model, each object had one and only one specific action associated with it. This is a clear limitation becuase for real organisms the same object can typically evoke several different responses, depending on the context. For example, an apple can be eaten, thrown, gifted, and so on. In other words, depending on the context, the same object can be categorized in different ways – in the apple example, as food, as a contundent thing, or as something that another individual would appreciate. This means that the internal representation of an object must be multi-functional, in the sense that it must allow the organism to consider the same object *as* belonging to different categories depending on the circumstances. My hypothesis is that linguistic labels help organisms to abstract from the ways in which an object can be categorized which are not relevant from the current situation, and to focus only on the categorization which is relevant.

Furthermore, labels can also induce hierarchical categorization. Actions can be hierarchically organized in the sense that two sets of sensory inputs can be responded to by two different actions, and therefore they constitute two distinct clouds of points, but there is a third action with which the organism responds to both sets of sensory inputs. Therefore there is a third cloud of points that includes both the first and second clouds of points. Language can favour the creation of hierarchies of clouds of points just because it provides hierarchies of labels: there are two linguistic signals, e.g., 'dog' and 'cat'", that correspond to two distinct clouds of points, and there is third linguistic signal, 'pet', that evokes the point located centrally in the larger cloud of points including the 'dog' cloud and the 'cat' cloud. Recall the behavioral studies on monkeys discussed in Section 5.2, which have demonstrated that chimps that have been not linguistically trained cannot discriminate between *relations of relations,* while chimps that have learnt a symbolic system which include words for relations can (Thompson et al., 1997). This suggests that training in linguistic tasks changes the way an agent perceives the world. The interactions with the world of a linguistically trained animal is mediated by linguistic forms, which render some of the aspects of experience more salient than others. Furthermore, this process is recursive: once you have learnt to see the world in certain ways you can also discover new, more abstract patterns. For example, recognizing that sometimes you are looking at the *colour* of objects while in other occasions you are looking at their *form* may allow the development of more and more abstract concepts, like the concepts of 'colour' and 'form'. In other words, while a non-linguistically trained animal which is able to react appropriately to stimuli of different colours (red vs. yellow) has surely different representations of those stimuli, it probably doesn't have the concept of 'redness' vs. 'yellowness'. And it surely doesn't have the more abstract concepts of 'colour' vs. 'form', not to speak about the still more abstract concept of 'property' (of which 'colour' and 'form' are two instances). Our abilities of constructing more and more abstract categories depends on our ability to label discovered categories and hence to reason on the categories themselves.

6.4. Memory

Another example of the importance of language for cognition concerns memory. One way of simulating short-term memory with neural networks consists in copying the activation pattern of a networks internal units in a special set of memory units and then connecting the memory units to the internal units in order to allow the network to retrieve the memory traces (see Elman, 1990 and Appendix A). This kind of short-term memory mechanism is present in both the non-linguistic part of our brain – represented in my model by the non-linguistic sensory-motor network – and in the linguistic part – represented by the linguistic network (Baddeley, 1992; Petrides et al., 1993; Becker and Morris, 1999). However, in contrast with what happens in my simplified model, in real brains the non-linguistic and the linguistic networks have a significantly different size. Specifically, the non-linguistic network is much greater than the linguistic one, in terms of the number of units and connections of which the two networks are composed. This simple fact provides a clear advantage for the specifically human linguistic memory system with respect to the older non-linguistic memory system which we share with other animals. It is in fact generally easier to remember words than the actual sensory-motor experiences to which words are associated. Hence,an individual possessing language can work more easily with linguistic (sound) information and translate this information into the associated non-linguistic information when necessary. Furthermore, possessing a linguistic memory system in addition to the older sensory-motor one has a second, fundamental advantage: delegating the memory function to the linguistic system leaves the sensory-motor system free to process other kinds of information which are useful for acting in the environment while linguistically remembering previous information.

These advantages of the possession of a linguistic memory for short-term processes extend also to long-term memory. Instead of memorizing full experiences a human being can label them and memorize their verbal description. Thanks to the abstracting power of language, what is to be remembered for us are just the most relevant features of a given experience. And thanks to the small size of the linguistic network those relevant features are coded in a very efficient way, so that they can be easily memorized and recalled. It is thanks to the possibility of this linguistic coding, memorizing, and recolling that human beings are able to recall events which happened in their very distant past. This hypothesis could also explain why our memories never go too far. Typically, the first rememberings never date back before about the third or fourth year of life. This is just the age at which private speech begins to appear in the child (Berk, 1994). I argue that we can't recall anything about our first three or four years of life because in those years we haven't learnt to label our own experiences and to memorize them in verbal form yet.

Finally, the improvement of memory which is made possible by language might be a pre-requisite for the development of another peculiar characteristic of human language, namely *displacement*. Animal signals tend to be deictic, that is, they communicate information which is only true given the current state of the sender and the receiver of the signal and their current location in space. Human language can communicate information about other places and other moments in time. I argue that the use of language for memorizing relevant information might have constituted a pre-requisite for the evolution of displaced communication. In other words, according to this view it is because we can memorize and recall relevant information in an extremely efficient way through language that we are able to communicate about things which are not present here and now. Furthermore, the linguistic coding of memories makes them ready to be communicated whenever the necessity to do so arises.

6.5. Attention

Language can also be a mechanism for directing attention to specific portions of the input arriving from the environment and for articulating or analyzing complex sensory inputs. All organisms need selective attention mechanisms since all organisms live in environments that send to their sensory organs many different inputs at the same time, and the organism must select which of these inputs to process in order to generate a response, while ignoring all the other inputs. Language can be such a selective attention mechanism. When an individual sees a complex scene, a word originating from another individual which accompanies the perception of the complex scene can help the individual to isolate some particular component of the complex scene and to respond to this component, ignoring the other components. This is a consequence of the co-variation of specific sounds with specific non-linguistic inputs: the same co-variation which gives linguistic sounds their meanings. Language can also help the individual to articulate a complex perceived scene into its elements. A sentence is a collection of linguistic sounds (words) each of which co-varies with a different component or aspect of a complex scene so that the sentence makes it easier for the individual that hears the sentence to isolate these different components and aspects and to respond to them more effectively.

Objects can have different properties that are relevant for different purposes and actions on the part of the organism. For example, objects can be of different sizes, shapes, colours, etc. These different properties are probably all included in the object's internal representation in the sensory-motor module. Language includes signals that evoke these distinct properties of objects rather than the entire object (this is the usual function of adjectives. These signals can direct the organism's attention to some specific property of an object rather than to the entire object. For example, the linguistic signal 'big' which accompanies the signal 'dog' directs the attention of the hearer (or of the speaker if the signal is self-generated by the organism) to one specific property of the dog.

Since one and the same property can be possessed by different objects, for example both dogs and cats can be big, one can hypothesize that linguistic signals evoking internal representations of properties, e.g., ’big’, rather than of objects, e.g., ’dog’, create a new type of clouds in the internal units of the sensory-motor module, i.e., clouds that include portions of the internal activation patterns evoked by different objects that belong to different clouds at the object level.

My hypothesis is that these clouds at the property level constitute a cognitive analysis of sensory inputs which is not accessible to organisms without language or with a language which does not include these type of linguistic signals. These linguistic signals, e.g., adjectives, by inducing a cognitive analysis of sensory inputs, allow an organism to respond in a more sophisticated way to these inputs. (Adverbs can play the same role with respect to verbs designating actions as adjectives with respect to linguistic signals designating objects.)

6.6. Problem solving

In human beings language does not mediate only the learning of categories, or, more generally, of *knowledge*, as discussed above. It also mediates the learning of *abilities*. Consider the case of a difficult practical problem, whose solution requires, for example, a complex sequence of actions. The discovery of such a solution may be too difficult to be found with classical trial and error learning. A child who has learnt language can nonetheless be *instructed*, through linguistic commands, about the sequence of actions that must be performed to solve the problem. Linguistic instructions can enable the child to produce the appropriate behavior in the appropriate circumstances, at least when the helping adult is present. But the importance of language does not stop here. If the problem’s solution is really complex, as we have supposed, it is extremely unlikely that the child can learn it in the few times she is helped by adults. But learning *sequences of words* it is much easier than learning compelx sequences of actions (see Section 6.4). If the child learns the words’ sequence that enables her in the problem’s solution, she can recall that sequence when needed thus guiding herself through the solution of the problem. This kind of linguistic auto-stimulation can successively follow the internalization process which is characteristic of talking-to-oneself. Furthermore, since verbal instructions about how to solve a complex problem can be passed, thorugh language, from an individual to another, they can undergo a process of cultural evolution, during which they can be refined and improved. Hence, through this learning process mediated by language the child is able to learn the solution of extremely difficult problems whose learning is precluded, not in principle, but in practice, to organisms that lack language.

But throught this process of internalization of the linguistic social aid the child does not only learn to solve specific complex tasks. She also learns the more general practice of including language in problem solving. This has effect of dramatically improving problem-solving abilities themselves. In fact, through the introduction of language, problem solving can benefit from all the above-mentioned advantages provided by language plus a few others. First, through language the child can introduce in the possible problem solution objects which are not immediately present to her attention. On the contrary, non-linguistic animal's problem solving is very much stimulus-driven. Second, language dramatically improves the human capacity to make predictions: in particular, to predict the consequences of one's own actions. Predictions becomes in fact more complex and more articulated if they are linguistically labeled, and the individual which has labeled his predictions can also work more effectively with them. She can recall them more easily, she can more easily compare the different predicted outcomes of different actions for deciding which action to produce, and she can more easily concatenate sequences of possible-actions and their relative predicted-consequences in a way which enables her to construct a *plan* for actions. In this way, language makes problem solving easier, more effective and more reasoned. And, again, linguistically labeled predictions and plans can be easily shared and discussed with others, making the overall predicting and planning capacity of single individuals and of groups even more effective.

6.7. Voluntary control

Another fundamental consequence of the internalization of linguistic stimulation for the purpose of guiding one's actions is the development of voluntary control. Animal's ability to voluntary control their own actions is very limited. Non-human animals are fundamentally stimulus-driven: besides the already-mentioned limitation in their capacity to include non-perceived object in their problem-solving, they are also easily distracted by non-relevant stimuli, in that they can hardly inhibit their istinctual responses to highly motivating stimuli. Of course, under the same conditions human beings experience the same kind of difficulties, but we are able to overcome them in ways which are not accessible to other animals. We can control our behavior, we can focus on our tasks, and we can inhibit our istinctual responses to even the most motivating stimuli. And we develop these abilities, I argue, through the internalization of the incitaments, suggestions, and commands which we receive from other individuals – most notably, our parents – during our infancy.

The idea is once more the same. The behavior of the child is constantly controlled, through linguistic stimulation, by other people: during all our infancy, we are continually instructed about all kinds of do's (wash your hand, clean your teeth, do your home-works...) and dont's (don't get dirty, don't eat sweeties, don't watch too much TV...). Once an individual has experienced the positive effects of being guided by linguistic stimuli produced by other individuals, she learns, by imitation, to linguistically stimulate herself in the same situations in order to produce the same effects. That is, she starts to talk to herself as a means for control her own behavior. It is in this way, I argue, that we learn to make ourself do what we know is important but we don't like much to do and, conversely, to prevent ourselves from doing what we are motivated to do but we know we shouldn't. This is what, later on, we call voluntary control, or Will. And, again, this is still another demonstration of the idea expressed by Vygotsky in the quotation reported at the beginning of this section. It is "the internalization of socially rooted and historically developed activies", typically mediated by language, which provides the individual human mind with new foundamental cognitive abilities which are absent in "animal psychology".

6.8. Mental life

Finally, by endorsing a Vygotskyan perspective on an idea which have been recently put forward by Daniel Dennett (1991), I argue that language plays a major role in the most striking and peculiar characteristic of human mind, that is, human mental life. Human beings have a very rich mental life which includes visual and motor images, rememberings, dreams, hallucinations and so on. Mental life can be considered as the self-generation of one's own input (Parisi, 2005b). For example, mental images are self-generated input (typically visual input, but we can imagine any kind of sensory-input, including the proprioceptive one) which generally we have not recently received but is actively produced by the nervous system itself. Rememberings are self-generated input which we know we have not received recently but we also know that we have received in the past. Dreams are self-generated inputs that occur when one is asleep. Hallucinations are self-generated input that we erroneously believe is coming from the external environment, and so on. The first and most obvious reason why language plays a fundamental role in human mental life is simply because a significant part of our self-generated input is linguistic: in other words, a large part of our mental life is constituted by internally talking to oneself, that is, by inner speech. But I think that the role of language in mental life goes beyond this. In what follows I will try to explain why.

Learning to self-generate one's own input might depend on learning to predict the consequences of one's own actions. By learning to self-generating one's own input externally one can learn how to control one's own behaviour. Afterwards, this capability can become internal by means of the association of the context to the sensory-motor traces of the effects of one's own actions. In other words, after I have learnt to induce to myself a sensory-motor experience by manipulating the external environment, I can learn to do the same internally, by predicting the sensory-motor experience I would have if I would do the appropriate actions. This is a possible explanation of how we learn to internally self-generate our own input. Indeed, it is possible that we learn to do that first of all with words, and then we can transfer this ability to all the other modalities. The reason is that language constitutes the easiest domain for learning to predict the consequences of one's actions.

The linguistic system is different from the other sensorymotor systems because in the linguistic system the mapping between the input and the output is far more direct, complete and stable than in other systems. In other words, when you produce a sound you can also predict very reliably the acoustic input that you will receive as a result. This is not true for other kinds of sensory-motor systems. In fact, it is true only for very simple movements (you can reliably predict the proprioceptive input given some motor command). And in fact we know that there are many circuits devoted to this kind of prediction (called internal models, see Wolpert et al., 1995). One of the most basic functions of these circuits is probably to provide self-generated proprioceptive feed-back in cases in which the real feed-back would be too slow for the behaviour to occur properly (see, for example, Clark and Grush, 1999). But the input which is self-generated by internal models is just proprioceptive, so it is useful just in the sensory-motor coordination of the organism itself, but not for other more complex kinds of activity.

The situation for the linguistic system is completely different, for several reasons. First, the linguistic input which we will receive as a consequence of our phono-articulatory actions is almost completely determined by the phono-articulatory action itself. Second, these consequences of our actions are always present (we cannot but hear what we say), and, thanks to the importance of language for our cognition, we always pay attention to linguistic input. This makes learning to predict our own linguistic stimuli given our linguistic action very straightforward. Furthermore, linguistic tokens have meanings, in the sense that they are associated with other sensory-motor experiences which are relevant to the organism and which they tend to restore. This simple fact renders the self-generation of linguistic stimuli particularly important in that self-generating a word will cause the self-generation of the non-linguistic sensory-motor experience which is associated to that word. In other words, both the association between the phono-articulatory movements and the resulting acoustic sounds and the association between a given sound (word) and its meaning (that is the sensorymotor experience associated with that word) are very systematic, reliable, and almost immediate for a developing child. This makes it easy for the child to learn to self-generate his or her own (non-linguistic) input by simply producing the words which are associated to the particular sensorymotor experience he or she wants to have.

On the other hand, how could a non-linguistic animal self-induce some experience? On the one side, it can self-generate proprioceptive inputs by its internal models. On the other side, it can control the input it receives by (overtly) directing its attention towards the desired objects. But how could it learn to direct its attention towards something which is not in the immediate surroundings? And hence, how could it learn to have mental images or rememberings, that is, to self-generate the experience of something which is not present here and now?

There is no real (external) action you can do in order to let you perceive the image of the Coliseum if you are not in front of it or very close to it. So there is no way in which you can internalize this ability by just thinking about that same action and self-generating (predicting) the consequences of it. But if you have learnt to associate the stimulus of the Coliseum and the word 'Coliseum' so that hearing that word will tend to re-activate the internal experience of seeing the Coliseum, then you just need to produce the word by yourself and listen to what you have produced and you can re-experience the Coliseum. Furthermore, since it is very easy to predict the acoustic effects of your phono-articulatory movements, the whole process can be easily internalized: the next time you want to re-experience the image of the Coliseum, just think about producing its name, and this will immediately trigger the desired internal image.

7. Conclusion

In this thesis I have presented artificial life simulations related (a) to general problems on the evolution of communication systems (including language), and (b) to the effects that acquiring and using language have on the development of human cognition. Here I will just summarize the major contributions that this thesis has given with respect to those two topics.

With respect to the evolution of communication, I addressed two related problems, a phylogenetic problem and an adaptive problem. The phylogenetic problem is the following: how can communication emerge in a species' phylogenetic history, given that the two abilities that communication requires – that is, speaking and hearing – are complementary? Traditional evolutionary thinking on this problem has focused on a possible solution to this problem which refers to the exhistence of receiver biases. But this seems to be a viable solution only for fundamentally manipulative communication systems. My original proposal is that the evolution of communication systems which provide benefits principally to hearers can be explained by an opposit (but possibly complementary) mechanism: the exploitation of preducer biases. In particular, my simulations have shown how the need for evolving organisms to categorize their environment in adaptive ways might result in a spontanous tendency, in speakers, to produce usefull signals, that is signals that systematically co-vary with adaptively relevant features of the environment. Recent empirical studies demonstrating the spontaneous differentiation of vocalizations in monkeys which were trained to produce general calls in two different conditions seem to provide an empirical support of my hypothesis. Of course, much more empirical evidence is needed for assessing the real empirical importance of this idea.

The other problem I addressed with respect to the evolution of communication is the following: how can communication systems which provide benefits only to one of the actors of communication, in particular the hearer, evolve? More specifically, why should speakers produce useful signals if they do not get any benefits in doing so? This question has been raised both with respect to the evolution of animal communication systems and with respect to the evolution of language, and has been addressed both through 'classical' verbal theories and mathematical modeling techniques and through computational models. The contributions of my artificial life simulations to this topic have been the following. First of all, I have demonstrated that the evolution of altruistic communication cannot be fully explained, as had been suggested by previous models, by the supposed cognitive pressure towards spontaneous good signal production. As I have shown, this cognitive pressure can act as a producer bias thus solving the phylogenetic problem of speaker-hearer coevolution, but it cannott solve, in itself, the adaptive problem. In particular, it cannott prevent the disruption of the communication system produced by the invasion of cheaters. Second, I have provided three distinct mechanisms which could explain the evolution of altruistic communication. The first mechanism is kin selection. If communicative interactions tend to happen principally among kin, then altruistic communication can emerge because kin tend to share genes, and hence altruistic speakers will tend to advantage hearers which possess the same altruistic genes. Kin selection seems to be the best candidate solution for explaining most of animals' communication systems, but it is still not clear which role could have it played in the evolution of human language. On the contrary, the other two mechanims which I have provided for explainig the evolution of altruistic (uses of) communication are specifically related to two characteristics of human language: the fact that it is, at least in part, culturally transmitted and the fact that it is used not only for social communication but also for talking to oneself. My docility simulation has shown that if organisms do not inherit behaviors directly, but only the genetic predisposition to learn from others, then the altruistic use of communication can emerge as a by-product of the evolution of organisms' docility, that is, of the organisms' tendency to being taught how to behave from others. Finally, my talking-to-oneself simulation has shown that if signals are used not only for social communication but also individually, in particular as memory aids, then the emergence of communication is favoured because talking-to-oneself poses a direct selective pressure towards good signal production.

The recognition that talking to oneself could have palyed a role in language evolution drove me to study the possible effects that using language for oneself can have on human cognition. In this context, I have first provided an original neural network model of early language aquisition (learning to map the first words with the first concepts). Then, I have provided a simple model of private and inner speech. Finally, I have shown how language, be it social, private or inner, can aid one of the most basic cognitive functions, namely categorization. Furthermore, I have discussed other possible consequences of the acquisition and use of language for the development of human cognition. In particular, I have developed several original ideas about the role of language in improving (or even generating) the following human capacities:

learning, abstraction, memory, attention, problem-solving, voluntary control of action, and mental life.

Notwithstanding its fundamental importance, the role of language as a cognitive tool has been almost neglected both in mainstream cognitive science and in the literature on language evolution. As I have argued, this topic is of the most importance because considering language (also) as a cognitive tool seems to provide the missing link for addressing the high-level cognitive capacities which characterize the human mind within the emerging framework in the cognitive sciences which considers cognition as "environmentally embedded, corporeally embodied, and neurally embrained." (van Gelder, 1999, pag. 244) Again, much more work needs to be done in order to understand the relashionships between language and cognition, both from the empirical and from the computational modeling point of view. I just hope that this thesis has provided both enough arguments for convincing the cognitive science community of the fundamental importance of this topic and some useful ideas for possible directions of future research.

Finally, with respect to the debate on the evolution of language, my major contribution has been to provide substantial arguments to the point that the use of language as a cognitive tool might have played an important role in the evolution of language itself. In fact, most of the advantages provided by talking to oneself do not seem to require a syntactic language, but just the 'symbolic' capacity to associate 'meanings' (inteded as internal representations of significant experiences) with labels. Hence, it is reasonable that the discovery of individual, cognitive uses of language could have happened quite early in language evolution. In particular, before the transition from an holistic proto-language to the full-blow compositional language of modern humans. Trying to sort out what could have been the consequences of this early use of language for oneself in the subsequent evolution of language is still another important topic for future research.

Appendices

A. Neural Networks

A.1. Natural and artificial neural networks

Artificial neural networks can be considered as very simplyfied models of the functioning of the brain, and, more generally, of the nervous system. The structure of the nervous system is quite well known. The basic, fundamental units of the nervous system are *neurons*, which are special types of cells capable of trasmitting elettrical signals. The numeber of neurons in the human brain is about 10^{11}-10^{12}, and each neuron is connected to about 10^{3}-10^{4} other neurons. There are a number of different kinds of neurons, but there is a general structure that underly all of them. This structure can be divided in four parts, namely the **dendrites**, the **soma**, the and the **axon**. The electrical signal emmitted by neurons are called **action potentials** or **spikes** and are constituted by rapid, binary, electical impulses with a duration of about 1ms. They are initiated in a specialized region at the origin of the axon and propagate along it. When the action potential reaches the end of the axon, it triggers the emission of some chemicals, called **neurotransmitters**, which are realised in the space between the axon and the dendrite of another neuron, which is called the **synaptic cleft** (the connection between two neurons are called **synapses**). The neurotrasmitters bind to the receptors of the **post-synaptic** neuron and cause, through a chain of events, either the **depolarization** or the **hyperpolarization** of the membrane of the receiving neuron. A depolarization corresponds to an **excitation** in that it favors the emission of a spike in the post-synaptic neuron, while a hyperpolarization corresponds to an **inhibition** in that it oppose spike emission. Changes in the polarization of the neurons propagate passively from the dendrites to the cell body, where their effects are integrated. If at the origin of the axon the depolarization reaches a certain threshold, an action potential is generated. After the spike, there is a brief refractory period in which the neuron is slightly hyperpolarized and cannot generate another action potential (for a detailed account, see Kandel et al., 1995).

There are several classes of neural models, that simulte neurons and neural networks at any scale and with any level of sophistication: from the detailed models of single neurons that simulate the effects of particular chemicals on ionic channels (which are the mechanisms through which the membran potential changes), to neural networks consisting of several thousand of abstract, idealized neurons (for an overview, see Floreano and Mattiussi, 2002). In the work presented in this thesis I have used probably the most common neural network models, which are very abstract and highly simplified ones,[1] but can be used to model interesting phenomena at the behavioral level.

An artificial neural nework is a collection of artificial **neurons** or **nodes**, linked to each other by **connection weights**, real numbers that correspond to the number and strenght of synapses between two neurons. Connection weights can be either positive, corresponding to **excitatory** synapses or negative, corresponding to **inhibitory** synapses. At anyh given moment the state of a neuron is represented by its **activity** (or **activation**), which correspond to the average firing rate of as real neuron. A neuron's activity is a function of the sum of the excitatory and inhibitory inputs that comes from all other neurons connected to it. The value of each excitation or inhibition arriving to a (post-synaptic) neuron through a connection is in turn calculated by multiplying the activity of the pre-synaptic neuron by the weight of the connection that link the two neurons. In sum, at any given moment, the activation a_i of a given neuron is given by the following formula:

$$a_i = f(\sum w_{ij} a_j)$$

where w_{ij} is the connection weight that connect the j_{th} neuron, with activation a_j, to neuron i, and $f(x)$ is the so called **activation function**, which determines how the neuron reacts to stimulation. There are a number of possible activation functions, some of which are depicted in figure A.1.

Neurons in neural networks are usually clustered in **layers** or **groups**, with two layers being either fully connected to each other (meaning that every neuron of the

---

[1] This simplification is necessary both for pragmatic and for theoretical reasons. From the pragmatic side, the simulation of populations organisms living in and interacting with an environment and evolving over thousand of generation can be very computationally expensive, and the use of more detailed models of neural networks could render the computational time too long. From the theoretical side, suffice it to say that semplification is just a necessary practice in science: the point is to do the *right* simplifications. Since the work presented here is not intented as a contribution to neuroscience but rather to cognitive science, I endorse the standard assumption that what is important from the behavioral point of view is how the network as a whole process information by the reciprocal influences of neurons, rather than the details of these influences. I am quite sure that in the future we will understand that some of these details are in fact important for understanding the behavior of the network as a whole, but this is something I leave to the (computational) neuroscientist to discover. In the meanwhile, the general trend in artificial life research (as in any science in general) is to try and go as far as possible with simple models, which are usually complex enough to work with.

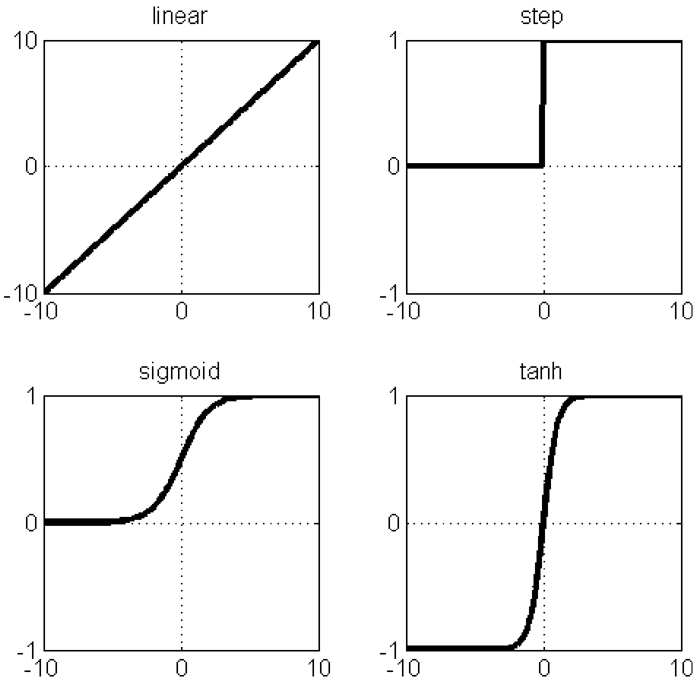


Figure A.1: Some examples of neurons' activation functions. Linear: $y = x$. Step: $y = \frac{segno(x)+1}{2}$. Sigmoid (or logistic): $y = \frac{1}{1+e^{-x}}$. Hyperbolic tangent (tanh): $y = \tanh(x)$.

sending group is connected to every neuron of the receiving group) or not connected at all. The number of groups, the number of neurons in each group and the pattern of connectivity between groups constitutes the **architecture** of a neural network. There are three kinds of neurons (and of groups of neurons): input, hidden and output neurons. Typically, groups of input units are connected to groups of hidden units which are in turn connected to groups of output units. Input units' activation depends on some event *outside* the neural network itself: in a classical connectionist network these activations are given by the researcher, while in a typical artificial life simulation the activation of the input units depend either on something outside the agent (the external environment) or on something in the body of the agent (the internal environment). The activation pattern of the input units, together with the weights that connect the input units to the hidden units, determine the activation pattern of the hidden units. Similarly, the activation pattern of the hidden units, together with the weights that connect the hidden units to the output units, determine the activation pattern of the output units. The output units represent the netwrok's response to the given input: in a classical connectionist network, this is again interpreted by the researcher as some high-level behavior, while in a typical artificial life simulation the output units correspond to the muscles of the agent, which determine agent's movements.

A.2. Learning algorithms

The way a neural network responds to inputs depends on (a) its architecture and (b), the connection weights. If some of the weights of a given neural network are changed, then its behavior will change. A fundamental feature of neural networks is their capacity to **learn**, that is, their capacity to adjust connection weights in such a way that the overall behavior gets better (according to some criterion)[2]. A number of learning algorithms have been developed in the leterature. Here, I will present just three of them, which are probably the most popular ones: the hebbian rule, the delta rule and the back-propagation of error.

In his most-famous book *The Organization of Behavior*, Donald Hebb proposed a possible rule for synaptic modification according to which the strenght of the connection between two units will be increased whenever the two units happen to fire together (Hebb, 1949). In neural network research, this is what is called the Hebb(ian) rule and can be formulated as follows:

$$\Delta w_{ij} = \eta a_i a_j$$

where Δw_{ij} represents the change in the weight that connect neuron j with neuron i, a_i and a_j are the activations of neurons i and j, respectively, and η is the **learning rate**, typically a number in the range $[0;1]$ which determines the rate of change of the connection.

This rule has the problem that if the activations of neurons are positive numbers (which is usually the case), then connections between neurons are bound to increase indefinitely. In order to solve this problem, a number of variations of this basic learning rule have been developed, among which the pre-synaptic rule:

$$\Delta w_{ij} = \eta a_i \left(a_j - \overline{a}_j\right)$$

the post-synaptic rule:

$$\Delta w_{ij} = \eta \left(a_i - \overline{a}_i\right) a_j$$

and the co-variation rule:

$$\Delta w_{ij} = \eta \left(a_i - \overline{a}_i\right) \left(a_j - \overline{a}_j\right)$$

---

[2] The change in connection weights of a neural network correspond to the increase or decrease of the number of synaptic connection (and their efficacy) that happens between two connected neurons of the real brain due to brain activity. These changes are the mechanisms that underlie brain's **plasticity**, that is, the capacity of brains to continually adapt to new circumstances.

where $\overline{a}_i$ and $\overline{a}_j$ represent the mean activations of the units i and j, respectively. The use of those rules guarantees that the weights do not increase indefinetely since they will be decrease every time the activation of the sending or receiving neurons is lower than usual (pre- and post- synaptic rules, respectively), or the differences between the mean and the present activations of the two neurons are of different sign (co-variation rule). Though not as powerful as other learning rules, all those variations of the Hebb rule are very interesting because (a) they are local, that is they work with information which is directly available at the synapsis between two nodes; and (b) the long term potentiation (LTP) and long term depression (LTD) seem to be based on mechanisms that instantiate exactly those rules (Rolls and Treves, 1998).

The delta-rule is a **supervised** learning algorithm that can be applied to **perceptrons**. A perceptron is a neural network composed of a single layer of weights which connect directly the input units with the output units, without any hidden group of units between them. A supervised learning algorithm is one in which the connection weights are adjusted as a function of some measure of the network's performance. In a typical supervised learning algorithm the researcher knows what is the correct network's response to a given input and use this knowledge to calculate an **error measure** between this correct response (also called the **teaching input**) and the actual output of the network. The learning rule is such that it tries to minimize the error, which is typically calculated as the **mean squared error** across all output nodes. The delta-rule is formulated as follows:

$$\Delta w_{ij} = \eta \left(t_i - a_i \right) a_j$$

where t_i is the teaching input for unit i. In this way, the weight between two neurons is increased if the actual response of the output unit is minor than the desired one, it is decreased if the activation of the output unit is greater than the desired one, and is left unchanged if the actual output matches the desired one. Furthermore, absolute value of the change between two units is proportional both to the discrepancy from the actual and desired output and to the activation of the input unit which send the connection, so that weights that contributed more to the error will be changed more.

Nothwithstanding its simplicity, the delta-rule is very powerful, since it can been shown that it guaratees to find a matrix of weights which minimizes the total squared error for a given set of input-output mappings, provided that the learning rate is small enough and the input-output pairs are presented in a random order. The problem is that it is applicable only to perceptrons, which have severe limitations. The absence of hidden layers of neurons prevents in fact perceptrons to perform input-output mappings which are not-linearly separable.[3] The demonstration of this shortcoming

---

[3] A linearly separable mapping is one in which points in the n-dimensional input space which must be

of perceptrons, provided in 1969 by Marving Minsky and Seymour Papert in their famouse *Perceptron* (Minsky and Papert, 1969), together with the lack of a learning rule for **multi-layer networks,** was the cause of the loss of interest for neural networks in the cognitive science community during the 1970s. On the other hand, a multy-layer netowrk can compute any number of arbitrary input-output mappings, provided a sufficient number of hidden units. Consequently, the development of learning algorithms for multy-layer networks has been a major breakthrough in neural network research. The first and most popular of this algorithms is the back-propagation of error. It can be considered as a generalization of the delta-rule in which the error on the output unit is back-propagated the hidden units. In this way, also the weights connecting the input to the hidden nodes can be changed so that global error is decreased.

The back-propagation is a very powerful learning rule but it has some problems, both theoretical and practical. First of all, it is not local, in that the back-propagation of error to the hidden units requires the global knowledge of the activations (indeed of the errors) of all output neurons. This is implausible from the biological point of view, since no mechanism has been found in the real brain that could support the transmission of such global information. Second, the provision of the teaching input to the network for each input pattern is in itself very problematic since it is not clear where such an information could come from in most of real life situations.[4] Finally, when neural networks are used in artificial life research, there is no 'correct output' for any given input. In this kind of research, in fact, what is important are not individual input-output mappings, but the macro-behavior generated by the continuous interactions between the agent and its environment. Consequently, the researcher cannot know what is the best output for an input pattern. In such situations, the connection weights can still be adapted to solve the task at hand by the use of a genetic algorithm (see Appindix B). This is the standard approach in most of artificial life research, and, in particular, of evolutionary robotics (Nolfi and Floreano, 2000).

A.3. Different kinds of networks

The neural networks described so far are just one of the various kind of neural networks used in artificial life research, namely **feed-forward**, **discrete time**, **continuous-activation** neural networks. In feed-forward networks the information flow goes al-

---

responded to in different ways can be separated by an hyperplane (that is, by a plane of dimension $n - 1$).

[4] Indeed, there are some kind of learning situations in which the presence of the teaching input is not so problematic. One example is instructional learning, in which the discent learn and is corrected by a real teacher. Another is learning to predict, in particular, to predict the perceptual consequences of one's own actions: in this case the teaching input is provided by the environment itself, since the consequence of one given action is the actual input pattern received by the network after it has produced that action.

ways in the same direction, that is, from input to output. This makes this networks completely *reactive*, in the sense that the response of the network in any given moment is completely determined by external stimula arriving to the network as input pattern (plus the connection weights, of course). As a consequence, feed-forward networks cannot process information distributed over time, nor produce a sequence of responses given just a single stimulation. This is a very serious limitation because the new 'embodied' cognitive science (see 5.1.2) is realizing that time is something very important for real organisms. Time is important not only because some information collected at time t^x can be useful at some later time $t^y = t^x + \Delta t$, which requires some sort of memory, but also in the collection of information itself: the information which is relevant for an organism's behavior usually almost never present to the organism's senses all at once; rather, it consists of a sequence of perceptual events distributed over time. As a result, an important task for present and future artificial life research is that of studying more complex neural networks which are able to internally process time. Some of those networks are briefly reviewed here.

A class of neural network which can integrate information over time is composed by **recurrent** networks (figure **??**), which are caharacterized by the presence of recurrent connections. Recurrent connections are connections that link either units at the same level (such as in Simple Recurrent Networks or Elman Networks, see Elman, 1990), or units of one higher level with those of a lower one (such as in Jordan nets, see Jordan, 1989).[5]

There are other ways to allow networks to process information over time which are not based on the presence of recurrences in the architecture of the network. One possibility is to directly introduce time in the way the neural network process the information. For example, **Time-Delayed** neural networks (Lang and Waibel, 1990) are networks in which the time at which a given neuron influences another one depends on the delay of the connection that link the two neurons : for example, a three time-steps delay in a connection that links one input neuron to one hidden neuron means that the information regarding the state of that input at time t_0 will reach the hidden neuron only at time t_3.

Another kind of networks which are able to process time information are **dynamic** neural networks (Nolfi and Marocco, 2001), characterized by neurons with the following activation function:

$$a^t = \gamma f\left(i^t\right) + (1-\gamma)\, a^{t-1}$$

where a^t and a^{t-1} are the activations of the neuron at time t and $t-1$, respectively,

---

[5] With 'level' I refer to the number of passes (vector transformations) that has to be done inside the network for reaching a given layer of neurons. In this respect, the 'lowest' level is that of input units while the highest is that of output units.

$f(x)$ is a standard activation function, typically the logistic one, i^t is the net input received by the neuron at time t (that is the weighted sum of the inputs arriving from the neuron's incoming connections) and γ is a parameter of the neuron codified as a real number in the range $[0;1]$ In particular, the parameter γ determines the neuron's dynamics: the lower its value, the less reactive the neuron is to incoming signals and the slower it is in changing its activation. In the extreme case of $\gamma = 0$, the the neuron's activation never changes as a function of incoming inputs and remains always equal to its initial value. On the contrary, for $\gamma = 1$ the neuron is a typical neuron, without any internal dynamics.

Another class of neural networks which are becoming popular in artificial life research are the continuous time recurrent neural networks or **CTRNN**s (Beer, 1995). Those networks combine recurrent connections with a **continuous-time** dynamics, typically implemented in the digital computer thorugh the application of the Euler method for the solution of the following differential equation, that determine neurons' activation changes:

$$\frac{da}{dt} = \frac{1}{\tau}\left(-a + f(i) + gI\right)$$

where a is the activation of the neuron, τ is the neuron's constant of decay, $f(x)$ is the classical activation function (typically the logistic), i is again the net input arriving to the neuron g is the *gain* and I is the intensity of the input perturbation. The reason for the presence of the factor gI is that in a typical CTRNN the input units are treated just as all the other units and their activation is determined by both the input arriving from the other neurons and the external input (represented by I and whose relative influence is determined by the gain factor g).

Finally, there are a number of even more biologically-oriented neural network models, for which the activation of the nodes are not represented as continuous values but, as happens in real neurons, are just yes-or-no inpulses, usually called *spikes*. Though these models are very interesting for their biological plausibility, they are both very difficult to analize and quite computationally expensive, so they are not commonly used in the artificial life community (but see Di Paolo, 2002; Floreano et al., 2005).

A.4. Representations (Space-state semantics)

The functioning of a neural network consists in the transformation of patterns of activation across different groups of units performed by the connection weights of the network. The pattern of activation of the input units is trasformed by the synaptic weights connecting the input units to the hidden units into the pattern of activation of the hidden units, which is in turn transformed by the weights that connect the hidden to the output units into the pattern of activation of the output units. If the pattern of activation in the input units corresponds to the network's perceptual state and the pattern of activation of the output units is the network's response to a given input, the pattern of activation of the hidden units can be considered as the network's **internal representation** of its perceptual state. Mathematically, this internal representation is constituted by a N-dimensional vector, namely, the vector of the activations of the N hidden units. Just as a two-dimensional vector can be conceived as a point in a two-dimensional Cartesian spece, with the two values of the vector corresponding to the two coordinates of the point in the space, an N-dimensional vector can be considered as a point in an N-dimensional *hyper-space*, whith each value of the vector corresponding to the N coordinates of the point in the hyper-space. So, the internal representation of a given input in a neural network is a point in the N-dimensional representational space of the network, with each neuron in the hidden layer constituting one dimension and its activation given the input corresponding to the point's coordinate of that dimension. Since spaces have metrics, we can now calculate *differences* between two internal representations (and, consequently, similarities): the difference between the representations of two input patterns U and V is just the *distance* between the two points representing the inputs in the hidden units space. The formula is the following:

$$d = \sqrt{\sum (u_i - v_i)^2}$$

where d is the distance between the two points (that is the difference between the two representations) and u_i and v_i are the activations of the i_{th} hidden unit given U and V, respectively.

The geometrical interpretation of internal representations also gives a concrete (operational) meaning to the concept of *prototype*. The prototipical representation of a given set of patterns corresponds to the baricentre of the points representing all the patterns: that is, the point whose coordinate, for each dimension, corresponds to the mean of the coordinates for that dimension of all the points representing the input patterns. Consequently, the prototypicity of a given representation can be calculated as the distance between the representation and the prototype itself.

Furthermore, we can conceptualize the process of *categorization* in neural networks in the following way. To categorize for a neural network means to respond to a set of input patterns with a given output pattern and to another set of input patterns with another output pattern.[6] Since the output of a neural network depends on the vector of the activations of the hidden units (together with the weights that connect the hidden to the output units), in order for a neural network to perform the appropriate

---

[6] This is a simplification that holds for classical neural networks; for embodied neural networks it means to produce different *sequences* of output patterns that constitute different *actions*.

categorization it has to *partition* its hidden units space, which is its representational space, in a useful manner. Specifically, this space must be partitioned so that each category is represented in one *region* of the space. In other words, all the input patterns that must be responded to in the same way must be represented as points close the one to the others, while input patterns that must be responeded to in different ways correspond to points distant the one to the others. We can also provide a measure of the *quality* of the categorization. The quality of a categorization depends on (a) how much patterns belonging to one category are represented as similar and (b) how much patterns belonging to different categories are represented as different. We can measure (a) as the average for all categories of the mean distance between each point belonging to one category and the category prototype; a measure for (b) is the average on all couples of categories of the distances between the categories' prototypes. If we want to have one single value as a measure of categorization quality we can either take the mean or the product of the two values (a) and (b), which must be previously normalized in $[0; 1]$.

B. Genetic algorithms

B.1. Natural and artificial selection

Genetic algorithms are a class of computational techniques inspired by the process of evolution by natural selection, which is the main explanatory principle for understanding the complexity of the living world.[1] The basic principle of evolution by natural selection is indeed quite simple. In order for evolution to take place, all you need is a population of individuals with:

1. **heredity** of traits (offsprings resemble parents);
2. **variability** in the traits (the process of eredity is not perfect and each individual is slightly different from the others);
3. scarcity of resources (the environment cannot support all individuals of the population).

If these conditions hold, then some traits will happen to be more **adaptive** to the environment than others, meaning that the individuals who possess those traits will tend to survive more and to produce more offsprings compared to individuals who possess other traits – simply because the environment does not support an indefinite grow of the population (3). As a consequence, given that traits are inheritable (1), more adaptive traits will spread in the population. And provided that heredity is not perfect (2), the constant introduction of new variants will assure that the process will keep on going indefinitely, producing individuals that are more and more adapted to their environment. This simple mix of Chance – that is the constant introduction of

---

[1] Other factors that play a role in evolution are, for example, genetic drift, genetic flow, phisical constrains and catastrophic events. Apart from Darwin's *On the origin of the Species by means of natural selection* (Darwin, 1859) two very good introductions of adaptationist thinking are Dawkins (1976) and Dennett (1995). In particular, Dennett's book discusses and clarifies, among many other things, most of the controversial points of evolutionary thinking in biology.

random variations in the population – and Necessity – that is the selective reproduction of those traits that are more adaptive to the environment (Monod, 1971) – can explain the complexities of the living world without the appell to a purposefull designer, that is God.

In the natural world, both the heritability and the variability of traits depend on organisms' reproduction system. Every cell of an organism contains the organism's DNA, which can be considered as a set of instructions for making a copy of that organism. The functional blocks of DNA, that is the single instructions for the copying of organisms are the **genes**. The collection of all the genetic material of an organism is its **genome**, while the specific genes contained in the genome are the organism's **genotype**. The **phenotype** is the organism's itself, which is constructed during development by following the instructions contained in its genome. The reason why organisms resemble their parents is that they inherit their parent's genes. The major reason of the variability in a population is that the copying process of the parent's genotype is never perfect: errors in the transcription of the genetyc material are called **mutations**. Another important factor of variability is due to the **recombination** of genetic material (or **cross-over**) which happens during **sexual** reproduction: the offspring inherit part of the genes from one parent and part from the other, with the result that the genotype of the new-born is a new combination of those of the parents, different from both. The adaptivity of an organism to its environment is often called its **fitness**, which can be considered as the probability for the orgnaism to survive and reproduce.

B.2. Algorithms

There are a number of computational techniques that are inspired by natural selection, such as evolutionary strategies (Back et al., 1991), evolutionary programming (Fogel et al., 1966) and genetic programming (Koza, 1992). While those techniques were invented in order to solve specific problems, genetic algorithms were developed by John Holland in order to study the process of adaptation by means of its simulation in the computer (Holland, 1975). Genetic algorithms can be implemented in many different ways, but the basic idea is always the same. First of all, you have to define the task that you want to solve by providing a **fitness function**, which will measure the adequateness of candidate solutions to the problem. Then, you have to specify a **genetic representation** of your candidate solutions. In particular, if you use the genetic algorithm for evolving populations of neural networks, as in the simulations presented in this work, you have to decide:

- which parts of the neural network will remain fixed and which will be encode in the genome. Some of the things you can evolve are the network's architecture, the connection weights, nodes' parameters such as activation functions, time-delays or time constants and network's learning algorithms or leraning parameters such as the learning rate or the number of learning epochs the network is to be trained (as in the 'docility' simulation of Chapter 4);

- what is the genetic enconding, that is how the parameters of the network you want to evolve are encoded in the genome. Usually the parameters are represented either as binary strings (binary encoding) or as vectors of real values (real-valued encoding), but there are other possibilities, such as treating the genome as a tree, of which the parameters are the branches (tree encoding);
- what is the genotype-phenotype mapping, that is how the genome (however encoded) produce the actual candidate solution to the problem. This depends strongly on the genetic encoding: in the simplest case you just use the values of the genes (encoded as binary strings or real values) as the network's connection weights; if you want to evolve not only the connection weights but also the network's architecture, you have to develop a more elaborate genotype-phenotype mappings, which can also involve some kind of developmental phase (see Cangelosi et al., 2003).

Once you have defined both the fitness function and the genetic representation, you construct a population of candidate solutions to the problem with random *genes*. The candidate solutions are the *individuals* of the population. They are tested on the problem and their fitness is calculated according to the specified fitness formula.

Selection schemes

After all individuals of one generation have been tested, the **selection scheme** determines which individuals are selected for reproduction according to the individuals' fitnesses. A number of possible selection schemes have been used, among which the most popular are the *roulette wheel*, the *rank* and the *tournament* selection schemes.

In the roulette wheel method one creates a roulette wheel such that for each individual the size of its slice of the roulette wheel is proportional to the individuals fitness and, for each wheel spin, the individual selected for reproduction is the one under the wheel marker. In this way, the probability of reproduction for each individual corresponds to the fitness of the individual divided by the average fitness of the population. This fitness-proportionate selection scheme can have the problem of premature convergence: at the beginning of the simulation the variance in the fitness is usually very high and the fittest individuals will tend to spread in the population very fastly; furthermore, when the population has converged — meaning that all individuals in the population are very similar to each other — all the individuals will tend to have appoximatly the same fitness value, and this will prevent the population to evolve further, since the selection scheme proportionate to fitness will tend to select individuals randomly. A possible solution to this problem is to perform what is called the *windowing* method before applying the roulette wheel selection scheme. Windowing consists in subtracting the fitness value of the worst individual of the population from all the individuals' fitnesses, so that each individual will have a fitness value between 0 and the fitness value of the best individual minus the fitness value of the worst individual. This is done in order to maintain a reasonable selective pressure even when the average fitness reaches a high value and the differences in fitness between individuals are very low.

The rank selection scheme consists in selecting individuals with a probability which is not proportional to their fitness, but which depends on the ranking of the individuals in the population. In the most simple and common rank method, one sorts the individuals according to their fitness and then selects the N best individuals for reproduction. This method avoids the population to converge too quickly by both preventing that the fittest individuals spread too quickly at the beginning of evolution, when the fitness variance is high, and keeping high selection pressure afterwords, when the fitness variance is low.

The tournament method can be described as follows: choose two individual randomly from the population, select the fittest for reproduction, return the two individuals to the population so that they can be choosen again and repeate the procedure untill you have selected the right number of individuals. This procedure produces a selective pressure similar to that produced by the rank method, but is usually less computationally expensive in that it avoids the sorting of the entire population, which can be very time-consuming.

Reproduction

Reproduction can be either sexual or a-sexual. The difference between these two reproduction schemes lies in the fact that the former include the application of cross-over between the genomes of two parents, while the latter is based on the cloning of single genomes. The simplest kind of cross-over is the single-point one (figure B.1a): take two selected individuals; choose randomly one point for dividing the genomes of the two in two parts; generate one offspring by taking the first part of the genome from the first parent and the second part from the second parent and onother by taking the first part from the second parent and second part from the first parent. One problem in the use of single-point cross-over is that it treats different points in the genome differently: in particular, the end-points of the genome strings are treated differently from the central ones in that they always will be exchanged. One solution is to adopt a double-point

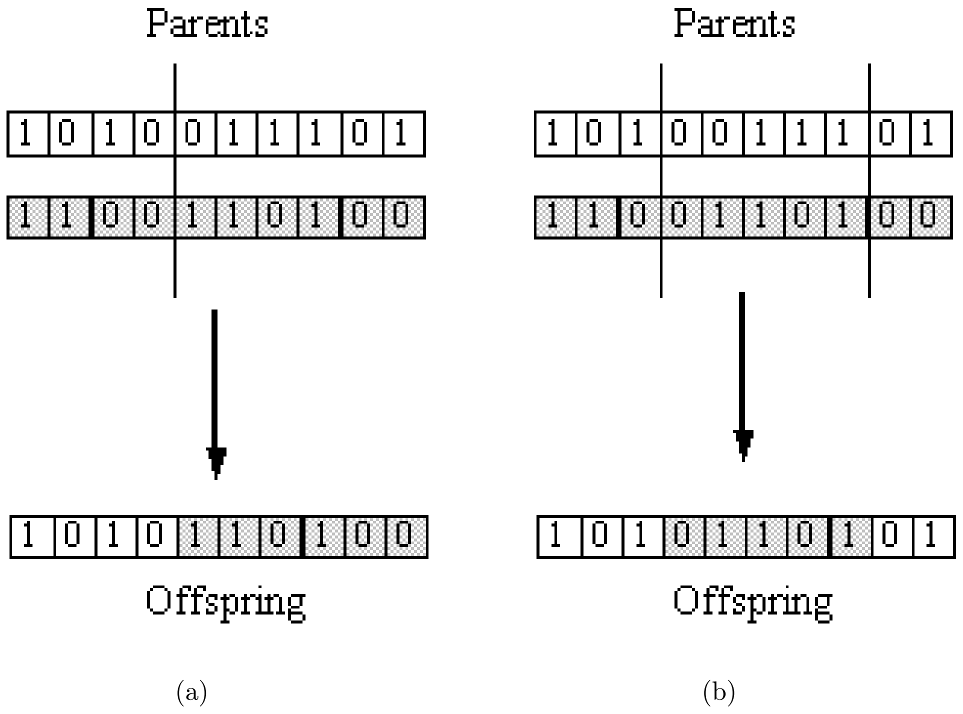


Figure B.1: Examples of single-point (a) and double-points (b) cross-over

cross-over (figure B.1b): two points are randomly selected and the segments which are exchanged are the two between those two points. Applying double-point cross-over is like treating the genome as a circle, so that there is no difference in the probability of cross-over between the centre and the periphery of the genome.

After having produced the right number of genomes by either cloning individuals or by applying cross-over between pairs of parents, mutations are applied to those genomes with a certain probability. There are a various ways of mutating a genes, which depend on the genetic representations one has chosen. If the genetic encoding is binary, mutations consist in changing the binary value; if genes consist of real values, then one can either replace the mutating gene with a randomly chosen value or change the current value by adding to it a random value in a certain range.

Whatever the selection and reproduction schemes chosen, one can prevent that good solutions are lost by not being selected or being distroyed by cross-over and mutation by retaining the best individual (or the best N individuals) and assuring that they are included into the next generation without any modification. This practice is called *elitism* and can greatly improve the effectiveness of the evolutionary search for good solutions.[2]

---

[2] The two best introductions to genetic algorithms are Mitchell (1996), a very good introduction to the field which considers in detail the use of genetic algorithms in scientific modelling, and Goldberg (1989), a more detailed and comprehensive account with an application-oriented perspective.

B.3. The geometrical interpretation of evolution

Genetic algorithms, like any other machine leraning simulation technique, can be considered as a means for finding some solution to a given problem. The algorithm performs a search for the optimal solution in what is called the **search space**. As neural networks' internal representations of input can be conceived as points in the multidimensional state-space defined by the vector of activations of the hidden units (see Appendix A), each candidate solution of a genetic algorithm can be considered as a point in the multi-dimensional search-space defined by the genome, where each dimension corresponds to the value of one gene. The metric of the search space depends on the genetic representation of candidate solutions: if genes are encoded as real values, the difference between two genotypes can be calculated as the distance between the two points that represent them, while if the enconding is binary, the difference can be calculated as the *Hamming distance*, that is the number of genes for which the two genotypes have different bits.

To each point of the search space — that is to each possible genotype — corresponds a fitness value. The **fitness landscape** is the representation of the search space against the fitness of the genotypes that 'inhabit' it. If genotypes have N genes, then the search space will have N dimensions, one for each gene, while the dimensions of the fitness landscape will be $N + 1$, where the last dimension corresponds to the genotypes' fitness (figure B.2). The conformation of fitness landscapes is what determinates the dynamics of the genetic algorithm in that what the genetic algorithm does is to move population along that landscape in particular ways in order to find the highest *peak*, which represent the optimal solution to the problem. The most difficult problem is to avoid getting stuck in local peaks, defined as points in the search space which correspond to a sub-optimal fitness but from which any small movement results in a lower fitness.

As explained in Appendix A, the point whose coordinates correspond to the mean values of a set of points is the geometric baricentre of the set, which can be considered as the prototype of set itself. If we consider the set of points which correspond to all the individuals in a given population, we can measure the genetic variability of the population as the mean distance of all individuals of the population from the populational baricentre. Finally, we can apply the same procedure also to sub-sets of the genome which encode aspects of the phenotype which we are interested in: by comparing the genetic variability of different parts of the genome we can analyse the different selective pressures that act on different aspects of phenotypes.

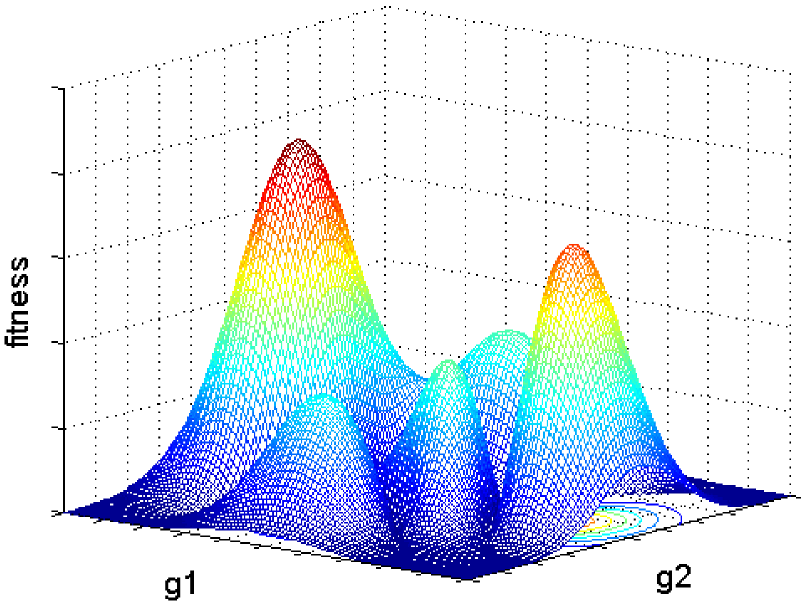


Figure B.2: An example of a fitness landscape for genotypes composed by only two genes (g1 and g2).

Bibliography

Aboitiz, F. and Garcia, R. 1997. The evolutionary origin of the language areas in the human brain: A neuroanatomical perspective. *Brain Research Review*, 25:381–396.

Ackley, D. H. and Littman, M. L. 1994. Altruism in the evolution of communication. In Brooks, R. A. and Maes, P., editors, *Artificial Life IV: Proceedings of the International Workshop on the Synthesis and Simulation of Living Systems*, pages 40–48, Cambridge, MA. MIT Press.

Adami, C., Ofria, C., and Collier, T. C. 2000. Evolution of biological complexity. *Proceedings of the National Academy of Sciences*, 97(9):4463–4468.

Axelrod, R. 1981. The evolution of cooperation. *Science*, 211(4489):1390–1396.

Azmitia, M. 1992. Expertise, private speech, and the development of self-regulation. In Diaz, R. and berk, L. E., editors, *Private speech: From social interaction to self-regualtion*, pages 101–122. Erlbaum, New Jersey, NJ.

Back, T., Hoffmeister, F., and Schwefel, H. P. 1991. A survey of evolutionary strategies. In Belew, R. and Booker, L., editors, *Proceedings of the 4th International Conference on Genetic Algorithms*, pages 2–9, San Francisco, CA. Morgan Kaufmann.

Baddeley, A. 1992. Working memory. *Science*, 255:556–559.

Baddeley, A., Chincotta, D., and Adlam, A. 2001. Working memory and the control of action: Evidence from task switching. *Journal of Experimental Psychology: General*, 130:641–657.

Baldassarre, G., Marocco, D., and Mirolli, M. 2006. La via italiana alla vita artificiale. *Sistemi Intelligenti*, 1:3–6.

Balkenius, C. 1993. The roots of motivations. In Mayer, J.-A., Roitblat, H. L., and Wilson, S. W., editors, *From Animals to Animats II: Proceedings of the Second International Conference on Simulation of Adaptive Behavior*, Cambridge, MA. MIT Press.

Batali, J. 1998. Computational simulations of the emergence of grammar. In Hurford, J. R., Studdert-Kennedy, M., and Knight, C., editors, *Approaches to the evolution of language*, pages 405–426. Cambridge University Press, Cambridge.

Bechtel, W. and Abrahamsen, A. 1991. *Connectionism and the Mind.* Blackwell, Oxford.

Bechtel, W., Abrahamsen, A., and Graham, G. 1998. The life of cognitive science. In Bechtel, W. and Graham, G., editors, *A companion to cognitive science.* Blackwell, Oxford, MA.

Becker, J. T. and Morris, R. G. 1999. Working memory(s). *Brain and Cognition*, 41:1–8.

Bedau, M. A. 1998. Philosophical content and method of artificial life. In Bynam, T. W. and Moor, J. H., editors, *The Digital Phoenix: How Computers are Changing Philosophy*, pages 135–152. Basil Blackwell, Oxford.

Bedau, M. A. 1999. Can unrealistic computer models illuminate theoretical biology? In Wu, A. S., editor, *Proceedings of the 1999 Genetic and Evolutionary Computation Conference Workshop Program*, pages 20–23, San Francisco. Morgan Kaufmann.

Bedau, M. A. 2003. Artificial life: organization, adaptation, and complexity from the bottom up. *Trends in Cognitive Sciences*, 7(11):505–512.

Beer, R. D. 1995. On the dynamics of small continuous-time recurrent neural networks. *Adaptive Behavior*, 3(4):471–511.

Beer, R. D. 2000. Dynamical approaches to cognitive science. *Trends in Cognitive Sciences*, 4(3):91–99.

Behrend, D., Rosengren, K., and Perlmutter, M. 1992. The relation between private speech and parental interactive stile. In Diaz, R. and berk, L. E., editors, *Private speech: From social interaction to self-regualtion*, pages 85–100. Erlbaum, New Jersey, NJ.

Berk, L. E. 1994. Why children talk to themselves. *Scientific American*, pages 78–83.

Berk, L. E. 2001. Trends in human development. In Halonen, J. S. and Davis, S. F., editors, *The many faces of psychological research in the 21st century.* The Society for the Teaching of Psychology, Web site: http://teachpsych.lemoyne.edu/teachpsych/faces/script/index.html.

Berk, L. E. and Garvin, R. 1984. Development of private speech among low-income appalachian children. *Developmental Psychology*, 20:271–286.

Berk, L. E. and Landau, S. 1993. Private speech of learning disabled and normally achieving children in classroom academic and laboratory contexts. *Child Development*, 64:556–571.

Bickerton, D. 1990. *Language and Species.* Chicago University Press, Chicago.

Bickerton, D. 1998. Catastrophic evolution: The case for a single step from protolanguage to full human language. In Hurford, James R.a nd Studdert-Kennedy, M. and Chris, K., editors, *Approaches to the Evolution of Language: Social and Cognitive Bases.* Cambridge University Press, Cambridge.

Bivens, J. and Berk, L. E. 1990. A longitudinal study of the development of elementary school children's private speech. *Merrill-Palmer Quarterly*, 36:443–463.

Block, N. 1978. Troubles with functionalism. In Savage, C., editor, *Perception and Cognition: Issues in the Foundations of Psychology.* University of Minnesota Press, Minneapolis.

Boden, M. A., editor 1996. *The philosophy of artificial life.* Oxford University Press, Oxford.

Bonabeau, E., Dorigo, M., and Theraulaz, G. 1999. *Swarm Intelligence: From natural to artificial Systems.* Oxford University Press, Oxford.

Bonabeau, E. W. and Theraulaz, G. 1994. Why do we need artificial life? *Artificial Life*, 1(3):303–325.

Bowerman, M. and Levinson, S. C., editors 2001. *Language acquisition and conceptual development.* Cambridge University Press, Cambridge.

Boysen, S. T., Bernston, G., Hannan, M., and Cacioppo, J. 1996. Quantity-based inference and symbolic representation in chimpanzees (pan troglodytes). *Journal of Experimental Psychology: Animal Behavior Processes*, 22:76–86.

Breyer, J., Ackermann, J., and McCaskill, J. S. 1998. Evolving reaction-diffusion ecosystems with self-assembling structures in thin films. *Artificial Life*, 4(1):25–40.

Briscoe, T., editor 2002. *Linguistic Evolution through Language Acquisition: Formal and Computational Models.* Cambridge University Press, Cambridge.

Brooks, R. A. 1990. Elephants don't play chess. *Robotics and Autonomous Systems*, 6:3–15.

Brooks, R. A. 1991. Intelligence without representation. *Artificial Intelligence Journal*, 47:139–159.

Bullock, S. 1998. A continuous evolutionary simulation model of the attainability of honest signalling equilibria. In Adami, C., Belew, R., Kitano, H., and Taylor, C., editors, *Proceedings of the Sixth International Conference on Artificial Life*, pages 339–348.

Burtsev, M. S. 2004. Tracking the trajectories of evolution. *Artificial Life*, 10(4):397–411.

Cangelosi, A., editor 2005. *The emergence of language. Special Issue of Connection Science*, volume 17.

Cangelosi, A., Greco, A., and Harnad, S. 2000. From robotic toil to symbolic theft: Grounding transfer from entry-level to higher-level categories. *Connection Science*, 12(2):143–162.

Cangelosi, A. and Harnad, S. 2000. The adaptive advantage of symbolic theft over sensorimotor toil: Grounding language in perceptual categories. *Evolution of Communication*, 4:117–142.

Cangelosi, A., Nolfi, S., and Parisi, D. 2003. Artificial life models of neural development. In Kumar, S. and Bentley, P., editors, *On Growth, Form, and Computers*, pages 339–354. Academic Press, London.

Cangelosi, A. and Parisi, D. 1998. The emergence of a language in an evolving population of neural networks. *Connection Science*, 10(2):83–97.

Cangelosi, A. and Parisi, D., editors 2002. *Simulating the evolution of language.* Springer-Verlag, London.

Cangelosi, A. and Parisi, D. 2004. The processing of verbs and nouns in neural networks: Insights from synthetic brain imaging. *Brain and Language*, 89(2):401–408.

Cangelosi, A., Smith, A. D., and Smith, K., editors 2006. *The Evolution of Language: Proceedings of the 6th International Conference on the Evolution of Language*, Singapore. World Scientific.

Carruthers, P. and Boucher, J., editors 1998. *Language and thought: Interdisciplinary themes.* Cambridge University Press, Cambridge.

Castelfranchi, C. 2006. Robotica o ia, è sempre 'artificiale' o comunque 'cognitiva'. *Sistemi Intelligenti*, 1:15–20.

Cheney, D. L. and Seyfarth, R. M. 1990. *How monkeys see the world: Inside the mind of another species.* University of Chicago Press, Chicago.

Cheng, K. 1986. A purely geometrical module in the rat's spatial representation. *Cognition*, 23:149–178.

Chomsky, N. 1972. *Language and Mind.* Harcourt, Brace and World, New York, NY.

Chomsky, N. 1975. *Reflections on Language.* Pantheon, New York, NY.

Chomsky, N. 1988. *Language and Problems of Knowledge.* MIT Press, Cambridge, MA.

Christiansen, M. H. and Chater, N. 1999. Connectionist natural language processing: The state of the art. *Cognitive Science*, 23:417–437.

Christiansen, M. H. and Kirby, S. 2003. Language evolution: The hardest problem in science? In Christiansen, M. H. and Kirby, S., editors, *Language Evolution: The States of the Art.* Oxford University Press, Oxford.

Churchland, P., Ramachandran, V., and Sejnowski, T. 1994. A critique of pure vision. In Koch, C. and Davis, J. L., editors, *Large scale neuronal theories of the brain*, pages 23–60. MIT Press, Cambridge, MA.

Churchland, P. M. 1981. Eliminative materialism and the propositional attitudes. *Journal of Philosophy*, 78(2):67–90.

Clark, A. 1989. *Microcognition: philosophy, cognitive science, and parallel distributed processing.* MIT Press, Cambridge, MA.

Clark, A. 1997. *Being There: putting brain, body and world together again.* Oxford University Press, Oxford.

Clark, A. 1998. Magic words: How language augments human computation. In Carruthers, P. and Boucher, J., editors, *Language and thought: Interdisciplinary themes*, pages 162–183. Cambridge University Press, Cambridge.

Clark, A. 2001. *Mindware: an introduction to the philosophy of cognitive science.* Oxford University Press, Oxford.

Clark, A. 2005. Beyond the flesh: some lessons from a mole cricket. *Artificial Life*, 11(1-2):233–44.

Clark, A. and Grush, R. 1999. Towards a cognitive robotics. *Adaptive Behavior*, 7(1):5–16.

Cliff, D. T., Harvey, I., and Husbands, P. 1993. Explorations in evolutionary robotics. *Adaptive Behaviour*, 2:73–110.

Darwin, C. 1859. *On the Origin of Species.* John Murray, London.

Dawkins, R. 1976. *The selfish gene.* Oxford University Press, Oxford.

Dawkins, R. 1986. *The Blind Watchmaker: Why the Evidence of Evolution Reveals a Universe Without Design.* Norton, New York, NY.

Dawkins, R. and Krebs, J. R. 1978. Animal signals: information or manipulation? In Krebs, J. R. and Davies, N. B., editors, *Behavioural Ecology: An Evolutionary Approach*, pages 282–309. Blackwell Scientific Publications, Oxford.

Dehaene, S. 1997. *The number sense: How the mind creates mathematics.* Oxford University Press, Oxford.

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., and Tsivkin, S. 1999. Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science*, 284:970–974.

Denaro, D. and Parisi, D. 1997. Cultural evolution in a population of neural networks. In Marinaro, M. and Tagliaferri, R., editors, *Neural Nets. Proceedings of Wirn-96*, pages 100–111, New York, NY. Springer Verlag.

Dennett, D. C. 1978. *Brainstorms.* Bradford Books, Montgomery, Vt.

Dennett, D. C. 1987. Cognitive wheels: The frame problem in artificial intelligence. In Pylyshyn, Z. W., editor, *The Robot's Dilemma: The Frame Problem in Artificial Intelligence.* Ablex, Norwood, NJ.

Dennett, D. C. 1991. *Consciousness Explained.* Little Brown & Co., New York, NY.

Dennett, D. C. 1993. Learning and labeling. *Mind and Language*, 8(4):540–547.

Dennett, D. C. 1994. Artificial life as philosophy. *Artificial Life*, 1(1):291–292.

Dennett, D. C. 1995. *Darwin's Dangerous Idea: Evolution and the Meanings of Life.* Simon and Schuster, New York, NY.

Dessalles, J.-L. 2000. Language and hominid politics. In Knight, C., Hurford, J. R., and Studdert-Kennedy, M., editors, *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, pages 62–79. Cambridge University Press, Cambridge.

Di Ferdinando, A. and Parisi, D. 2004. Internal representations of sensory input reflect the motor output with which organisms respond to the input. In Carsetti, A., editor, *Seeing, thinking and knowing*, pages 115–141. Kluwer, Dordrecht.

Di Paolo, E. A. 1999. A little more than kind and less than kin: the unwarranted use of kin selection in spatial models of communication. In Floreano, D., Nicoud, J., and Mondada, F., editors, *Advances in Artificial life. Proceedings of the 5th European Conference on Artificial Life*, pages 504–513, Lausanne. Springer-Verlag.

Di Paolo, E. A. 2002. Spike timing dependent plasticity for evolved robots. *Adaptive Behavior*, 10:243–263.

Di Paolo, E. A., Noble, J., and Bullock, S. 2000. Simulation models as opaque thought experiments. In Bedau, M., McCaskill, J., Packard, N., and Rasmussen, S., editors, *Artificial Life VII*, pages 497–506, Cambridge, MA. MIT Press.

Diaz, R. and Berk, L. E., editors 1992. *Private speech: From social interaction to self regulation.* Erlbaum, New Jersey, NJ.

Dobzhansky, T. 1973. Nothing in biology makes sense except in the light of evolution. *American Biology Teacher*, 35:125–129.

Duncan, R. and Pratt, M. 1997. Microgenetic change in the quantity and quality of preschoolers' private speech. *International Journal of Behavioral development*, 20:367–383.

Eldredge, N. and Gould, S. J. 1972. Punctuated equilibria: an alternative to phyletic gradualism. In Schopf, T. J. M., editor, *Models in Paleobiology*, pages 82–115. Freeman Cooper, San Francisco.

Elman, J. L. 1990. Finding structure in time. *Cognitive Science*, 14:179–211.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., and Plunkett, K. 1996. *Rethinking Innateness: A Connectionist Perspective on Development.* MIT Press, Cambridge, MA.

Emerson, M. J. and Miyake, A. 2003. The role of inner speech in task switching: A dual-task investigation. *Journal of Memory and Language*, 48:148–168.

Fitch, W. T. 2004. Kin selection and 'mother tongues': A neglected component in language evolution. In Oller, K. and Griebel, U., editors, *Evolution of Communication Systems: A Comparative Approach*, pages 275–296. MIT Press, Cambridge, MA.

Fitch, W. T. 2005. The evolution of language: A comparative review. *Biology and Philosophy*, 20(2-3):193–203.

Floreano, D. and Mattiussi, C. 2002. *Manuale sulle Reti Neurali.* Il Mulino, Bologna.

Floreano, D., Zufferey, J., and Nicoud, J. 2005. From wheels to wings with evolutionary spiking neurons. *Artificial Life*, 11(1–2):121–138.

Fodor, J. 1975. *The language of thought.* Harvard University Press, Cambridge, MA.

Fogel, L., Owens, A., and Walsh, M. 1966. *Artificial Intelligence through Simulated Evolution.* John Wiley, New York, NY.

Frauenglass, M. and Diaz, R. 1985. Self-regulatory functions of children's private speech: A critical analysis of recent challanges to vygotsky's theory. *Developmental Psychology*, 21:706–708.

Frisch, K. v. 1967. *The Dance Language and Orientation of Bees.* Harvard University Press, Cambridge, Mass.

Furusawa, C. and Kaneko, K. 2002. Origin of multicellular organisms as an inevitable consequence of dynamical systems. *The Anatomical Record*, 268(3):327–342.

Gadanho, S. C. and Hallam, J. 2001. Robot learning driven by emotions. *Adaptive Behavior*, 9(1):42–64.

Gallese, V. and Lakoff, G. 2005. The brain's concepts: The role of the sensory-motor system in reason and language. *Cognitive Neuropsychology*, 22:455–479.

Gell-Mann, M. 1995. What is complexity? *Complexity*, 1(1):16–19.

Gentner, D. 2003. Why we are so smart. In Gentner, D. and Goldin-Meadow, S., editors, *Language in mind*, pages 195–235. MIT Press, Cambridge, MA.

Gentner, D. and Goldin-Meadow, S., editors 2003a. *Language in mind.* MIT Press, Cambridge, MA.

Gentner, D. and Goldin-Meadow, S. 2003b. Whither whorf. In Gentner, D. and Goldin-Meadow, S., editors, *Language in mind*, pages 3–14. MIT Press, Cambridge, MA.

Gleitman, L. and Papafragou, A. 2005. Language and thought. In Holyoak, K. and Morrison, R., editors, *Cambridge Handbook of Thinking and Reasoning.* Cambridge University Press, Cambridge.

Goldberg, D. 1989. *Genetic Algorithms in Search, Optimization and Machine Learning.* Addison-Wesley, Reading, MA.

Gould, S. J. and Eldredge, N. 1993. Punctuated equilibrium comes of age. *Nature*, 366(6452):223–227.

Gould, S. J. and Lewontin, R. 1979. The spandrels of san marco and the panglossion paradigm: a critique of the adaptationist program. *Proceedings of the Royal Society of London B*, 205(1161):581–598.

Gould, S. J. and Vrba, E. S. 1982. Exaptation: A missing term in the science of form. *Paleobiology*, 8:4–15.

Grafen, A. 1990. Biological signals as handicaps. *Journal of Theoretical Biology*, 144:517–546.

Gruber, O. 2002. The co-evolution of language and working memory capacity in the human brain. In Stamenov, M. and Gallese, V., editors, *Mirror neurons and the evolution of brain and language*, pages 77–86. Benjamins, Amsterdam.

Gruber, O. and Goschke, T. 2004. Executive control emerging from dynamic interactions between brain systems mediating language, working memory and attentional processes. *Acta Psychologica*, 115:105–121.

Hamilton, W. D. 1964. Genetic evolution of social behavior. *Journal of Theoretical Biology*, 7(1):1–52.

Harnad, S., Hanson, S., and Lubin, J. 1995. Learned categorical perception in neural nets: implications for symbol grounding. In Honavar, V. and Uhr, L., editors, *Symbol processors and connectionist network models in artificial intelligence and cognitive modelling: steps toward principled integration*, pages 191–206. Academic Press, New York, NY.

Haugeland, J. 1985. *Artificial Intelligence: The Very Idea.* MIT Press, Cambridge, MA.

Hauser, M. D. 1996. *The Evolution of Communication.* MIT Press, Cambridge, MA.

Hauser, M. D., Chomsky, N., and Fitch, W. T. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298:1569–1579.

Hauser, M. D. and Fitch, W. T. 2003. What are the uniquely human components of the language faculty? In Christiansen, M. H. and Kirby, S., editors, *Language Evolution: The States of the Art.* Oxford University Press.

Hebb, D. O. 1949. *The Organization of Behavior.* John Wiley, New York, NY.

Hermer, L. and Spelke, E. S. 1994. A geometric process for spatial reorientation in children. *Nature*, 370:57–59.

Hermer-Vazquez, L., Moffett, A., and Munkholm, P. 2001. Language, space, and the development of cognitive flexibility in humans: The case of two spatial memory tasks. *Cognition*, 79:263–299.

Hermer-Vazquez, L., Spelke, E. S., and Kasnelson, A. 1999. Sources of flexibility in human cognition: Dual-task studies of space and language. *Cognitive Psychology*, 39:3–36.

Hihara, S., Yamada, H., Iriki, A., and Okanoya, K. 2003. Spontaneous vocal differentiation of coo-calls for tools and food in japanese monkeys. ; :. *Neuroscience Research*, 45:383–389.

Hockett, C. F. 1960. The origin of speech. *Scientific American*, 203:88–96.

Holland, J. H. 1975. *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor.

Holland, J. H. 1995. *Hidden Order: How Adaptation Builds Complexity*. Perseus Books, Reading, MA.

Hunt, E. and Agnoli, F. 1991. The whorfian hypothesis: A cognitive psychology perspective. *Psychological Review*, 98:377–389.

Hurford, J. R. 1999. Expression-induction models of language evolution: Dimensions and issues. In Briscoe, T., editor, *Linguistic evolution through language acquisition*, pages 301–344. Cambridge University Press, Cambridge.

Hurford, J. R. 2002. The roles of expression and representation in language evolution. In Wray, A., editor, *The Transition to Language*. Oxford University Press, Oxford.

Hutchins, E. 1995. *Cognition in the Wild*. MIT Press, Cambridge, MA.

Hutchins, E. and Hazlehurst, B. 1995. How to invent a lexicon: the development of shared symbols in interaction. In Gilbert, N. and Conte, R., editors, *Artificial Societies: the computer simulation of social life*, pages 157–189. UCL Press, London.

Jackendoff, R. 1996. How language help us think. *Pragmatics and Cognition*, 4:1–34.

Jordan, M. I. 1989. Serial order: A parallel, distributed processing approach. In Elman, J. L. and Rumelhart, D. E., editors, *Advances in Connectionist Theory*. Erlbaum, Hillsdale, NJ.

Kandel, E. R., Schwartz, J. H., and Jessell, T. M., editors 1995. *Essentials of Neural Science and Behavior*. Appleton and Lange, Norwalk.

Kauffman, S. A. 1993. *The origins of order: Self-organization and selection in evolution.* Oxford University Press, Oxford.

Kelso, J. 1995. *Dynamic Patterns.* MIT Press, Cambridge, MA.

Kirby, S. 2000. Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners. In Knight, C., Hurford, J. R., and Studdert-Kennedy, M., editors, *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, pages 303–323. Cambridge University Press, Cambridge.

Kirby, S. 2002. Natural language from artificial life. *Artificial Life*, 8(2):185–215.

Kit, C. 2002. How does lexical acquisition begin? a cognitive perspective. *Cognitive Science*, 1:1–50.

Koza, J. R. 1992. *Genetic Programming: On the Programming of Computers by Means of Natural Selection.* MIT Press, Cambridge, MA.

Krebs, J. R. and Dawkins, R. 1984. Animal signals: Mind-reading and manipulation. In Krebs, J. R. and Davies, N. B., editors, *Behavioral Ecology: An Evolutionary Approach*, chapter 15, pages 380–402. Blackwell Scientific Publications, Oxford, second edition edition.

Kuhn, T. S. 1977. A function for thought experiments. In Kuhn, T. S., editor, *The Essential Tension*, pages 240–265. University of Chicago Press, Chicago.

Kvasnicka, V. and Pospichal, J. 1999. An emergence of coordinated communication in populations of agents. *Artificial Life*, 5:319–342.

Lang, K. and Waibel, A.and Hinton, G. E. 1990. A time-delay neural network architecture for isolated word recognition. *Neural Networks*, 3:23–43.

Langton, C. G. 1989a. Artificial life. In *Artificial Life*, pages 1–47, Redwood City, CA. Addison-Wesley.

Langton, C. G., editor 1989b. *Artificial Life: The Proceedings of an Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems*, Redwood City, CA. Addison-Wesley.

Langton, C. G. 1990. Computation at the edge of chaos. *Physica D*, 42:12–37.

Langton, C. G., editor 1995. *Artificial Life: An overview.* MIT Press, Cambridge, MA.

Lenaerts, T., Chu, D., and Watson, R. 2005. Dynamical hierarchies. *Artificial Life*, 11(4):403–405.

Lenski, R. E., Ofria, C., Pennock, R. T., and Adami, C. 2003. The evolutionary origin of complex features. *Nature*, 423:139–144.

Levinson, S. C., Kita, S., and Haun, D. 2002. Returning the tables: language affects spatial reasoning. *Cognition*, 84:155–188.

Levy, S. 1992. *Artificial Life: The quest for a new creation.* Jonathan Cape, London.

Li, P. and Gleitman, L. 2002. Turning the tables: Language and spatial reasoning. *Cognition*, 83:265–294.

Loewenstein, J. and Gentner, D. 2005. Relational language and the development of relational mapping. *Cognitive Psychology*, 50:315–353.

Lupyan, G. 2005. Carving nature at its joints and carving joints into nature: How labels augment category representations. In Cangelosi, A. and Bugmann, G.a nd Borisyuk, R., editors, *Modelling Language, Cognition and Action: Proceedings of the 9th Neural Computation and Psychology Workshop*, pages 87–96, Singapore. World Scientific.

Lupyan, G. 2006. Labels facilitate learning of novel categories. In Cangelosi, A., Smith, A., and Smith, K., editors, *The Evolution of Language: Proceedings of the 6th International Conference on the Evolution of Language*, pages 190–197. World Scientific Publishing.

MacWhinney, B. 1998. Models of the emergence of language. *Annual Review of Psychology*, 49:199–227.

Majid, A., Bowerman, M., Kita, S., Haun, D. B., and Levinson, S. C. 2004. Can language restructure cognition? the case for space. *Trends in Cognitive Sciences*, 8(3):108–114.

Maki, H. 2002. Origins of spontaneous mutations: specificity and directionality of base-substitution, frameshift, and sequence-substitution mutageneses. *Annual Review of Genetics*, 36:279–303.

Marocco, D., Cangelosi, A., and Nolfi, S. 2003. The emergence of communication in evolutionary robots. *Philosophical Transactions to the Royal Society A: Mathematical, Physical and Engineering Sciences*, 361(1811):2397–2421.

Maynard Smith, J. and Harper, D. 2003. *Animal Signals.* Oxford University Press.

Maynard-Smith, J. and Szathmary, E. 1995. *The Major Transitions in Evolution.* W.H.Freeman Press, New York.

Minsky, M. and Papert, S. 1969. *Perceptrons: An introduction to Computational Geometry.* MIT Press, Cambridge, MA.

Mirolli, M. 2006. Vita artificiale e linguaggio. *Sistemi Intelligenti*, 1:133–142.

Mirolli, M. and Parisi, D. 2003. Artificial organisms that sleep. In Banzhaf, W., Christaller, T., Dittrich, P., Kim, J. T., and Ziegler, J., editors, *Advances in Artificial Life. Proceedings of the 7th European Conference on Artificial Life*, pages 377–386, Berlin. Springer Verlag.

Mirolli, M. and Parisi, D. 2004. Language, altruism and docility: How cultural learning can favour language evolution. In Pollack, J. B., Bedau, M., Husbands, P., Ikegami, T., and Watson, R. A., editors, *Artificial Life IX: Proceedings of the Ninth International Conference on the Simulation and Synthesis of Living Systems*, pages 182–187, Cambridge, MA. MIT Press.

Mirolli, M. and Parisi, D. 2005a. How can we explain the emergence of a language which benefits the hearer but not the speaker? *Connection Science*, 17(3-4):325–341.

Mirolli, M. and Parisi, D. 2005b. Language as an aid to categorization: A neural network model of early language acquisition. In Cangelosi, A., Bugmann, G., and Borisyuk, R., editors, *Modelling language, cognition and action: Proceedings of the 9th Neural Computation and Psychology Workshop*, pages 97–106, Singapore. World Scientific.

Mirolli, M. and Parisi, D. 2006. Talking to oneself as a selective pressure for the emergence of language. In Cangelosi, A., Smith, A., and Smith, K., editors, *The Evolution of Language: Proceedings of the 6th International Conference on the Evolution of Language*, pages 214–221. World Scientific Publishing.

Mitchell, M. 1996. *Introduction to Genetic Algorithms.* MIT Press, Cambridge, MA.

Miyake, A., Emerson, M. J., Padilla, F., and Ahn, J. 2004. Inner speech as a retrieval aid for task goals: The effects of cue type and articulatory suppression in the random task cuing paradigm. *Acta Psychologica*, 115:123–142.

Mondada, F., Pettinaro, G. C., Guigrard, A., Kwee, I. W., Floreano, D., Denebourg, J.-L., Nolfi, S., Gambardella, L. M., and Dorigo, M. 2004. Swarm-bot: A new distributed robotic concept. *Autonomous Robots*, 17(2–3):193–221.

Monod, J. 1971. *Chance and necessity: an essay on the natural philosophy of modern biology.* Alfred A Knopf, New York.

Munroe, S. and Cangelosi, A. 2002. Learning and the evolution of language: The role of cultural variation and learning costs in the baldwin effect. *Artificial Life*, 8(4):311–339.

Nazzi, T. and Gopnik, A. 2001. Linguistic and cognitive abilities in infancy: When does language become a tool for categorization? *Cognition*, 80:303–312.

Newell, A. and Simon, H. 1976. Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, 19:113–126.

Nicolis, G. and Prigogine, I. 1985. *Exploring Complexity.* Piper, Munchen.

Noble, J., Di Paolo, E. A., and Bullock, S. 2001. Adaptive factors in the evolution of signalling systems. In Cangelosi, A. and Parisi, D., editors, *Simulating the Evolution of Language*, pages 53–78. Springer, Heidelberg.

Nolfi, S. and Floreano, D. 2000. *Evolutionary robotics. The biology, intelligence, and technology of self-organizing machines.* MIT Press, Cambridge, MA.

Nolfi, S. and Floreano, D. 2002. Synthesis of autonomous robots through evolution. *Trends in Cognitive Science*, 6(1):31–36.

Nolfi, S. and Marocco, D. 2001. Evolving robots able to integrate sensory-motor information over time. *Theory in Biosciences*, 120:1–24.

O'Brien, G. and Opie, J. 2002. Radical connectionism: Thinking with (not in) language. *Language and Communication*, 22:313–329.

Oden, D., Thompson, R., and Premack, D. 2001. Can an ape reason analogically? comprehension and production of analogical problems by sarah, a chimpanzee (pan troglodytes). In Gentner, D., Holyoak, K., and Kokinov, B., editors, *The analogical mind: Perspectives from cognitive science*, pages 471–498. MIT Press, Cambridge, MA.

Oliphant, M. 1996. The dilemma of saussurean communication. *Biosystems*, 37(1-2):31–38.

Ono, N. and Ikegami, T. 2003. Selection of catalysts through cellular reproduction. In Standish, R. K., Bedau, M. A., and Abbass, H. A., editors, *Artificial Life VIII*, Cambridge, MA. MIT Press.

Parisi, D. 1997. An artificial life approach to language. *Brain and Language*, 59:121–146.

Parisi, D. 1999. *Mente: I nuovi modelli della vita artificiale.* Il Mulino, Bologna.

Parisi, D. 2001a. L'intero elefante. *Sistemi Intelligenti*, 3:485–508.

Parisi, D. 2001b. *Simulazioni: La realtà rifatta nel computer*. Il Mulino.

Parisi, D. 2004. Internal robotics. *Connection Science*, 16(4):325–338.

Parisi, D. 2005a. Dodici differenze tra l'intelligenza artificiale e la vita artificiale. *Sistemi Intelligenti*, 1:155–157.

Parisi, D. 2005b. Mental robotics. Paper presented at the Conference on Artificial Consciousness, Agrigento, Italy, November 2005.

Parisi, D. 2006. Robot futuri. *Sistemi Intelligenti*, 1:7–14.

Parisi, D., Cecconi, F., and Nolfi, S. 1990. Econets: Neural networks that learn in an environment. *Network*, 1:149–168.

Parisi, D. and Mirolli, M. 2006. The emergence of language: How to simulate it. In Lyon, C., Nehaniv, C., and Cangelosi, A., editors, *Emergence of Communication and Language*. Springer Verlag, Berlin.

Pattee, H. H. 1989. Simulations, realizations, and theories of life. In Langton, C., editor, *Artificial life*, pages 63–77, Redwood City, CA. Addison-Wesley.

Penrose, R. 1989. *The emperor's new mind: Concerning computers, minds, and the laws of physics*. Oxford University Press, Oxford.

Petrides, M. E., Alivisatos, B., Evans, A. C., and Meyer, E. 1993. Dissociation of human mid-dorsolateral from posterior dorsolateral frontal cortex in memory processing. *Proceedings of the National Academy of Sciences*, 90:873–877.

Pfeifer, R. and Scheier, C. 1999. *Understanding intelligence*. MIT Press, Cambridge, MA.

Piaget, J. 1972. *The Principles of Genetic Epistemology*. Basic Books, New York, NY.

Piattelli-Palmarini, M. 1989. Evolution, selection and cognition: From 'learning' to parameter setting in biology and in the study of language. *Cognition*, 31:1–44.

Pinker, S. 1994. *The Language Instinct: How the Mind Creates Language*. William Morrow, New York, NY.

Pinker, S. 2003. Language as an adaptation to the cognitive niche. In Christiansen, M. H. and Kirby, S., editors, *Language Evolution: The States of the Art*, pages 16–37. Oxford University Press.

Pinker, S. and Bloom, P. 1990. Natural language and natural selection. *Behavioral and brain sciences*, 13:707–784.

Plunkett, K., Sinha, C. G., Moller, M., and Strandsby, O. 1992. Symbol grounding or the emergence of symbols? vocabulary growth in children and a connectionist net. *Connection Science*, 4:293–312.

Port, R. F. and van Gelder, T., editors 1995. *Mind as Motion.* MIT Press, Cambridge, MA.

Putnam, H. 1963. Brains and behavior. In Butler, R., editor, *Analytical Philosophy.* Basil Blackwell, Oxford.

Putnam, H. 1975. The meaning of 'meaning'. In *Philosophical Papers Vol. 2: Mind, Language and Reality.* Cambridge University Press, Cambridge.

Ramirez, J. 1992. The functional differentiation of social and private speech: A dialogic approach. In Diaz, R. and berk, L. E., editors, *Private speech: From social interaction to self-regualtion*, pages 199–214. Erlbaum, New Jersey, NJ.

Rasmussen, S., Chen, L., Nilsson, M., and Abe, S. 2003. Bridging nonliving and living matter. *Artificial Life*, 9(3):269–316.

Rattermann, M. J. and Gentner, D. 1998. The effect of language on similarity: The use of relational labels improves young children's performance in a mapping task. In Holyoak, K., Gentner, D., and Kokinov, B., editors, *Advances in analogy research: Integration of theory & data from the cognitive, computational, and neural sciences*, pages 274–282. New Bulgarian University, Sophia.

Rolls, E. T. and Treves, A. 1998. *Neural Networks and Brain Function.* Oxford University Press, Oxford.

Rosch, E. 1973. Natural categories. *Cognitive Psychology*, 4:328–350.

Rumelhart, D., McClelland, J., and the PDP Research Group 1986. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, volume 1 & 2. MIT Press, Cambridge, MA.

Ryan, Michael J., . R.-A. S. 1993. Sexual selection and signal evolution: the ghost of biases past. *Philosophical Transactions of the Royal Society of London: Biological Sciences*, 340:187–195.

Ryan, M. J. 1990. Sexual selection, sensory systems and sensory exploitation. *Oxford Surveys in Evolutionary Biology*, 7:157–195.

Schyns, P. G. 1991. A modular neural network model of concept acquisition. *Cognitive Science*, 15(4):461–508.

Searle, J. R. 1980. Minds, brains and programs. *Behavioral and Brain Sciences*, 3:417–458.

Seeley, T. D. 1992. The tremble dance of the honey bee: message and meanings. *Behavioral Ecology and Sociobiology*, 31:375–383.

Simon, H. A. 1990. A mechanism for social selection and successful altruism. *Science*, 250:1665–1668.

Sipper, M. 1998. Fifty years of research on self-replication: An overview. *Artificial Life*, 4(3):237–257.

Smith, K. 2002. Natural selection and cultural selection in the evolution of communication. *Adaptive Behavior*, 10:25–44.

Smith, L. B. and Thelen, E., editors 1993. *A Dynamic Systems Approach to Development.* MIT Press, Cambridge, MA.

Smolensky, P. 1988. On the proper treatment of connectionism. *Behavioral and Brain Sciences*, 11:1–23.

Sober, E. 1992. Learning from functionalism: Prospects for strong artificial life. In Langton, C. G., Taylor, C., Farmer, J. D., and Rasmussen, S., editors, *Artificial Life II*, pages 749–765. Addison Wesley.

Spelke, Elizabeth, S. 2003. What makes us smart? core knowledge and natural language. In Gentner, D. and Goldin-Meadow, S., editors, *Language in mind*, pages 277–311. MIT Press, Cambridge, MA.

Steels, L. 1994. The artificial life roots of artificial intelligence. *Artificial Life*, 1(1):89–125.

Steels, L. 2003. Language-reentrance and the Ínner voice'. *Journal of Consciousness Studies*, 10(4-5):173–185.

Steels, L. 2005. The emergence and evolution of linguistic structure: from lexical to grammatical communication systems. *Connection Science*, 17(3–4):213–230.

Steels, L. and Belpaeme, T. 2005. Coordinating perceptually grounded categories through language: A case study for colour. *Behavioral and Brain Sciences*, 28(4):469–529.

Steels, L. and Brooks, R., editors 1994. *The artificial life route to artificial intelligence: Building Situated Embodied Agents.* Lawrence Erlbaum Ass., New Haven.

Steels, L. and Kaplan, F. 2002. Bootstrapping grounded word semantics. In Briscoe, T., editor, *Linguistic evolution through language acquisition: Formal and computational models*, pages 53–73. Cambridge University Press, Cambridge.

Struhsaker, T. T. 1967. Auditory communication among vervet monkeys (cercopithecus aethiops). In Altmann, S. A., editor, *Social Communication among Primates.* University of Chicago Press, Chicago.

Sugita, Y. and Tani, J. 2005. Learning semantic combinatoriality from the interaction between linguistic and behavioral processes. *Adaptive Behavior*, 13(1):33–52.

Szathmary, E. and Maynard Smith, J. 1995. The major evolutionary transitions. *Nature*, 374:227–231.

Szostak, J. W., Bartel, D. P., and Luisi, P. 2001. Synthesizing life. *Nature*, 409:387–390.

Takagi, H., Kaneko, K., and Yomo, T. 2000. Evolution of genetic codes through isologous diversification of cellular states. *Artificial Life*, 6:283–306.

Taylor, C. and Jefferson, D. 1994. Artificial life as a tool for biological inquiry. *Artificial Life*, 1(1–2):1–13.

Tesfatsion, L. 2002. Agent-based computational economics: Growing economies from the bottom up. *Artificial Life*, 8(1):55–82.

Thelen, E., Schöner, G., Scheier, C., and Smith, L. B. 2001. The dynamics of embodiment: A field theory of infant perseverative reaching. *Behavioral and Brain Sciences*, 24:1–34.

Thompson, D. W. 1917. *On Growth and Form.* Cambridge University Press, Cambridge.

Thompson, R. K. R., Oden, D. L., and Boysen, S. T. 1997. Language-naive chimpanzees (pan troglodytes) judge relations between relations in a conceptual matching-to-sample task. *Journal of Experimental Psychology: Animal Behavior Processses*, 23:31–43.

Tinbergen, N. 1952. Derived activities: Their causation, biological significance, origin and emancipation during evolution. *Quarterly Review of Biology*, 27:1–32.

Tinbergen, N. 1963. On aims and methods of ethology. *Zeitschrift fur Tierpsychologie*, 20:410–433.

Trivers, R. L. 1971. The evolution of reciprocal altruism. *Quarterly Review of Biology*, 46:35–57.

van Gelder, T. J. 1998. The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences*, 21:1–14.

van Gelder, T. J. 1999. Dynamic approaches to cognition. In Wilson, R. and Keil, F., editors, *The MIT Encyclopedia of Cognitive Sciences*, pages 243–245. MIT Press, Cambridge MA.

Vogt, P., editor 2005. *Language Acquisition and Evolution. Special Issue of Adaptive Behavior*, volume 13.

Vygotsky, L. S. 1962. *Thought and language.* MIT Press, Cambridge, MA.

Vygotsky, L. S. 1978. *Mind in society.* Harvard University Press, Cambridge, MA.

Vygotsky, L. S. and Luria, A. 1994. Tool and symbol in child development. In van der Veer, R. and Valsiner, J., editors, *The Vygotsky Reader*, pages 99–174. Blackwell Publishers, Oxford.

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

Waxman, S. and Markow, D. 1995. Words as invitations to form categories: Evidence from 12 to 13-month-old infants. *Cognitive Psychology*, 29(3):257–302.

Waxman, S. R. 2004. Everything had a name, and each name gave birth to a new thought: Links between early word-learning and conceptual organization. In Hall, D. G. and Waxman, S. R., editors, *From many strands: Weaving a lexicon*, pages 295–335. MIT Press, Cambridge, MA.

Welder, A. N. and Graham, S. A. 2001. The influence of shape similarity and shared labels on infants' inductive inferences about nonobvious object properties. *Child Development*, 72:1653–1673.

Williams, G. C. 1966. *Adaptation and Natural Selection.* Princeton University Press, Princeton, NJ.

Wittgenstein, L. 1953. *Philosophical Investigations.* Blackwell, Oxford.

Wolpert, D. M., Ghahramani, Z., and Jordan, M. I. 1995. An internal model for sensorimotor integration. *Science*, 269:1880–1882.

Zahavi, A. 1975. Mate selection. a selection for a handicap. *Journal of Theoretical Biology*, 53:205–214.