Project no. 001940

ECAGENTS

Embodied and Communicating Agents

Sixth Framework Programme (FP6)

Future and Emerging Technologies of the Information Society Technologies (IST-FET Priority)

White Paper: Target Problems and Grand Challenges for Developing Embodied and Communicating Agents

Deliverable 1.1

Due date of deliverable: 31 December 2004
Actual submission date: 31 December 2004

Start date of project: January 1, 2004 Duration: 48 months

Organisation name of lead contractor for this deliverable: SONY-CSL

Table of Contents

Preface

Luc Steels

Sony Computer Science Laboratory Paris

1.1 The ECAgents Project and the White Paper

The ECAgents project proposes fundamental research to understand the role of communication in collections of embodied and situated agents using the methods of complex systems science and computer science. We aim at studying how communication arises, what different types of communication systems there are or can be, what the different pre-conditions are that must be satisfied for the emergence of different types of communication systems, what kind of performances at the collective level different communication systems make possible. In more concrete terms, the project considers basic properties of different communication systems, from simple communication systems in animals to human language and technology-supported human communication.

In studying the role of communication in collections of embodied agents, the project has adopted the following methodological choices:

- By collections of agents is meant a plurality of agents that by interacting together exhibit collective performances that no single agent would be able to generate by acting alone.
- The agents are embodied and physically situated, that is, they are physical agents interacting not only among themselves but also with the physical environment. The agents do not only exchange messages with other agents but they also move or are carried around in physical space and they interact non-symbolically with the physical environment.
- The communication system of the agents is not pre-designed from the outset and is not fixed but it emerges spontaneously from the interactions of the agents among themselves and with the external environment and is in a constant state of flux due to the changing conditions of the agents, their tasks, and their environment.
- Not only the communication conventions but also the underlying ontologies are assumed to self-organise and evolve as the population of agents, the media they use, the environment, and the topics of mutual interest keep changing.
- The research strategy adopted for the project is not the traditional one of studying, analysing, and experimenting with communicating agents that exist in nature but is to construct collections of artificial agents, both simulated in a computer and actual physical artefacts, and to do experiments and test hypotheses with these collections of artificial agents.

– The research strategy includes not only the actual construction of both simulated and real artefacts but also the study of the more abstract and general properties of collections of interacting and communicating agents, e.g, the role of the interaction and communication network topology, the more abstract properties of their communication system (combinatiorial vs non-combinatorial, grammaticalized vs nongrammaticalized, etc.), its dynamical systems properties, etc.
– Given the research strategy just described, the project is not only advancing our scientific understanding of embodied communicating agents existing in reality but it is also suggesting new technologies that consist in collections of physical devices (robots and robot-like artefacts, wireless devices, ubiquitous computing, etc.) that interact with an external environment and communicate both among themselves and with human users.

The ECAgents project, which is firmly rooted in the most innovative and advanced IT-technology that will become widespread in the coming 10 years, includes partners that are already doing concrete experiments with robots, wireless devices, ubiquitous environments, and living systems including humans. However, its main focus is on the development of scientific foundations by using methods, insights, and techniques from complex systems research. An evolving communication system and its underlying adaptive ontology is viewed as a complex adaptive system and evolutionary theory, game theory, network theory, and dynamical systems theory can all significantly contribute to its study. We believe that there is today still a tremendous gap (with some notable counter-examples) between complex systems researchers and IT, but this project is determined to bridge this gap for an issue of major importance. Through the results of this project, we expect significant breakthroughs in many future and emergent technologies, from self-developing robots to the semantic web and ubiquitous wireless devices.

In the original workplan of the ECAgents project, we decided to focus the first year on an analysis of the challenges posed by Embodied Communicating Agents and document the results of this analysis in a White Paper. The reasons for this action are as follows:

1. The ECAgents Project (as indeed the 'Complex Systems FET action' in general) is a highly multi-disciplinary endeavour and one of the major goals of the project is to bridge the gap between different disciplines. A profound discussion of challenges and explicit communication of these challenges among the partners from different disciplines has proven to be a very effective way to achieve this inter-disciplinary discussion and the writing of the White Paper has been a major stimulus.
2. The evolution of communication in embodied agents is still a very new terrain and the actual challenges are far from understood. By mapping out the terrain we are not only in a much better position to refine the workplan for the remainder of the project but also give guidance to people outside the ECAgents project on what good research topics are. This way, the ECAgents project is achieving a multiplicator effect.

We have already communicated results of the White Paper in various fora, such as the Complex Systems Conference organised in Torino by the 'Exystence' network of excellence in complexity (with contributions from 5 ECAgents partners), the IST conference in The Hague (with contributions from 3 ECAgents partners), and a dozen other scientific workshops and conferences in the various disciplines involved in the study of Embodied Communicating Agents. The goal of the White Paper is to capture a large part of our ongoing discussion which is far from finalised. They are reported by the principle researchers involved, although other partners in ECAgents were always engaged in the preparatory discussion for each topic.

The White Paper is an evolving document and the present version is a snapshot. We are committed to continue improving this document as our understanding of the field and its various challenges matures. At some point, when sufficient maturity has been reached, we plan to publish the White Paper as a public document, intended to give guidance to young researchers on challenging open problems in the field. As such, the White Paper will be the basis of other dissemination actions of the ECAgents project.

Communication is an extremely broad field which ranges from very simple animal style communication to extraordinarily complex human style communication. The ECAgents project covers the full range of this broad spectrum, but obviously different challenges apply depending on the nature of the communication system. Hence the document is organised in two parts. The first one focuses more towards animal style communication systems and the second one on human style communication systems, even though we recognise that there is a continuum between the two. The potential for applications based on the results of investigating these two forms of communication is contained in a third major part of this document. The remainder of this preface briefly surveys the various contributions.

1.2 Animal/Animat-Like Communication

The first part of the White Paper focuses on animal/animat-like communication. These are communication systems inspired by animal communication and synthesised on physically embodied animal-like robots, often called animats. It covers the following topics: (1) Definition of communication from a biological perspective, (2) Challenges for achieving emerging communication in animats and (3) Theoretical challenges in the synthesis of Human-Like Embodied Communication systems.

Defining Communication Defining communication is a difficult issue because each related field has their own way of looking at it. The first note, contributed by Eors Szathmary (Collegium Budapest) defines communication from a biological perspective and hence sets the stage for the synthesis of animal-like communication systems.

Challenges for Achieving Animat-like Communication This section introduces the main challenges for the synthesis of communication in animats. There is first a contribution by Stefano Nolfi (CNR) which gives a broad overview of the state of the art. Next, Magnenat and Floreano (EPFL) provide more specific challenges if an evolutionary robotics approach is used to synthesise animat-like communication systems. They also discuss technical opportunities and challenges from the viewpoint of robotics.

Theoretical Challenges Two sections have been contributed. The first one by Peter Hammerstein (Univ of Berlin). focuses on insights obtained from the study of cooperation and conflict in biology, using game theory as a way to frame the issues and provide a reasoned approach towards a solution. These insights are viewed as providing constraints on the design of ECAgents. The second section surveys self-organising communication systems in animals and draws some lessons and theoretical challenges for animats.

1.3 Human-Like Communication

The second part of the White Paper focuses on human-like communication. It covers the following topics: (1) Definition of human-like communication, (2) Challenges for achieving the prerequisites for human-like communication, (3) Stages in achieving human-like language communication, (4) Computational, Neurobiological and Theoretical challenges in the synthesis of Human-Like Embodied Communication systems.

Defining Human-like Communication The first section (reported by Domenico Parisi, CNR Rome) focuses on defining human-like communication as distinguished from animal-like communication. It is inspired by earlier research by Hockett. Parisi introduces and elaborates on eight defining characteristics: Human language has syntax and, more generally, has signals which are made up of smaller signals, is culturally transmitted and culturally evolved, is used to communicate with oneself and not only with others, is particularly sophisticated for communicating information about the external, environment, uses displaced signals, is intentional, is the product of a complex nervous system, and influences human cognition. These characteristics provide the challenging benchmarks against which human-like communication systems are to be measured.

Prerequisites for Human-like Communication The next two sections elaborate on the prerequisites for human-like communication, focusing in particular on attention sharing, and origins of the communication medium.

Human-like communication has been defined by Tomasello, et.al. as manipulation of attention so that it becomes shared. It follows that we need a good understanding of what attention is and how it can be manipulated, particularly because attention sharing can also be achieved without explicit communication and by non-verbal communication such as through pointing. Such an analysis is reported by Hafner and Kaplan (Sony CSL Paris). They identify four aspects: Attention detection, Attention manipulation, Social coordination, and Intentional understanding. Each aspect is related to developmental stages in humans and the state of the art in the synthesis of these prerequisites as well as specific remaining challenges are reported.

It is obvious that communicating agents need a physical medium in which they can carry out their communication, and if the language system is highly complex, this medium needs to be sufficiently complex to allow combinatorial possibilities. Human languages use predominantly speech but in principle other media are possible and may be better for artificial agents (e.g. we can envision that a molecular medium as explored in the Pace project might be usable). Recent work is beginning to show that a population of agents can self-organise a sound medium into a combinatorial sound system similar to human like languages. The contribution by PY Oudeyer (Sony CSL Paris) analysis the current state of the art and the remaining challenges in this area.

Stages The next section (contributed by Luc Steels (Sony CSL Paris)) proposes first a series of six stages that go from a minimal communicating system to a human-like language. Each stage introduces a more complex form of conceptualising reality and hence increased complexity on the meaning side which requires a more complex way of expression on the form side. The transition between one stage and the next each time requires a major 'breakthrough', such as the origins of compositionality, the origins of syntax, level formation for the emergence of a meta-grammar, etc. For some of the early stages there are already quite solid computational mechanisms which make already applications in embodied communicating agents possible, whereas for other stages very little is at this point achieved. They will require major advances both technically and conceptually with respect to the current state of the art.

Neurobiological Challenges The section of this document which presents the stages also contains the challenges in terms of computational modeling of agent architectures at each stage. One of the goals of the ECAgents project is to turn these computational models into neurobiologically realistic models so that it becomes potentially possible to trace out an evolutionary scenario for explaining the brain functions that contributed to the human evolution of language. The challenges related to this effort are reported in a contribution by Eors Szathmary (Collegium Budapest)).

Theoretical Challenges Once there is a good understanding of different stages in the complexity of human language-like communication, it becomes possible to identify very clear theoretical challenges that can be solved using the methods of complex systems science. These theoretical challenges concern (1) Prediction of macroscopic properties of a system of communicating agents based on microscopic behaviors of the agents (e.g. predict that agents will reach a shared lexicon) (2) Understanding of powerlaws underlying the behavior of these systems (e.g. relation between increase in population size versus time towards convergence) and (3) Understanding of limitations and constraints that would push agents from one stage towards the next stage. The ongoing discussion on these challenges in the ECAgents project is reported by Steels, Loretto, et.al. This section also reports the potential methods from complex systems science that could be used to tackle them. The ability to predict properties of embodied communicating agents is obviously a prerequisite for their practical employment.

1.4 Applications

The final part of the White Paper is concerned with applications of Embodied Communicating Agents and reported by Holmqvist (Viktoria Institut), Kaplan and Steels (Sony CSL Paris). Three applications are being discussed which each generate their own challenges: (1) Robotic applications, where agents are embodied in physical form as autonomous robots, with potential applications in service robots and entertainment robots, (2) Ubiquitous applications, where agents can move between different forms of physical embodiment so that agents can exist on devices such as PDA's and digital cameras, and (3) peer-to-peer applications, such as music file sharing, where the problem is semantic interoperability. Each of these applications is currently on the edge of becoming a reality but several basic issues still remain to be solved.

Part I

Animal/Animat-like Communication

Definition of Communication - A Biological Perspective

Szabolcs Szamado

Collegium Budapest

2.1 Definition of Communication

There are numerous definitions of communication, see table 1. for examples. One reason of this bewildering variety is that each discipline and sub-discipline of science has its own definition of communication. Since these disciplines vary with respect to their subject of study and methodology, definitions of markedly different rigour and scope can be found. While it might be impossible to find a definition that would fit all the disciplines, it might be a good idea to favour some of the definitions in order to achieve a common ground that can be shared by all project participants.

Here we would like to introduce and dissect one of the most common definitions used in biology and show the advantages that this viewpoint can offer.

According to Krebs and Dawkins (1984), 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." In our everyday usage (and according to many of the listed definitions), however, communication usually implies the transmission of some information. Thus, as Maynard Smith and Harper (1995) argues there must be a connection between these two approaches because: "it is not evolutionarily stable for the receiver to alter its behaviour unless, on average, the signal carries information of value to it." That is, we can expect that at signalling equilibrium both the signaller and the receiver benefits on average from the use of signals. This can be stated analytically by saying that the value of signals should be always greater than zero at equilibrium (Lachmann and Bergstrom, 2004). Here the value of signals is defined as the difference expectations: the expected fitness given signal minus the expected fitness without signal (Lachmann and Bergstrom, 2004).

In the spirit of the Maynard Smith and Harper (1995) extension of the definition we can differentiate between signals and cues. "Signals are stimuli that convey information and have been moulded by natural selection to do so; cues are stimuli that contain information but have not been shaped by natural selection specifically to convey information." (Seeley, 1989) That is, a cue is an accidental source of information, while signals are selected. Examples for signals: dominance displays, threat displays, courting dance, begging, etc. Examples for cues: size, speed, etc. Both size and speed may convey useful information to the observer about the strength or health of the other animal; however, they were

"Communication occurs when the action of or cue given by one organism is perceived by and thus alters the probability pattern of behaviour in another organism in a fashion adaptive to either one or both of the participants." (Wilson, 1975)
"Communication is the transfer of information via signals sent in a channel between a sender and a receiver." (Hailman, 1977)
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." (Krebs and Dawkins, 1984)
"Communication is a matter of causal influence . . . the communicator [must] construct an internal representation of the external world, and then . . . carry out some symbolic behaviour that conveys the content of that representation." (Johnson-Laird, 1990)
"The term is used here in a narrower sense, to refer to the behaviour by which one member of a species conveys information to another member of the species." (Kimura, 1993)
"Communication occurs, if and only if, information moves from the input to one process to the output from a second process, the latter process being the reserve of the first process." (Losee, 1999)
"Communication is the activity of communicating; the activity of conveying information" (word reference.com)

Table 1. Various definitions of communication

selected for a different reason. As for size, the reasons are to fend off rivals, offer protection against cold, etc., while with regard to speed the reasons include catching prey or escaping predators, etc. The exact terminology is not important in a sense that some author might call cues signs and call permanently on signals cues (Hauser, 1996). The important thing, however, is to make a distinction between those features that were designed/selected for transmitting information, and those features that were designed/selected for some different function, but as a side-effect, they might convey some piece of information useful for potential receivers. While one would like to call the first type of interaction communication (i.e. the one that involves signals), the second type is only observation or eavesdropping (i.e. the one that involves cues).

All in all, the following criteria should hold for any communication system:

- Signals should be used, where a signal is an object/behaviour/morphological feature that was selected/designed to convey that particular piece of information.
- There should be a positive feedback between the efficiency of information transfer and the fitness/reproductive success of the given agent/animal.

Why is this last point important? Because it is this positive feedback process that can guarantee the maintenance and the long-term stability of a given communication system. Of course, such a definition would exclude one-shot interactions; however it might not be a great loss. In fact, if the aim of the project is to design embodied communicating agents that can communicate reliably, then excluding one-shot interactions might be a very desirable research strategy.

Challenges for Synthesizing Animat-like Communication

Stefano Nolfi

Institute of Cognitive Sciences and Technologies, CNR

3.1 Introduction

In this chapter we will focus on how an effective communication system might arise among a collection of initially noncommunicating embodied agents. Answering this question is important for both scientific and technological reasons. From a scientific point of view, understanding how communication might emerge in a population of interacting embodied agents might shed lights on the evolution of animal communication and on the origin of language. From a technological point of view, understanding the fundamental principles involved might lead to the development of innovative communication methods for multi-agent software systems, robotic systems and ubiquitous computing devices.

Existing models of emergence of communication often focus on specific aspects, such us (a) how a shared communication system can emerge in a population of interacting agents (e.g. Steels, 1999; Cangelosi and Parisi, 1998) (b) how a structured form of communication can emerge from a simpler unstructured communication system (e.g. Kirby, 2001; Cangelosi and Parisi, 2002), (c) language acquisition and transmission (e.g. Billard and Dautenhahn, 1999; Steels and Kaplan, 2000a; Sugita and Tani, 2004). In this chapter, instead, we will focus on the more general question of how a population of embodied and situated agents that have to solve a given adaptive problem might develop forms of interaction and communication that enhance their adaptive capability.

The motivation of this choice is twofold. The theoretical motivation is that communication and communication systems are adaptive capabilities shaped by their function. What, when and how agents communicate (and whether agents do or do not communicate) depends on the adaptive function of communication. Similarly the type of communication system that might self-organize in a population of interacting agents will strongly depends on the type of behaviour that individuals display in isolation and on the complementary functions that interactions and communications might have. The underlying assumption is that communication and language can be properly understood by taking into account their relation with other important behavioural, social, and cognitive processes. The practical motivation is that, from an application point of view, the possibility to develop embodied agents able to solve real life problems by exploiting complex forms of interaction and communication might have huge application potentials.

In this perspective, three additional aspects play a crucial role.

1. We are interested in models that not only allow the emergence of a communication ability and a shared communication system but that also allow the *discovery of categories (or coupled internal/external dynamical processes) that are useful from the communication and cognitive point of view* and that are not already explicitly or implicitly identified in the experimental set. This claim is based on the theoretical assumption that one of the main reasons that explain why the acquisition of a communication ability might enhance the cognitive/adaptive abilities of interacting agents are indeed: (1) the fact that the indirect adaptive advantages of communication might force the development of useful and compact ways to categorize the continuous flow of sensory-motor information, and (2) the fact that the discoveries of useful categories might be easier through social interactions than in isolation.

2. We are interested in models in which individuals, beside from reach signalling and interaction capabilities, also have a *reach sensory and motor non-communicative repertoire* that might allow them to improve their ability to solve their cognitive/adaptive problems by improving both their individual and their social/communication capabilities. This claim is based on the assumption that only by co-adapting their behavioural non-communicative and communicative abilities, individuals might develop a really useful communication system grounded in the physical and behavioural characteristics of communicating individuals and able to exploit active perceptual capabilities. Moreover, this claim is based on the assumption that one of the key aspects of communication is the possibility to rely on implicit information that does not need to be communicated.

3. Finally we are interested in models in which *forms of communication of different complexity might be used*. By forms of communication we refer to the protocol with which individuals interact during communication and to the way with which communication signals are structured. Forms of communication might range from simple continuous broadcasted signalling to complex regulated communication protocols in which, for instance, communication acts are episodic and asynchronous, communication protocols are negotiated on the fly between the two communicating agents, and communication acts consists of sequences of signals organized according to a grammar. This claim is based on the assumption that more complex forms of communication are not effective in general terms. Therefore agents should be left free as much as possible to select the communication form that is most useful, given their current behavioural/cognitive capabilities.

It is important to point out that our aim here is not that to discuss how animal like communication can be modelled but rather how embodied artificial agents can develop forms of communication that are functionally equivalent to those that can be observed in animal-like communication. Moreover, although we will focus on the question of how relatively simple forms of communication might emerge from scratch, we will also address the issue of how complex forms of communication (including human-like communication forms) might emerge from simpler forms.

3.2 State of the art

In this section we will review the research works that are more relevant to the perspective outlined in the previous section. In section 3.2.1, we will review experiments in which agents, that are asked to solve simple tasks that require cooperation and coordination, develop simple forms of ritualised social interactions and/or signalling capabilities. In section 3.2.2, we will review experiments in which agents interacting according to predetermined ritualised interaction schemes and able to modify their internal states on the basis of the result of such interactions, develop an ability to successfully categorize external objects according to a self-organized shared vocabulary and ontology. The aim of this section is not that to provide an exhaustive review of the area (for a broader review see Wagner et al., 2003a; Steels, 2003a; Cangelosi and Parisi, 2002) but rather to identify theoretical and experimental contributions that might lead to the development of more powerful models and/or to models in which aspects previously studied in isolation can be integrated.

3.2.1 How simple forms of communication might emerge in teams of adaptive interacting agents

One interesting demonstration of how behaviours with communicative functions might emerge from the attempt to solve a task that requires cooperation and coordination has been provided by Quinn et al. (Quinn, 2001; Quinn et al., 2003). The author evolved a team of mobile robots for the ability to move by remaining close to one another. Robots are only provided with proximity sensors (that also allowed robots to avoid colliding with one another) and therefore do not have dedicated communication channels. Evolved individuals are able to solve the coordination problem by communicating through a sequence of sensory-motor interactions. For instance, in a simple case described in Quinn (2001), two evolved agents coordinate according to the following sequence of behaviours: (1) both agents rotate clockwise, (2) the agent that first faces the other agent with its front (agent B) moves toward the other agent (agent A), (3) agent B remains close to A by moving backward and forward in order to compensate A's movements, (4) once agent A faces agent B with its front, it reverses his direction and then it starts to move forward by being followed by agent B. Agents A and B thus assume the roles of leader and follower respectively.

The motor action of the first aligned agent (i.e. the back and forth behaviour that allows agent B to stay close to agent A and that, consequently, produces a high activation of A's infrared sensors) serves as a signal for the other agent (as reported by Kirby (2002) we might gloss it in English as "after you"). In fact, "if the agent perceives the signal while it is still rotating, it will adopt the leader role. However, if it becomes aligned without having perceived the signal, it will perform the signalling action and subsequently take the follower role" (Quinn, 2001).

By analysing how the evolved behaviour originated evolutionarily, the authors observed how the behaviour of one agent that produces sustained proximity and that triggers the reverse behaviour in the other agent (i.e. the behaviour that has a communication value) resulted from the adaptation of other elementary behaviours (the obstacle avoidance behaviour and the back away behaviour) that did not have communicative functions. Indeed, by analysing the evolutionary process, the authors observed four phases:

1. Initially (20-50 generations) agents just turn both motors on thus moving in straight lines
2. Later on (50-100 generations) agents develop an ability to avoid each other. During this phase, the turning and halting responses displayed by the agents to avoid each other often result in 'deadlock' situations in which the two agents remains close one another.
3. Later on (110-370 generations) deadlock situations are broken as a result of the fact that one of the two agents backs away from its partner after some time allowing the partner to move towards it for a while. The continuation of this process leads to a slow and jerkily movement of the couple.
4. Finally (from generation 370 on) agents display an ability to reverse in response to sustained proximity. This new reversing behaviour that allow agents to start moving in a coordinated manner capitalises on the straight movement and avoiding behaviour that previously served other functions.

It might be questionable whether this form of interaction is a form of communication or not. Indeed, this is a paradigmatic case in which actions in general and communication actions can hardly be differentiated. This difficulty can be explained by considering that the term communication does not have a clear and uncontroversial definition (Di Paolo, 1998; Castelfranchi, in preparation), and that distinguishing between communicative and non-communicative actions is especially difficult in the cases of simple forms of communication. For the purpose of this section it is sufficient to say that we will attribute a communication value to all actions or sequences of actions that, by influencing the sensory-motor flow of other agents, enhance the adaptive ability of the group as a whole. The reason why we do not simply call these actions communication acts is that, in addition to a communication value, they might have other functions (e.g. they might allow agents to avoid obstacles, an ability that does not necessarily influence the behaviour of other agents).

In another recent work, teams of 4 mobile robots have been evolved for the ability to aggregate and to move together towards a light target (Baldassarre et al., 2003). Robots are provided with two motors controlling the two wheels, a speaker continuously emitting a sound, infrared sensors, and directional microphones. As in the case of the Quinn's experiments described above, evolved individuals display an ability to coordinate by interacting/communicating so as to assume and maintain different roles. In particular, robots are able to form a

square-like formation in which each individual robot maintains its relative position with respect to the light and to the other robots, while the whole group moves straight toward the light. Interestingly evolved robots are able to assume different roles despite teams are constituted by identical reactive individuals (i.e. agents that always react in the same way to the same sensory state).
By evolving teams of robots for the ability to solve a collective navigation problem, Nolfi and Marocco (submitted) showed how robots develop communication abilities and a vocabulary including 4 signals that influence both the motor and signalling behaviour of other robots. Robots are asked to find and remain on two feeding areas by equally subdividing themselves between the two areas. The team consists in wheeled robots provided with infrared and sound sensors and actuators controlling the two wheels and a sound speaker.

In this experiment: (1) the number, the form and the meaning of signals (i.e. the effects of signals on other agents) are not implicitly determined in the experimental setting but rather emerge during the evolutionary process, (2) non-communicative and communicative actions are tightly co-adapted so as to maximize useful properties emerging from their interactions, and (3) evolving individuals also display an ability to develop a simple form of communication protocol that allows them to switch signalling behaviours on and off.

Other researchers focused on the emergence of mutual interaction between two cooperating agents. Di Paolo (2000) reported the results of a set of experiments in which two simulated agents moving in an arena have been evolved for the ability to approach each other and to remain close together as long as possible. Agents are provided with: (1) two motors controlling two wheels, (2) a sound organ able to produce sounds with different intensities located in the centre of the agent's body, (3) two sound sensors symmetrically placed at ?45 degrees with respect to the frontal side of the agent that detects the intensity of the sound, and (4) a recurrent dynamical neural controller with four internal neurons. Evolved agents successfully approach each other by later remaining close to one another. Moreover:

1. evolved agents self-stimulate themselves through their own sounds. By reducing agents' capacity to ear their own sounds, in fact, the author observed that agents' performance deteriorated.
2. the intensity of sounds produced by the two agents has a marked rhythmical shape that results from the interactions of the two agents. After some time, in fact, signals are phase-locked at some value near perfect anti-phase (i.e. a form of simple turn taking occurs in signalling behaviour) and the movements of the two robots become highly coordinated. This coordination between motor and signalling behaviours of the two agents cannot be explained by the ability of one of the two agents to adapt to the behaviour of its partner only, but rather by the achievement of a dynamical co-adaptation process (entrainment). As shown by the author, in fact, non-plastic beacons producing rhythmical signals are unable to trigger the same type of coordination process.

In a related work, Iizuka and Ikegami (2003a,b) evolved two populations of simulated agents living in couple in an unstructured arena that should exchange their roles (chaser/evader) so as to produce a form of turn-tacking behaviour. Chasing and evading are defined as staying or not staying behind the other agent, respectively. Evolving agents are provided with a feed-forward neural network with three layers including: (1) three sensory neurons encoding the other agent relative position and orientation and three context units whose activation value is copied from that of the activation state of three additional output units at time $t-1$, (2) ten internal neurons, and (3) two motor neurons encoding the desired speed of the two wheels and three additional output units that are used to predict the activation state of the three sensory units at time $t+1$. Evolving agents are selected for the ability to alternate their roles and to predict each other's behaviour. Individuals are evaluated in pairs and each individual of a population is evaluated, in different trials, with all the individuals of the other population. The sensory state at time $t+1$ is used to compute a prediction error that is then used to change the connection weights according to the back-propagation learning rule.

The analysis of obtained results shows how in early evolutionary phases agents tend to produce regular turn taking (i.e. the two agents display regular trajectories that allow them to exchange their role periodically). In successive evolutionary phases, instead, agents tend to display chaotic turn-taking (i.e. the two agents display non-geometrical and an always changing trajectory without fixed periodicity). Regular turn-takers are comparatively insensitive to noise (probably due to their simple dynamics) with respect to chaotic turn-takers. However, chaotic turn-takers are better capable to adapt online to the other agent's behaviour with respect to regular turn-takers. Tests made by using passive agents (i.e. agents unable to adapt their behaviour on the fly) showed how the evolved turn-tacking behaviours are not simply forms of oscillator but rather forms of dynamic coupled behaviours resulting from ongoing two-directional interactions.

The visual inspection of the agents trajectories and the analysis reported above seem to indicate that interesting forms of interactions and communication occur. Moreover, although the role of prediction learning is not analysed in detail, obtained results seem to indicate that the ability to predict the other agent's behaviour might constitute an important pre-requisite for the possibility to develop effective turn-taking behaviour. Finally, as pointed out by the authors, turn-taking is certainly an important pre-requisite for the emergence of complex forms of communication.

Overall, the experimental results above demonstrate how individuals selected for the ability to perform a cooperative task might not only develop forms of communication but also primitive forms of communication protocols that in turn enhance their communication/interaction abilities.

Although these models provide important insights and demonstrate how simple forms of communication might emerge from scratch,

however, they only lead to the development of simple forms of communicative and non-communicative behaviours. How these models can be extended in order to deal with more complex and reach situations is an open research issue that will be discussed in section 3.3.1 and 3.3.

3.2.2 How a population of communicating agents might lead to the self-organization of an ontology and a shared lexicon

In the Talking Head experiment, Steels (1999) demonstrated how the interaction between a population of embodied and communicating agents might lead to the self-organization of a shared lexicon as well as a perceptually grounded categorization of the world. Although the goal of this research is not that to observe how communication might emerge as an indirect result of the need to accomplish a collective task, this model represents an important reference point and provides important insights on crucial aspects that are simplified in the models reviewed in the previous section.

In the Talking Head experiment the environment consists of an open-ended set of geometrical figures (objects) pasted on a white board. The population consists of a number of software agents that are sequentially embodied into two robots provided with a pan-tilt camera and a simulated sound auditory and production systems (for a similar model implemented on mobile LEGO robots, see Steels and Vogt, 1997b). The two robots look toward the white board and interact by playing a language game in which they assume the role of the speaker and the hearer, respectively. During each game, the speaker identifies a randomly selected object on the white board and produces a word or a sequence of words that should allow the hearer to identify the corresponding object. The hearer then tries to identify the area to which the speaker is referring to by visually pointing to the area itself. The speaker finally responds by pointing to the selected area thus allowing the hearer to identify whether communication was successful or not, and, in the latter case, which was the correct target area. As a result of each game and on the basis of the course of the game (e.g. the fact that the hearer already has in its vocabulary the words produced by the speaker or not, the fact that the hearer did or did not successfully identify the target area), agents modify their internal vocabulary and ontology (i.e. the meaning associated to the words of their vocabulary). The continuation of this process leads to: (a) an increase of successful games (up to almost 100%), and (b) to the development of an effective lexicon and an ontology shared within the population (i.e. a lexicon and an ontology that allows agents to play the language game successfully). Such self-organized lexicon and ontology also fulfils the environmental and body characteristics experienced by the agents (e.g. the discrepancy between the two agents' field of view, the reliability of the robots visual system, the specific type of objects and configurations of objects located on the white board).

Agents are provided with hand crafted sensory pre-processing routines and with predefined motor skills and schemas of interactions. Sensory pre-processing routines consist in: (1) software routines that allow an agent to extract a sequence of perceived objects and their relative properties (such as the horizontal and the vertical position of the object, its average grey scale value, its area, the number of edges etc.) from a visual scene, (2) software routines and position sensors that detect the point to which the speaker robot is visually pointing to, (3) software routines that allow the hearer to receive as input the sequence of words produced by the speaker. Motor skills consist in, for example, a software routine that allows an agent to identify a unique area on the visually perceived scene on the basis of a sequence of words with their associated meanings. Schemas of interactions consist, for example, in: (1) routines that create a new word with its tentative associated meaning in the vocabulary of the hearer when it hears a word that it is not included in its vocabulary, (2) a routine that creates a new word in the vocabulary of the speaker when none of its current words uniquely identify the current selected object of the white board, (3) a routine that updates the communication success rate associated to words, etc.

What results from the changes in agent's internal structures occurring during agent's interactions are: (1) a perceptually grounded categorization of the world (consisting of a lexicon and a corresponding ontology), and (2) the convergence of the population toward a sufficiently shared lexicon and ontology. As an example of word/meaning formation, consider that the horizontal position of an object ranging from 0.0 to 1.0 might be categorized into two categories/words (corresponding, for example, to the two halves of the range) or into finer and finer categories with their corresponding words. As a second example consider that one object (i.e. a red triangle located in the top-left side of the board) might be discriminated in different ways (e.g. by using words that indicate its shape and colour or its position). Finally, consider that the same meaning can be associated with two or more words and two or more words might have the same meaning (both at the level of the single agent or at the level of the population). Indeed, by analysing the frequency of words used to express a single meaning in one experiment, one can observe a struggle in which different words compete until the population settles on a single dominant word. This winner-take-all effect is due to a positive feedback loop between use and success. The more agents prefer a particular word, on the average, the more they use this word and the more success this word has.

In a successive work, Steels and Kaplan (2000b) used a similar approach to study how a Sony AIBO robot might acquire a lexicon and a corresponding ontology by a human mediator with whom it plays a similar language game. The use of a mobile autonomous robot (rather than a pan-tilt camera placed on a fixed position as in the case of the Talking Head experiment) introduces significant new complexity from the point of view of the categorization problem given that objects are almost never seen in their entirety and objects' perceived images significantly vary on the basis of the robot/head/object relative positions and orientations. The robot/human interaction is regulated on the basis of a predefined sequence of elementary behaviours (a language game). More precisely:

1. The human mediator first shows an object to the robot by placing the object in the robot's field of view and by saying "look", a word that helps the robot to focus its attention on the current visual scene. The robot then concentrates on the object by trying to track it and touch it.
2. The human label the object with a word ("ball" for example).
3. The robot tries to pronounce the same word. The human mediator then provides a positive feedback (i.e. pronounce the word "yes") or repeats the original word if the word it hears is different from the one it previously produced. If the word is a new one for the hearer robot, it creates a new word in its vocabulary.
4. The robot stores in its memory a perceived instance of the object and associates it with the corresponding word. The comparison of a new perceived image with the labelled images previously stored later allows the robot to identify and name an object without the help of the human mediator.

As pointed out by the authors, several problems might arise during these human/robot/environment interactions. For example, the robot might have heard a wrong word due to problems with speech recognition or the robot might not have been paying attention to the right object. The impact of these problems, however, is minimized by the interactions with the human mediator regulated by the language game script (i.e. the human mediator repeats the word if it has not been properly understood by the robot or tries to bring the robot's attention on the right object when the robot pays attention to something else). For a related model that addresses how a communication ability can be socially transmitted from a robot with a predetermined lexicon and other robots see (Billard and Dautenhahn, 1999).

These models present two important advantages with respect to the models described in the previous section, namely: (1) the ability to exploit social learning, and (2) the ability to exploit ritualised interactions between agents (language games). The implication of these aspects will be discussed in the next sections. The main limitation of these models is that, aside from the content of communication acts, the behaviours of agents is rather predetermined and fixed. This prevents the possibility to exploit a co-adaptation between communicative and non-communicative forms of behaviour. Moreover, this makes these models not suitable to solve general co-cooperative problems (e.g. cooperatively explore an unknown area) or to study how ritualised interactions, language games and vocabularies might have originated.

3.3 Open research problems: identifying and integrating crucial cognitive/behavioural capabilities

The attempt to model how a population of embodied agents trying to solve problems that require cooperation and coordination might develop complex forms of communication and a shared communication language is a formidably complex enterprise. The research works reviewed in the previous section show how several aspects that might allow to achieve such a goal can be modelled (e.g. how signalling behaviours and primitive forms of communication protocols can emerge, how communicative and non-communicative behaviour can co-adapt, how a population of interacting agents might develop a shared lexicon and ontology). However, the modelling of other crucial aspects (e.g. compositional languages and grammar) is only at a very preliminary stage (Steels (2003a)). Moreover, a significant challenge is constituted by the need to integrate aspects that have been successfully modelled in different experimental settings into a single coherent model. In the rest of the section we will discuss how important aspects that might represent an important prerequisite for the emergence of complex forms of communication can be modelled and how all the necessary aspects might be integrated into a single model.

From an evolutionary and developmental perspective the most straightforward way to approach the issue of how complex forms of interaction and communication can emerge is to start from simple but open-ended models that might lead to the emergence of progressively more complex forms of communication and cognitive capacities. After all, this is how these abilities emerged in natural life. This possibility, however, can reasonably be pursued only as a long-term research goal. On the short term, it is reasonable to assume that progresses might be only achieved by predefining, in the starting conditions, crucial elements or capacities that although in theory could spontaneously emerge in the course of the process, in practice, would very unlikely do so. These elements or capacities might consist in agent's pre-determined architectural constraints, learning algorithms, interaction schemas, etc. From this point of view our problem becomes that of identifying the crucial minimal set of pre-requisites that might trigger the emergence of complex forms of interactions and communications.

3.3.1 Adaptation processes

A promising way to develop ECAgents is that to rely on a self-organization process based on evolutionary and/or learning techniques. On the relation between self-organization at the level of the adaptation process and at the level of behaviour exhibited by a collection of interacting individuals see the chapter 6. The models reviewed in section 3.2.1 rely on an evolutionary process (i.e. a process based on selective reproduction and random variation) while the models described in section 3.2.2 rely on a form of ontogenetic learning (i.e. a process in which agents modify their free parameters as a result of their interaction with the physical and social environment). These two forms of adaptive processes have complementary characteristics and can be effectively integrated (see Nolfi and Tani, 1999). In this section we briefly discuss some of the potential advantages of integrating an evolutionary and a learning process.

Artificial evolution, by only requiring an overall evaluation of the performance of an agent or of a group of agents, is a straightforward method to select solutions in which different characteristics co-evolve and co-adapt. For example, as clearly shown in the models reviewed in section 3.2.1, it is an effective way to co-evolve communicative and non-communicative behaviours. Learning, on the other hand, by being based on changes introduced as the result of the continuous interaction with the physical and social environment, can potentially exploit the huge amount of information that agents collect through their sensors during their lifetime. This information does not provide direct cues on how agents should change to solve their adaptive problems. However, combined with additional evolved or handcrafted mechanisms able to transform sensory information into teaching or reinforcement signals (Ackley and Littman, 1991; Nolfi and Parisi, 1997) or able to channel changes on the basis of genetically encoded constraints (Floreano and Urzelai, 1998) can lead to powerful ontogenetic adaptive processes.

Evolution and learning operate on different time scales. Evolution is a form of adaptation capable of capturing relatively slow environmental changes that might encompass several generations. Learning, instead, allows an individual to adapt to environmental changes that are unpredictable at the generational level. Indeed, the combination of evolution and learning can lead to an ability to develop the required behavioral capabilities and to an ability to select on the fly the right strategy on the basis of the current environmental circumstances (Nolfi and Parisi, 1997; Nolfi and Floreano, 1998 and Floreano and Urzelai, 1998).

More generally, the interaction between evolution and learning deeply alters the dynamics of the two processes so that their dynamic in interaction is very different from their dynamic in isolation. Indeed, evolving plastic individuals tend to develop a predisposition to acquire their capabilities through learning rather than, directly, an ability to behave effectively as in the case of evolving non-plastic individuals. This predisposition to learn may consist of: (1) the presence of starting conditions that canalise learning in the right direction, and/or (2) an inherited tendency to behave in a way that maximizes the chance to be exposed to useful learning experiences. Similarly, while in non-evolving individuals the value of free parameters prior to learning is a constraint that should be overcome, in evolving individuals inherited genetic parameters prior to learning represent an opportunity to be exploited during learning (Nolfi, 2002a).

Finally, as we will discuss in Section 3.4.3, social learning (i.e. learning from others) might potentially allow evolving individuals to acquire capabilities independently discovered by other different individuals.

3.3.2 Agents' sensory-motor structure

Another aspect that strongly affects the potential outcome of experiments involving a population of interacting agents is the type of sensors and motors (actuators) with which agents are provided. We will not discuss here the possibility to co-evolve/co-adapt the body and the control system of agents although this possibility certainly provides potential advantages (I. et al., 1994; Sims, 1995; Bongard and Pfeifer, 2003). Rather we will try to identify general criteria that the experimenter might follow in determining a suitable sensory-motor structure.

1. The first aspect that should be stressed is that sensors and actuators should not be conceived as independent entities that have independent functions. Indeed, by interacting with the external environment (i.e. by modifying their own position or orientation with respect to the environment or by modifying the environment itself) agents might greatly simplify the problem of categorizing environmental situations that require different motor reactions (Scheier et al., 1998; Nolfi, 2002b; Nolfi and Marocco, 2002; Beer, 2003; Nolfi, in press). Moreover, the possibility to interact with the environment by producing simple stereotyped behaviour, might allow agents to indirectly detect complex environmental regularities (Nolfi and Marocco, 2002; Nolfi, in press). In other words, reach sensing capabilities might be more likely obtained by complementing a set of sensors with motors that allow agents to interact with their environment rather than by simply adding additional sensors. It should be noted, however, that to really exploit sensory-motor coordination agents should not only be provided with sensors and effectors but should also be able to modify (through an adaptation process) the relation between sensors and motors. In the Talking Head experiment reviewed in section 3.2.2, for example, agents are provided with motors controlling the pan-tilt movement of the camera. However, given that the motor behaviour of these agents is predefined and fixed, the way in which they interact with the environment cannot be co-adapted with their current ontology.
2. A second important aspect that should be stressed is that communicative and non-communicative sensory-motor channels cannot and should not be separated. In fact, elementary behaviours that initially do not have any social functions and that have an impact on the sensory systems of other agents might later on assume a social/communicative function. These forms of preadaptations (in which traits evolved for a non-social function later assume a social/communicative function eventually loosing, later on, their original non-social function) might play an important role in the emergence of communication. Indeed, they seems to have played a crucial role in the origin of the communicative behaviour described by Quinn (2001) and reviewed in section 3.2.1.

The fact that in natural organisms (and probably in self-organizing artificial agents) sensors and actuators tend to have both non-communicative and communicative functions, however, does not imply that some type of sensors and actuators and some sensory-motor modalities might potentially have a strong communication potentials. This is the case, for example, of the sensory-motor structures that allow pointing, detection of pointing (e.g. gazing, head-movements, arms and fingers movements etc.). Moreover, some types of sensors and actuators or sensory-motor modalities might be especially suited for communication for their ability to convey information ready to be used from other agents. As an example of this category consider pheromone that: (1) by lasting a significant amount of time can be detected over a significant time range, (2) by remaining in the physical area in which it has been synthesized can convey spatial information in a ready to use way, (3) by summing up the trace left by different individuals can provide compact information on what several individuals did.

3.3.3 Cognitive capacities

In addition to suitable sensors and actuators, embodied and communicating agents should be provided with a control system that determines the activity of the actuators on the basis of the current and previously experienced sensory-motor states. Although simple forms of communication might be developed by relying on very simple control systems (e.g. reactive neural networks in which sensory neurons are directly linked with motor neurons and motor actions are only based on current sensory states), the development of more complex forms of communication might require much more complex "cognitive" abilities.

Two basic capabilities that embodied and communicating agents should have are: (1) the ability to form internal categories by mapping sensory patterns or sequences of sensory patterns that require similar motor reactions into similar internal states or into similar internal dynamics, and (2) the ability to generalize, that is the ability to react to new sensory patterns (or sequence of sensory patterns) on the basis of their similarities with previously experienced sensory patterns (or sequence of sensory patterns).

While the possibility to form categories based on single sensory states and the ability to generalize on the basis of these categories have been successfully modelled (Cangelosi and Parisi, 1998; Steels, 1999; Steels and Kaplan, 2000b; Marocco et al., 2003), the possibility to form categories based on regularities that can only be detected by looking at how sensory states change in time is still far from being well understood. Consider, for example, cases in which agents have to discriminate different locations of the environment on the basis of the occurrence of different sequences of sensory cues (Nolfi, 2002c), or select moving objects to be caught on the basis of their trajectories (Beer, 2003). To perform these categorization processes agents should be able to takes into account aspects such us the duration of an event or the sequence with which different events occur that can only be detected by looking to how sensory states change in time. For recent results that indicate how the availability of internal states that change at different time rates might represent an important pre-requisite for solving this problem, see (Nolfi, 2002c; Gers et al., 2002; Croon et al., in press). Recent results also indicate the importance of viewing categories as dynamical internal processes rather than as fixed-point attractors in agents' internal dynamics (Beer, 2003; Sugita and Tani, 2004; Iizuka and Ikegami, in press). For an attempt to model categorization as a bi-directional coordination between the dynamics resulting from the agent/environment interaction and the agent's own internal dynamics see (Di Paolo, 2000; Iizuka and Ikegami, in press)

The emergence of complex forms of communication might also require other more complex cognitive capacities such as the ability to predict the sensory-motor consequences of agents' own actions (Nolfi and Tani, 2002; Clark and Grush, 1999), the ability to predict changes in the physical and social environment, the ability to establish shared attention (see chapter on joint attention 8), the ability to learn from others or to imitate other agents' behaviour (Billard, 2000; Tani et al., 2004) etc. The later issue will be discussed in more details in section 3.4.3.

An additional interesting aspect that might be investigated is whether the ability to have access to their own communication acts (i.e. talking to themselves (Steels, 2003b)) might improve the ability of agents to communicate and/or the ability to acquire complex cognitive abilities.

Finally, the emergence of complex forms of communication very likely requires selective attention mechanisms and/or an ability to modify communication behaviours on the basis of the potential targets of communication acts. This aspect will be discussed in more details in the next section.

3.3.4 Interaction/communication protocols

The adaptive potential of social interaction/communication significantly depends on the protocol that regulates communication between agents. Indeed, communicative actions might have counter-adaptive effects on other agents' behaviour and on the adaptive capability of the population as a whole. For instance, communication acts might interfere with other agent's behaviours thus preventing or delaying the ability of these agents to accomplish their current tasks.

In general terms, one can expect that the adaptive potential of communication depends on the ability of communicating agents to regulate the communication dynamics on the basis of a suitable interaction/communication protocol and specifically:

1. The ability of agents of limiting communication acts (i.e. actions that have an effect on other agents behaviours) to those that can increase the adaptive capability of the team. Interestingly, this aspect might lead to an adaptive pressure to use dedicated communication channels (i.e. to detach communication actions from non-communicative behaviours).
2. The ability to detect the potential target agents of communication and to filter and/or re-code communication so as to provide to receivers relevant, useful, and ready to use information. This ability to modify communication on the basis of receivers' needs might include, for example, the ability to recode spatial information on the basis of the relative position of the 'speaker' and the 'hearer' or the ability to detect the adaptive needs of target agents.
3. The ability to approach other agents in order to communicate, to potentially receive communicative information, to select good learning experiences or to achieve joint shared attention (on the last aspect see Billard and Dautenhahn, 1999).
4. The ability to regulate the communication flows by taking turns (Iizuka and Ikegami, 2003a,b) or more generally the ability to carry on communication behaviours consisting of several bi-directional communication acts.

5. The ability to increase communication success through a ritualised form of interaction (Steels, 1999) between communicating agents (e.g. a communication protocol in which the hearer repeats the detected communication signal and waits for a confirmation from the speaker).
6. The ability to communicate through signals with time-varying properties or sequences of signals eventually structured according to a given grammar.

Obviously, the full set of abilities is only required in complex forms of communication. Simple communication forms, such us signalling of danger situations, in which: (1) few different signals are needed to communicate the relevant information, (2) communication acts occur only sporadically, and (3) communication acts have a priority on all other types of activities and are relevant for all members of the population; communication might successfully emerge without the need of any communication protocols.

3.4 Open research problems: identifying the conditions that might lead to the emergence of ECAgents

While in the previous Section we tried to identify the functional components that should be integrated to lead to complex forms of communication, in this section we will try to identify the conditions that might lead to complex forms of interaction and communication. Given the difficulty of the enterprise, our goal is not the attempt to answer to this question, but simply to identify open problems and sketch some interesting research directions.

3.4.1 How communication can emerge as a result of indirect selective pressure

One first important open question concerns whether non-trivial forms of communication can evolve as a result of an indirect selective pressure originating from the need to solve a given adaptive problem. This question involves two aspects: (1) the identification of the structural, cognitive and behavioural prerequisites for the emergence of complex forms of communication, and (2) the identification of the situations (i.e. the class of problems and/or the environmental and social conditions) that might exert an adaptive pressure to communicate. While in the previous section we focussed on the former issue, in this section, we will focus on the latter.

As we claimed in the introduction, the attempt to evolve communication without explicitly rewarding communication is crucial in order to allow the emergence of a self-organization process in which: (a) communication abilities and communication systems are not indirectly predetermined by the experimenter, (b) communicative and non-communicative behaviour can freely co-evolve and co-adapt, and (c) individuals are free to determine the most effective way to categorise sensory-motor information. However, this leaves open the problem of determining the conditions in which indirect selective pressure on communication can be expected.

In their pioneering work on evolution of communication Werner and Dyer (1992) suggest that an evolutionary pressure on agents to communicate should be expected in cases where "animals [agents] have information that other animals needed to know but were not capable of finding out by themselves" (Werner and Dyer, 1991 pp.661). This general hypothesis might be further detailed by identifying the conditions in which this situation occurs. Indeed, we might identify at least the following cases:

1. Information related to the internal states of an individual beyond the nervous system (e.g. hormones, internal organs, immune system, emotional states etc.). This information might be highly valuable in order to determine how socially interact properly. Moreover, information related to the internal states of an individual might indirectly provide compact cues on the previous sensory-motor experiences of that individual.

2. Information related to the current sensory state experienced by an individual (e.g. sensory information indicating the presence of a predator). This form of information might be useful to other individuals that, by being located in different positions and orientations or by not being provided with the same sensing capabilities might not have access to it.

3. Information related to what an agent is going to do (e.g. information related to the action that an agent is going to perform or related to more abstract intentions of an agent).

4. Information about the external environment collected by an agent during its previous interaction with the environment (e.g. information on the location of a food source that is no longer in the agent's sight).

Other aspects that might co-determine whether or not an indirect selective pressure on communication could be expected regards the relation between individual and collective interests (an issue that will be discussed in the next section), the nature of the problem (i.e. whether or not the problem requires cooperation), and the relative organization of the interacting agents (whether the problem requires specialization and whether agents can assume different specialized roles). With respect to the last aspect, a selective pressure on the emergence of communication might more likely be expected in a team of homogeneous rather than in non-homogeneous agents. As showed by (Haynes and Sen, 1997, 1995) in fact, while agents that are not specialized might need to communicate to negotiate their role on the fly, specialized agents do not need to communicate in order to negotiate their relative roles.

3.4.2 Adaptive factors in the evolution of communication

Beside the problem of determining how a given problem might exert an indirect adaptive pressure on the emergence of communication, we should be able to identify the conditions in which communication might emerge evolutionarily. The emergence of communication in fact, requires the development of two complementary but independent abilities: an ability to produce signals (from the point of view of the signaller) and an ability to appropriately react to received signal (from the point of view of the receiver). When selection operates at the level of individuals, two aspects might prevent the emergence of communication, namely: *the lack of an adaptive benefit for the signaller* and *the conflict between individual and collective interests*. For a detailed account of conflicts of interests in evolutionary theory, see chapter X by Hammerstein.

The first problem is due to the fact that in many cases, also occurring in natural communication (e.g. in the case of alarm calls), signalling behaviours provide an adaptive advantage for the receivers but not direct benefits for the signaller. This lack of an adaptive advantage from the point of view of the signaller, might prevent the preservation of genetic characters that lead to signalling behaviours that are useful for the receivers and for the group as a whole. The second problem is due to the fact that, even in cases in which communication emerges, the evolved strategies are not stable and are easily invaded by mutants that produce different signals. In this condition in fact, mutant's fitness will remain the same while the fitness of the other members of the population, that are unable to correctly interpret mutant's signals, will decrease. This selective advantage gathered by the mutant to the expenses of the other individuals and of the population as a whole will allow mutant individuals to leave more offspring and will consequently lead to the loss of the ability to communicate according to the previously evolved communication system. For a simple demonstration of how communication fails to evolve in a population of disembodied agents in which communication only provides an adaptive advantage for the receivers, see Oliphant (1996). For a demonstration of how the evolutionary dynamics might lead to an instable situation in which an ability to communicate periodically evolves and then is lost due to mutant signallers invading the population, see Batali (1995) and Mirolli and Parisi (in preparation).

As demonstrated in several experimental studies, however, other factors might counter-balance these adaptive problems and might lead to the emergence of a stable communication system. For instance a stable communication system emerges in experiments in which: (1) the population is spatially distributed and individuals are more likely to communicate and mate with those close to them (Oliphant, 1996), (2) the same set of internal neurons of agents' controller determine both the motor and signalling behaviour of the agent and receive both sensory and communicative information (Cangelosi and Parisi, 1998), (3) agents (provided with the same neural architecture described above) receive communication signals only from their parents and are allowed to communicate only after a first evolutionary phase in which they can develop their individual capabilities (Marocco et al., 2003). In any case, although these and other ecological factors (see Di Paolo, 1998; Noble et al., 2002) might counter-balance the lack of direct benefit for signalling and the advantage for individuals to deceive, these two factors will in any case tend to prevent the emergence or the preservation of communication. Indeed, if we compare the experiments described in Cangelosi and Parisi (1998) and Marocco et al. (2003) that differ with respect to the complexity of the problem, we can see that why in the former the constraint on agents' neural architecture were enough, in the latter communication only emerged by also restricting communication acts between parents and by allowing individuals to evolve their individual ability before communicating. The question of how complex communication systems can emerge without a direct benefit for the signaller therefore largely remains an open problem.

Obviously, these adaptive problems do not affect (or at least are much less important) in cases in which communication provides an adaptive benefit for both producers and receivers. This is the case, for example, of mating signals (for an example of how this type of communication might emerge in a population of artificial agents, see Werner and Dyer (1991)).

Finally, these adaptive problems do not affect (or at least are much less important) in cases in which agents are selected on the basis of their collective performance (Baldassarre et al., 2003; Quinn et al., 2003). Interestingly, a similar situation occurs in colonies of some social insects (e.g. in bees) in which most of the individuals are sterile and in which individuals are very genetically related.

For a discussion of coordinated behavior and cooperation in social insects and their relation to selective and behavioral factors see chapter 4 by Floreano and Keller and chapter 6 by Deneubourg.

3.4.3 Social Learning and Culture

Agents might develop an ability to communicate and a shared communication system phylogenetically (i.e. through changes occurring over generations) or ontogenetically (i.e. through changes occurring during agents' lifetime). While in the former case characters that allow communication are encoded genetically and are transmitted and varied during agents' reproduction, in the latter case the characters that allow communication are transmitted and varied through social learning. These two modalities are also referred to with the terms: genetic evolution and cultural transmission or cultural evolution (for an example of how cultural evolution might lead to the emergence of an ability from scratch through variations arising during social imitation and selective reproduction, see (Denaro and Parisi, 1996)). Cultural transmission and evolution plays a central role in human language but it also plays a role in some forms of animal communication (e.g. in monkeys, squirrels, birds etc., see Wagner et al., 2003b). Moreover, when both genetic and cultural factors are present, communication emerges as a result of the interaction between three adaptive processes: genetic evolution, individual learning, and cultural evolution (or social learning) that have different characteristics and operate at different time scales.

The issue of how artificial evolution, online adaptation, and social learning techniques might be effectively combined together is a largely unexplored research area in this field. Indeed, although methods that combine evolutionary and learning algorithms (e.g. evolutionary algorithms with reinforcement learning algorithms or with hebbian learning algorithms) have been already proposed and investigated by several authors (see Nolfi and Floreano, 1999; Nolfi, 2002a), the study of social learning in situated agents is an area that is gathering an increasing research attention but that it is still in its infancy (Lindblom and Ziemke, 2003). The attempt to combine social learning and evolutionary techniques is a largely unexplored area (for a preliminary attempt in this direction and a critique of obtained results see MacLennan and Burghardt (1993); Noble and Cliff (1996)).

Advances in social learning techniques and methods for combining evolutionary and social learning techniques might produce significant insights on how complex forms of communication(s) might emerge from the interaction between situated agents. Indeed, social learning has specific features that might greatly enhance agents' ability to acquire complex skills. As an example of these features we should consider that in social learning agents play two roles (a student role and a teacher role) and consequently might improve both their ability to learn from others and their ability to facilitate other agents' learning. In other words, agents that learn socially might exploit the fact that the social environment with which they interact during learning, unlike the physical environment, has been co-evolved to favour the ability to acquire adaptive skills through learning (at least in the case in which interacting agents have an interest in cooperating). As an example of the advantages that might arise by combining evolutionary and social learning adaptive processes, we should consider socially acquiring skills from different agents allow individuals to combine several adaptive characters independently discovered by different individuals and resulting from both genetic and ontogenetic variations. Genetic assimilation (Baldwin, 1896; Waddington, 1942) might later assure the genetic fixation of characters previously acquired ontogenetically, where appropriate.

3.5 Discussion

The attempt to develop agents able to solve collective problems by cooperating and communicating through a self-organizing process is an extremely ambitious goal. Achieving this goal, in fact, imply to understand which initial conditions might lead to the emergence of a complex behavioural, cognitive, and social abilities. Moreover, the attempt to develop these abilities in embodied and situated agents introduces other important challenges (e.g. the need to deal with noisy and incomplete information, the need to extract regularities by integrating information in time and/or the need to produce sequential behaviours).

Despite this enormous complexity, the promising preliminary results that we reviewed and the possibility to integrate important aspects that are actually studied in isolation in different models indicate that the time is now ripe for investigating this challenging problem without necessarily rely on shortcuts or simplifications (e.g. models in which communication involve the exchange of a predefined list of signals and/or 'meanings' or in which the function of communication is predetermined and fixed).

Evolutionary Pre-requisites for Emergence of Communication

Dario Floreano and Laurent Keller

EPFL and University Lausanne, Switzerland

4.1 Introduction

Communication is based on the establishment of a convention between a signaller and a receiver (Maynard Smith and Harper, 1995). The emergence of such convention requires some degree of coordinated activity and cooperation among individuals. It is therefore necessary to understand the evolutionary conditions that enabled relatively simple (with respect to humans) organisms to display non-trivial levels of coordination and cooperative behaviour, such as ants and honeybees. In this section we review the open questions, challenges, and current understanding on the emergence of cooperation, communication, and division of labour in societies of biological and artificial agents. We also suggest possible experimental methods to tackle such questions and challenges.

Ants compose about 15 percent of the entire animal biomass of most terrestrial environments (Hölldobler and Wilson, 1990). The ecological impact of their diversity and abundance is commensurate with their dominance: they turn more soil than earthworms, and act as keystone species in a wealth of predatory and mutualistic interactions with other organisms (Hölldobler and Wilson, 1990). Without question, the primary feature of their biology that has contributed to their enormous success is their complex social organisation (Wilson, 1971; Seger, 1991; Bourke, 1991; Keller and Chapuisat, 2001), which features coordinated behaviour, signalling, cooperation, altruism, and division of labour. We think that at least the first three features are necessary (although they may not be sufficient) for the emergence of communication in societies of organisms. It is therefore crucial to understand how such complex social organisation evolved in order to infer principles and mechanisms that may be applicable to societies of artificial agents.

Because all types of sophisticated social organisations occur in insects and vertebrates, which invariably have relatively long generation time (Keller and Genoud, 1997; Bourke, 1999), it is impossible to study the origin of complex social organisation by guided evolution (a process where the evolution of a trait can be followed by breeding organisms under different selection regimes). To circumvent this problem, one could use a system of artificial ants implemented as small mobile robots with simple vision and communication abilities that are evolved by means of genetic algorithms (Nolfi and Floreano, 2000). In order to do so, it would be interesting to evolve robot control systems and study their behaviour under two different levels of selection (individual and colony level) and under different types of group structure (high and low relatedness) for reasons that are explained more fully below.

Such experiments should not only help to understand the evolutionary conditions that led to the emergence of complex social organization suitable for communication, but also generate guidelines for the design of autonomous agents capable of efficient cooperation and task self-allocation. So far, researchers have mainly equipped robots with hand-designed mechanisms to achieve emergent cooperation and communication. The use of artificial evolution to synthesize robots able to display collective behaviour and communication is still a rather unexplored area.

The following sections present the state of research in the two main facets of this research endeavour, namely, a) cooperation and division of labour in social insects, and b) cooperative robotics.

4.2 Cooperation and division of labour in social insects

Social insects (ants, wasps, bees and termites) fascinate due to their extreme levels of cooperation and social cohesion. They indeed provide some of the most remarkable examples of altruistic behaviour, with a worker caste whose individuals forego their own reproduction to enhance reproduction of the queen. The level of such worker self-sacrifice can be extreme, as exemplified by the evolution of kamikaze weapons, such as detachable stings and exploding abdomens used in defence of the colony (Wilson, 1971). Workers also collectively exhibit highly organised, sophisticated behaviour that is adaptively fine-tuned to ecological conditions. For example, workers of some ants react to the presence of workers from other colonies, and the heightened risk of conflict, by increasing the production of soldiers, which are specialised in colony defence (Passera et al., 1996). The honeybee waggle dance communication system provides yet another example of the sophistication of the aggregate behaviour of social insect workers (Seeley, 1995). Such examples of group harmony and cooperation have given rise to the concept that colonies are harmonious fortress-factories in which individual-level selection is muted, with the result that colony-level selection reigns. In other words, the colony often appears to behave as a super-organism operating as a functionally integrated unit (Wheeler, 1928). In this vein, Seeley (1997) has described the elegant group-level adaptations of honeybee societies.

However, the concept of a super-organism as being the only unit at which natural selection operates has been challenged both on theoretical grounds and by the observation that life within the colony is not always as harmonious as it may first appear. The more we come to understand the dynamics of life in the society, the more we realise that the colony is far from simply being a super-organism emerging from a nexus of self-sacrificing altruism (Ratnieks and Reeve, 1992; Bourke and Franks, 1996; Crozier and Pamilo, 1996; Bourke, 1999; Keller and Reeve, 1999). Social life may involve conflicts of genetic self-interest, resulting in tactics of coercion, manipulation and even deadly aggression between colony members in the name of genetic self-interest. These conflicts arise because colony members should favour individuals that are more closely related (share more identical gene by common ancestry) to maximise their inclusive fitness (Hamilton, 1964a). Since the pattern of relatedness to a set of individuals differs for different colony members within genetically heterogeneous insect societies, and since the colony has finite amounts of resource to allocate, the stage is set for a multitude of potential conflicts (Hamilton, 1964b). These conflicts, in turn, have negative effects at the colony level because they may decrease the overall productivity.

These costs at the level of the colony are expected to lead to counter-strategies to suppress selfish behaviours (Ratnieks, 1988; Reeve et al., 1997). In other words, the actual conflict should generally be lower than the potential conflict (Ratnieks and Reeve, 1992). Understanding exactly how and to what degree actual conflict is suppressed and how this increases overall group productivity is the key to understanding the extent to which social insect colonies can be viewed as adaptively organized group-level units (Seeley, 1997).

To study the evolutionary outcomes of within colony conflicts it is helpful to use a multi-level selection approach (Keller and Reeve, 1999). Although genes are the entities that are ultimately transmitted over generations, it is important to keep in mind that genes are packaged in organisms, organisms in groups, and groups in populations, and that selection theoretically may act at any of these levels. Selection acting within and between colonies can be analysed by using a multi-level analysis of selection. However, it is very difficult to empirically quantify the selective forces acting at the different levels of organisation of the colony. An alternative approach would be to use guided selection and modify the strength of selection at the individual and colony levels. Unfortunately, this is not possible due to the relatively long generation time of social insects.

To circumvent this problem, one may resort to a new experimental system consisting of colonies of artificial ants implemented as small mobile robots with simple vision and communication abilities. Such a system would allow the investigation of the role played by the level at which selection acts (colony versus individual) as well as of the group composition (relatedness between individuals) on the evolution of communication and altruism. Relatedness is known to have played a major role in favouring the evolution of altruism in social insect (Bourke and Franks, 1996; Sundstrm et al., 1996; Keller and Reeve, 1999) and other animals and such experiments would help to determine whether the role of relatedness can be experimentally demonstrated with artificial agents such as robots engaged in tasks where communication provides a selective advantage.

Such an experimental setup would also allow one to study the evolution of another unusual characteristic of social insects, namely division of labour between workers (the sterile individuals performing almost all the work in the colony). That is, in addition to the reproductive division of labour between queens and workers, work is further partitioned among workers (Wilson, 1971; Bourke and Franks, 1996). Workers can specialise on particular tasks (e.g., brood care, nest construction, nest defence, foraging).

The basic problem faced by a colony is to dynamically allocate the right number of workers to the various tasks. One mechanism for dynamic task allocation has been observed in some species of ants and bees, where workers change their tasks according to age. Young individuals typically work on internal tasks (brood care and nest maintenance), while older individuals forage for food and defend the nest (Wilson, 1971). But recent research has shown that ant workers are usually able to switch tasks when needed, for example if one behavioural caste is experimentally removed (reviewed in Bourke and Franks, 1996). Hence, a forager can become a guard, or a nurse can become a forager independently of their age. This plasticity of individual behaviour allows colonies to adapt to changes in their environment and also to demographic changes in the colony. Colony members might change their behaviour according to a self-organising mechanism where workers perform a task when a specific stimulus for this task exceeds its individual threshold (Deneubourg et al., 1987; Bonabeau et al., 2000; Anderson and McShea, 2001; Anderson, 2002; Gautrais et al., 2002). Tasks and stimulus are linked in a negative feedback loop that regulates the system: when an individual performs a task, it decreases the stimulus for this particular task. To test these hypotheses one should measure whether behavioural plasticity and negative feedback loops can evolve in groups of artificial ants and, if so, determine the conditions that are conducive to their evolution.

4.3 Cooperative robotics

Synthesis and analysis of collective behaviour from individual interactions represent a major challenge in both ethology and artificial intelligence (Bonabeau et al., 1999; Camazine et al., 2001). Accomplishing tasks with a system of multiple robots is appealing because of its analogous relationship with populations of social insects. Researchers argue that cooperating teams of simple robots can accomplish useful tasks that a single robot could not possibly do. In a survey, Cao et al. (1997) defined cooperative behaviour as follows: given some task specified by a designer, a multiple-robot system displays cooperative behaviour if, due to some underlying mechanism (i.e., the mechanism of cooperation), there is an increase in the total utility of the system.

We can distinguish three mechanisms that have been used so far for achieving cooperation among robots: a) self-organization (emergent cooperation), b) explicit cooperation through task-related communication (intentional model of cooperation), and c) cooperation by learning.

a) Emergent cooperation. Interest has focussed on how to develop a decentralized approach to control a multi-robot system without explicit communication among the robots. The hypothesis is that a decentralized, non-communicating system should scale more easily with the number of robots (Kube and Zhang, 1993). Cooperation here emerges according to a principle where a robot's action is determined or influenced by the consequences of another robot's previous action, similar to the phenomenon of *stigmergy*

(Bonabeau et al., 1999). Experiments with robots have successfully exploited this principle to perform a number of tasks. For instance, Holland and Melhuish (1999) used a group of physically-identical robots to cluster and sort objects of two different types. Using a behavioural rule set much simpler than those proposed so far, and having no capacity for spatial orientation or memory, robots were able to achieve effective clustering and sorting, similar to ant brood sorting where ants sort their brood so that items at similar stages of development are grouped together. In another experiment, Kube and Bonabeau (2000) worked on a transport task, where the objective was to locate a brightly-lit box and move it to a goal location. The box was weighted such that at least two robots were needed to move the box. The robots did not explicitly communicate and were not centrally controlled. Although an optimal solution was not found, the robots always managed to push the box towards the goal. Finally, a third experiment addressed a stick-pulling task (Ijspeert et al., 2001 after Martinoli and Mondada, 1995). The objective is to locate sticks in a circular arena and to pull them out of a hole in the ground. Because of the length of a stick, a single robot is not capable of pulling it out alone. Collaboration between two robots is thus necessary for pulling a stick completely out. By providing robots with a simple time latency, cooperative stick-pulling emerged without explicit communication and coordination as a probabilistic phenomenon depending on the size of the working area and the number of robots.

b) Explicit (intentional) cooperation. Gerkey and Matarić (2002) developed a model where robots cooperate explicitly and with purpose, often through task-related communication. The authors claim that intentional cooperation is better suited than emergent cooperation for tasks that humans would like robots to perform. They presented a novel method of dynamic task allocation for multi-robot systems based on simple auctions to allocate tasks. The results indicated that the system could take into account various changes in the environment. When compared to emergent cooperation, this approach requires a bigger overhead at the level of communication.

c) Cooperative robot learning. Some researchers have explored the idea of providing robots with learning capabilities to coordinate robot interactions and improve team performance. However, learning in physically embedded robots is known to be a difficult problem, due to sensor and actuator uncertainty, partial observability of the robot's own environment, and dynamic properties of the environment, especially when multiple learners are involved (Mataric, 1998). Moreover, most learning approaches used in cooperative robotics do not appear to be scalable, because they imply for each learning robot a state-space growth exponential in the number of team members. Touzet (2000) recently proposed to exploit robot awareness, defined as the perception of the locations and actions of other robots, in order to improve cooperative learning. However, in real environments (as opposed to simulations) this approach requires a reliable radio communication and rich information on the state of each robot in the team.

Experimental studies indicate that another important factor in the emergence of cooperation among robots is the composition of the group. Two major factors are a) the size of the group and b) the physical and behavioural diversity of individual robots.

Krieger et al. (2000) studied the effect of group size in teams of robots programmed with ant-inspired algorithms in a foraging task. To determine the relationships between group size and efficiency, they compared groups of 1, 3, 6, 9 and 12 robots. Groups of intermediate size performed better, which is probably due to a trade-off between the positive and negative effects of robot-robot interactions. Because robots were programmed to avoid each other, groups of robots exhibited a more efficient coverage of the foraging arena than single robots. When the number of robots increased, however, negative interactions among robots (here defined as robots interfering with each other when trying to perform a task) also increased. A similar conclusion was recently reached by Lerman et al. (2002) in the case of the stick-pulling task mentioned above.

Balch and Parker (2000) argued that heterogeneity is an important focus of multi-robot systems research; one of the most compelling is the observation that it is nearly impossible in practice to build a truly homogeneous robot teams because several copies of the same model of robot can vary widely in capabilities due to differences in sensor tuning, calibration, etc. Based on this hypothesis, the Alliance architecture (Parker, 1998) has been proposed as an approach to dynamically assigning tasks to different members of a robot groups and was demonstrated on a group of robots dividing a clean-up task. Parker used a group of robots with a priori hard-wired heterogeneous capabilities (Parker, 1994).

Finally, Fontan and Mataric (1998) addressed both group size and group diversity in a problem of dynamic task assignment (see also Anderson and Ratnieks, 1999; Labella et al., 2004). They studied a territorial approach to a task where robots are assigned individual territories that can be dynamically resized if one of the robots fails. Using a collection of experimental robot data, they reported a decline of performance of space division strategy with increasing group sizes. Similar to the results obtained by Krieger et al. (2000), a medium-size group of robots was the most efficient choice given the trade-off between interference and workload in the particular territorial division.

To summarize, the prevailing approach in collective robotics is to provide robots with a set of predefined algorithms for cooperation and communication and observe the team performance by varying environmental variables and/or team size. The question of the relationship between team similarity (physical and behavioural) and emergence of cooperative behaviour and communication remains still open. The same applies to the emergence of division of labour in groups of robots. To the best of our knowledge, there have not yet been attempts to use an evolutionary approach to investigate these issues in a systematic framework. Cooperation, division of labour, and communication are major challenges in collective robotics where teams of robots are expected to autonomously coordinate in order to carry out tasks that a single robot could not do.

Game-Theoretic Challenges

Peter Hammerstein

Humbold University Berlin

5.1 Introduction

Modern evolutionary biological theory is, in large part, a theory of conflicts of interest and how they are managed. Conflicts exist among, and even within, organisms. Gene-level conflict arises, for example, if some genes could increase their rate of reproduction at the expense of others. For example, gene-level conflict can be caused by differing modes of reproduction. If some genetic elements can only reproduce via the egg cytoplasm, these elements will have little 'interest' in male reproduction.

Organism-level conflict arises when organisms compete for the same, limited resources, like food or mates. Alleles that code for successful competitive strategies will out-reproduce alternative alleles. Organisms will therefore tend to acquire abilities and strategies for competition both within, and across, species. Conflicts of interest are deeply embedded in the very foundations of life.

Important transitions in evolution occur when strategies arise that can overcome these conflicts (e.g., Maynard Smith and Szathmáry, 1997). Modern biology has focused considerable effort on formulating theories that can explain cooperation and reliable communication despite inherent conflicts of interest—biologists *assume* conflict, and then investigate how nature has overcome it. A biological theory is considered a success when it successfully explains how organisms manage conflict, even if many other details of organism interaction are left unexplained.

IT infrastructure, on the other hand, is often not designed with conflicts of interest as a founding principle. Instead, the infrastructure is assumed to serve the needs of multiple parties, and measures to manage conflict, like passwords and encryption, are added after the fact. IT infrastructure is engineered to efficiently solve problems like reliable and efficient communication in large networks of increasingly heterogeneous agents. These are exactly the problems that behavioral ecologists in biology tend to sweep under the rug—if conflicts of interest can be overcome, then the remaining problems are simply engineering 'details'.

We believe biologists studying communication and cooperation have a lot to learn from IT research in these areas. IT, on the other hand, might also be able to learn something from biologists' obsessive concern with conflicts of interest. The internet, for example, is ridden with viruses, worms, and other malware that threaten its utility to the public.

Here we outline what we consider to be potentially important applications of biological theory to the design and implementation of embodied, communicating agents that are robust (but not entirely immune) to parasitism, deception, and free-riding. We focus not on mechanistic details, but on broad, strategic insights.

5.2 Challenge I: Avoiding infection—Immune system analogies

Organisms harvest energy and other resources from the environment, converting them into useable form. Because organisms are concentrated packets of energy and other nutrients, and because organisms can also provide other benefits, like transportation and a regulated environment, they themselves represent an extremely valuable resource that other organisms could exploit. Indeed, virtually no complex, multi-cellular organism is pathogen free. But, pathogens damage their hosts, and impose other costs, so there has been a strong selection pressure to deter and eliminate infections.

5.2.1 Self identity and cross-agent diversity

There are several important properties of organisms that deter parasites. The following are two of the most important of these. First, many complex organisms have an ability to identify self. This allows an organism to identify and eliminate non-self, thus disposing of many parasites and toxins. Second, in many species, individual organisms differ somewhat from one another at the molecular level. This decreases the ability of parasites to efficiently exploit the organism, because what works well in one organism might not work so well in another with a somewhat different composition. A single strategy would have difficulty successfully infecting and exploiting multiple hosts. The second property means that organisms have to *learn* to identify themselves (because they will differ from their parents).

Ideally, an organism would have a perfect ability to identify self, and despite functioning similarly, or identically, to other members of its species, would differ substantially from them in its 'implementation' details. Real organisms achieve these goals in some of the following ways. Gene pools typically contain a significant degree of variation. This means that several alleles at each locus produce proteins with similar functionality, but somewhat different structures. Sexual recombination results in offspring with a random assortment of these different alleles. Thus, when an individual learns to identify itself, its protein 'fingerprint' is different from that of every other member of its population. Parasites find it difficult to mimic host protein patterns (which would allow them to evade the immune system), because there is no single pattern in the population. In addition, the protein structure of each individual is unique, so infection and exploitation strategies cannot hone themselves to a uniform host environment.

Here we sketch one way to achieve similar properties in EC agents. First, for the sake of discussion, assume an agent contains 30 software modules with predefined functions. Code each of these modules to spec using independent teams for each module, each team using a different programming language (e.g., C#

and Java) and/or different compilers or virtual machines. This results in 30 pairs of modules, with each module in a pair performing the same function, but using code that is completely different, at least at the binary level, but also perhaps at higher levels. Form each EC agent by randomly selecting one module from each pair, for each of the 30 pairs. Each EC agent will then function similarly, but will contain a unique combination of modules (30 pairs of modules can be combined in about 1 billion different ways). Each EC agent can then learn what its unique combination of modules is, either by computing a hash on the module binaries, by analyzing subtleties of module behaviour (modules should function almost, but perhaps not quite, identically), or by other means.

However it is achieved, EC agents must have a robust ability to identify self. This will become especially important if EC agents reproduce. Self reproduction means that it will be difficult, if not impossible, to ensure that agents are initially infection-free (because agents will not be produced in 'pristine' environments). Agent parents might have to provide offspring with 'antibodies' to parasites they themselves are infected with. EC agents must also differ substantially from each other at the implementation level, however closely they resemble themselves at the functional level. For similar work on immune system analogies in IT security, see, e.g., Kephart (1994), Somayaji et al. (1997), Somayaji and Forrest (2000), and Forrest et al. (1997).

5.2.2 Virulence

Keeping EC agents free from infection might be an unreasonable goal. Infection with 'mild' strains of pathogens might even be beneficial to the extent that they exclude more virulent strains. There are a number of arguments and counter arguments regarding the evolution of virulence (see Ewald, 1994 for one influential view). Almost all agree that parasites will evolve traits that maximize their own reproduction. In some cases, this will involve a devastating exploitation of the host, quickly leading to host death. In other cases, host exploitation is minimal. And in yet other cases, parasites and hosts co-evolve a mutually beneficial symbiosis. If parasite transmission requires proper host functioning, then the ecology of parasites infecting EC agents will necessarily tend to exclude those that seriously impair EC agents. EC agent communication systems should therefore carefully consider the modes of communication. If a communicative act can transmit, e.g., code, and this code could become infected, then it is important that communication require that the EC agent exercise many of its functional abilities. Here is a simple illustration of this idea. If an EC agent transmitted information by executing a complex dance, then any parasite replicating via this channel could not significantly impair the ability to dance. If dancing ability were dependent on functionality that was essential for other key tasks, like obtaining energy, then parasites could not seriously impair these key abilities because by doing so they would impair the ability to dance, and therefore their own ability to reproduce. By tightly linking modes of infection (e.g., communication channels) with other essential functions, populations of EC agents should be able to tolerate a broad spectrum of infections.

5.3 Challenge II: Signaling and communication when there are conflicts of interest

A signal is "an action or structure that increases the fitness of an individual by altering the behaviour of other organisms detecting it, and that has characteristics that have evolved because they have that effect" (Maynard Smith and Szathmáry, 1997). Crucially, signals evolve because they provide fitness benefits to the signaler, not to the recipient (Dawkins and R., 1978). Between-organism signals therefore often contain a substantial dose of deception. Mimicry and crypsis, for example, are extremely common in vertebrates, arthropods and opisthobranch gastropods (Starrett, 1993). EC agents may wish to employ various forms of deception, like camouflaging themselves to avoid danger, and they will certainly need mechanisms for detecting deception.

Deceptive signals select for ever better signal discrimination on the part of signal receivers, which selects for more effectively deceptive signaling, and so on. Deceptive signaling/detection arms races need not always characterize signaling systems, however. A number of mechanisms exist whereby honest signaling can evolve. The best known is 'costly signaling' (Spence, 1973; Zahavi, 1975). If a signal has an inherent fitness cost, then signals can only be sent by individuals who can afford the signal. The signal is then an honest indicator of the quality of the sender. Male red deer, for example, engage in costly bouts of roaring when competing with other males. Because roaring is physiologically taxing, animals in poor condition cannot roar at the same rate as animals in good condition. The roaring rate is an honest indicator of body condition, and thus fighting ability (Clutton-Brock and Albon, 1979).

More generally, if a signal has a necessary correlation with a property of the sender, then the signal is an honest indicator of this property, even if the signal is not costly (Maynard Smith and Szathmáry, 1997). Funnel web spiders, *Agelenopsis aperta*, fight over web sites (Riechart, 1978). Spiders contesting on a web vibrate the web, and the vibrations transmit reliable information about relative body mass. If the difference in body mass exceeds about 10%, the smaller spider withdraws (Hammerstein and E., 1988).

The essence of honest signaling theory is what is variously termed private, incomplete, or asymmetrical information—some participants in a social interaction have difficult-to-observe qualities that are critical to the decision-making of other participants. One example of private information that could be potentially important to EC agents is whether or not they are infected with a virus. If so, other agents might not want to interact with them. EC agents might only interact with other agents that are able to demonstrate that they do not suffer any major infection. If infections impair functionality, then demonstrating a large range of complex functionality would reassure potential social partners that no major infections were present. Returning to the dance example, a complex dance that exercised numerous abilities with precise timing could honestly signal that no major, disabling infection was present. After successful dance displays by potential communication partners, communication could then commence with (relative) confidence.

Honest signals can also evolve when there is little or no conflict of interest between sender and receiver, or when sender and receiver have a mutual interest in coordination. Signalling systems in eusocial insects like ants and bees are typical examples. In these cases, the signals should have exceptionally low fitness costs (Markl, 1985). Such 'cheap' signals can also evolve when there are future interactions between sender and receiver, and when the receiver can punish false or deceptive signals by, e.g., attacking the sender of false signals or defecting from repeated future cooperative interactions to the deficit of the sender (e.g., Silk et al., 2000; and references therein). In addition, low-cost signals can evolve when signalers receive no benefits from deceiving their partners (Dawkins and Guilford, 1994).

Human language is a highly sophisticated system for 'cheap' signalling—speaking imposes few costs on the speaker. Further, there is no necessary correlation between properties of the speaker and the semantic content of speech. The existence of language therefore testifies to an evolutionary history in which interactions among individuals commonly involved few conflicts of interest, depended heavily on coordination, or were repeated with opportunities to punish lying. More likely, some combination of these factors permitted the evolution of language (Lachmann et al., 2001; Silk et al., 2000).

5.3.1 Lies, exaggerations, and misrepresentations

Although most speech should be basically honest, speech can be deceptive when "on average, the incentive to the signaller to misrepresent the state of the world [is] outweighed by the incentive not to do so" (Lachmann et al., 2001, p. 13189). Lying can confer social benefits, such as misleading competitors, extracting additional resources from social partners, or avoiding punishment for proscribed behaviour.

There are obviously strong selection pressures for the evolution of mechanisms to discriminate real signals from deceptive and erroneous signals. These mechanisms should attend to cues that correlate with signal veracity. IT has access to very sophisticated error detection and correction mechanisms. Deception detection mechanisms, however, are another matter.

Because there are incentives to manipulate information about the social world for personal benefit, mere repetition of a signal from the same, potentially deceptive, source does not decrease the probability of deception in the signal. Receiving the same information from multiple, independent sources, however, decreases the probability of deception for several reasons. First, if an individual manipulates information to serve his or her own individual interest, then the probability that other signallers share exactly the same interest, and thus an incentive to manipulate their signals in exactly the same way, will decrease as the number of signallers increases. Second, deception imposes a severe coordination problem on multiple deceivers. Whereas there is only one truth, there are many different deceptions that can serve the same end. Multiple sources who share exactly the same interest in deception could each generate radically different stories serving this interest. Without sophisticated, costly coordination, it would be very difficult to keep their stories straight. Third, lying invites punishment. Multiple signallers colluding to deceive would also have to share the same willingness to risk punishment for lying. The probability that multiple signallers are equally willing to suffer punishment declines as the number of signallers increases. Thus, receiving information from multiple sources should be a reliable cue of veracity.

Deceivers do share interests, and can coordinate stories, so receivers should attend to cues that multiple sources of gossip are independent—that is, one gossiper did not simply receive the information from another, and, importantly, the one gossiper does not have an obvious incentive to coordinate a deception with another gossiper. EC agents will need to be able to detect cues that signal the sources of information.

Finally, the cost of punishment implies that gossipers should only deceive if they have an incentive to do so. Gossip receivers should therefore have mechanisms that attend to cues that gossipers have ulterior motives, like competition for resources.

5.4 Challenge III: Cooperation when there are conflicts of interest

Competing organisms can often mutually benefit by engaging in various forms of cooperation. Cooperation should not be seen as an alternative to competition, but instead as a competitive strategy. Conflicts of interest present severe problems for cooperative ventures, like the division of benefits and free-riding. Consequently, biologists have invested enormous effort investigating strategies which solve these problems.

5.4.1 Mutualism

One of the simplest and most effective solutions is what we will here call mutualism: a cooperative goal can only be reached by the mutual, concurrent actions of participants, and all participants benefit once the goal is reached. Imagine, for example, a group of individuals in a boat that would all benefit if the boat could reach a distant island before other boats did (perhaps because there is buried treasure on the island). There is no division of benefit problem, because all benefit if the boat reaches the island first, and there is no free-rider problem if all must exert maximum effort to have any chance of reaching the island first. This is the solution 'adopted' by the genome. Potentially competing genes all benefit equally if the organism—which they mutually construct—out reproduces other organisms, and the organism has the greatest chance of out-reproducing others if the genes each contribute maximally to the reproductive capabilities of the organism. Further, this solution permits substantial efficiencies to be realized by division of labour. Just as the boat would have the best chance of reaching the island first if each person took on tasks that best matched their abilities—the strongest taking over the oars, those with the best vision navigating the boat—so, too, do genes benefit by each contributing different capabilities to the organism (genes for eyes, genes for muscles, etc.). EC agents with potential conflicts of interest would often benefit if cooperative tasks were structured so that all agents on a 'team' were 'stuck in the same boat', as it were.

5.4.2 Bargaining

Sometimes even if one is stuck in the same boat for a time, one won't be stuck there forever. In this case, division of benefits can be a problem. When one's boat wins the race to treasure island, for example, the treasure must still be divided. Perhaps the strong will now assert their power to take most of the treasure. There are situations, however, where even the weak can get their fair share.

Viscous social markets and monopoly power. When there are many resource providers (when there is a market instead of a monopoly), one has little need to pay a premium to social partners because one can always obtain the necessary resources elsewhere (resource costs are then determined by the supply and demand curves of standard economic theory). If, however, there is a high degree of mutual dependence, and the cost of switching partners is high, a fair division of benefits can be accomplished by bargaining. By withholding services that are needed by others, individuals can compel a fair division of benefits. In the boat example, if individuals need to not only reach the island first, but then escape before others arrive, individuals who do not receive a fair share of the treasure can threaten to withhold their services unless the treasure is fairly divided. Since no one will escape the island with the treasure unless all cooperate, there is a strong incentive to fairly divide the treasure.

Such bargaining is necessary and effective when 1) at least one participant is not benefiting from the current social contract, 2) others are benefiting from the social contract, and 3) participants have a monopoly or near monopoly on the resources they provide—otherwise, disaffected parties could simply choose to cooperate with someone else (see Kennan and Wilson, 1993 for a review).

Private information and credible signalling: the function of delay. When the value of cooperation decreases with time, withholding benefits can also credibly signal that one truly is suffering costs to those who might not otherwise recognize those costs. It is difficult for group members to accurately assess the costs and benefits incurred by their social partners: she claims she is not benefiting from a relationship, but perhaps she really is and just wants more than her fair share; her true valuation is private information.

The discount factor, ?, is the fraction of cooperative benefits still available after each round of bargaining, and is thus a measure of the costs of delay due to multiple rounds of bargaining. Kennan and Wilson (1993) argue that quick agreements are usually possible in most models of bargaining where valuations and discount factors are common knowledge (i.e., no private information). Informally, if each participant knows what the other participants know, each will come to the same conclusions about how any sequence of bargaining rounds will proceed; each participant will also come to the same conclusions about the 'optimal' outcome for other participants, and so this outcome can be offered in the first round. In a simple game of alternating offers by a buyer and seller, if 0<?<1, then the maximum benefit decreases as $?^t$, where t = the number of rounds, so the seller must make an offer just sufficiently generous such that the buyer cannot do better by waiting another round—when delay is costly, each party has an incentive to minimize the number of rounds of bargaining in order to maximize benefits. It can be shown that if the seller makes the first offer, she will offer a price that gives her 1/(1+?) of the benefits, which the buyer accepts immediately (Rubinstein, 1982).

If, on the other hand, the participants in a cooperative venture do not know how other participants value the potential benefits or the costs they will suffer from delays, it will be impossible for all participants to reach the same conclusion about the 'optimal' agreement. If participants could credibly signal to other participants their true valuations and discount factors, then an agreement could be reached. Kennan and Wilson (1993) argue that the willingness of a participant to suffer the costs of multiple rounds of bargaining (due to discount factors less than one), coupled with the sizes of the offers made each round, represents credible information about that participant's true valuation—a greater willingness to delay signals lower valuations (because the more valuable the potential benefits from cooperation are to a participant, the less she can afford to delay). Once each participant acquires a relative level of certainty about the other participants' private valuation by observing their willingness to incur delays, the bargaining game becomes equivalent to one where valuations and discount factors are public knowledge, and an agreement can be quickly reached.

In conflicts over division of benefits, EC agents may well have to bargain by withholding services until an equitable deal is reached.

5.4.3 Reciprocal altruism

Cooperative benefits can often be realized via gains in trade. If, for example, two individuals have resources that are more valuable to the other than to themselves both benefit by trading these resources. But, if neither can guarantee that the other will reciprocate a transfer of a resource, there is a dilemma (the famous prisoner's dilemma): it is in the interest of each to accept a resource but not to provide one in return (i.e., defect or cheat); therefore neither offers a resource to the other; therefore both are worse off then if they had traded. If these individuals could guarantee that the other would reciprocate a resource transfer, both could realize gains from trade. The best known such 'partner control' mechanism is TIT-FOR-TAT: If there are indefinitely repeated opportunities to reciprocate, then individuals can garner most of the gains of trade while avoiding most of the costs of cheating if they cooperate (transfer a resource) on the first round, and then, in all future rounds simply copy their social partner's behaviour in the previous round. A major problem of TIT-FOR-TAT is that it does not scale to even moderately large social groups. The problem is the following. The only response to defection by a social partner is to defect oneself. Thus, in a group of size n, if one person defects, all must defect. But, as n increases, the probability that there will be at least one defector increases significantly. Thus, with even moderately large groups, there is a high chance that there will be at least one defector in the group, so everyone defects, preventing cooperation. TIT-FOR-TAT will only allow small groups of EC agents with potential conflicts of interest to cooperate.

5.4.4 Punishment

Many attempts have been made to discover partner control mechanisms that can scale to large groups. One such mechanism is punishment. If defectors can be punished, and if they respond to punishment by then cooperating, cooperation can be sustained in large groups. The reason is that, unlike TIT-FOR-TAT where the only response is to defect from the entire group (in effect, punishing the entire group for the defection of one individual), punishment can be directed at the defector alone. Regrettably, punishment also has some serious problems. If punishment is costly, then which group member should pay the cost of punishment? He who punishes does everyone else a favour that is not repaid. Most solutions to this problem have invoked some sort of group selection.

5.4.5 Costly signalling

Costly signalling is another solution to cooperation like the widespread food sharing seen in many hunter-gatherer groups. Perhaps individuals provide benefits to others as a costly and therefore honest signal of some hidden quality (Gintis et al., 2001; Hawkes, 1991; Smith and Bliege-Bird, 2000). If the hunter who kills the dangerous buffalo gets to marry the prettiest girl in the group, then it is in his interest to kill the buffalo, feeding everyone, even if others do not reciprocate by killing buffalo in the future. Although this strategy can scale to large social groups, the problem with this explanation for cooperation is that it does not explain why the hunter shares the meat with anyone; neither even why it is that he hunts. He could equally well signal his skills by killing the buffalo but keeping the meat for himself, or by engaging in some other dangerous task that benefited no one.

5.4.6 Reputation

Reputation is central to several theories of human sociality. In the indirect reciprocity theories (Leimar and Hammerstein, 2001; Nowak and Sigmund, 1998), benefits are provided to an individual based on information about his or her past contributions to others in the group—generous individuals are rewarded by receiving benefits from group members. This strategy can scale to large social groups. In the 'health-insurance' theories (Gurven et al., 2000; Sugiyama and Chacon, 2000), individuals increase the likelihood that they will be taken care of when ill or injured by generously providing benefits to group members when they are well. In the 'show-off' or 'costly-signalling' theories discussed above, individuals engage in behaviour, such as big-game hunting, that signals their quality as mates or social partners, and consequently reap valuable mating or social benefits.

Reputation can also play an important role in reciprocal altruism models (e.g., Cox et al., 1999; Enquist and Leimar, 1993; Pollock and Dugatkin, 1992). In these models, individuals learn whether future social partners previously defected or cooperated with other social partners, and benefit from this knowledge. In more sophisticated versions of these models, if the values of benefits that individuals provide vary, then individuals should attempt to cooperate with those who can provide the greatest benefits at the lowest cost.

In each of these models, as several authors have noted (e.g., Enquist and Leimar, 1993; Leimar and Hammerstein, 2001, *information* about key behaviours (such as generosity to others or a successful hunting expedition) must be reliably transmitted to group members. In other words, one must achieve and maintain a reputation for being able to provide valuable benefits to others in order to maximize the benefits one acquires from others. This process requires that information about one's capabilities be transmitted among other group members. Although direct observations are obviously informative in the indirect reciprocity models, key behaviours may also be communicated to other group members by the few observers of individual acts of generosity. The show-off/costly signalling and health insurance models assume that the key behaviours will be directly observed by those who ultimately provide benefits. But with models too, most group members will not directly observe who killed the buffalo, but will have to rely on reports (as well as seeing the dead buffalo) to properly assign credit to the successful hunter or hunters. Further, although the health insurance models posit that beneficiaries of past generosity will have a fitness interest in caring for providers when they are injured, it would be reasonable to extend this model in the following way: it would be in the fitness interests of all *potential* beneficiaries to care for an injured provider (even if some had not been personal beneficiaries in the past), because they could benefit from the future generosity of the provider when she is well. In this extended version, information about individual acts of generosity must be transmitted to other group members by observers of these acts.

Both theory and empirical studies suggest that, within groups, access to resources provided by others is mediated by reputation. This strongly implies that in order to sustain cooperation, groups of EC agents with conflicts of interest will have to gossip. But, this raises the specter that agents might manipulate gossip to increase their own reputations and decrease those of their competitors. Agents must also have mechanisms to verify gossip (see above section on lies, exaggerations, and misrepresentations).

5.5 Challenge IV: Identity

Most of the signalling and partner control mechanisms, especially those that involve reputation, rely on being able to reliably identify individuals. In humans, one's identity is most reliably signalled by one's face. Although human faces are extremely similar, humans have also evolved to detect subtle but stable differences that uniquely identify individuals. These differences arise from the genetic differences discussed earlier. All humans differ somewhat from each other genetically, and these differences manifest themselves via ontogeny as stable differences in facial features. Providing EC agents with stable and difficult to fake identities will be one of the most important and most challenging tasks for the project. If there are conflicts of interest, cooperation will simply not be possible without reliable identification procedures. As it has in humans, the cross-agent diversity discussed earlier might provide a means to uniquely identify individuals. For example, some function of an agent's unique combination of modules might serve to identify that individual. To fake another's identity, an agent would need access to the other's entire set of modules. Because it may often be in the interest of agents to fake their identities, much of the burden of identification will reside with the consumers of this information.

In sum, EC agents with conflicts of interest will find it difficult to either communicate or cooperate without addressing the problems described here. We believe these issues must be addressed at every phase and level of EC agent design.

Bio-inspired Self-organized Communicating Agents

Jose Halloy, Jesus Millor, Claire Detrain, and Jean-Louis Deneubourg

Universite Libre de Bruxelles

6.1 Introduction

One of the challenges of ECagents is building groups of artificial agents. Classically, in man-made systems, problem-solving is based on the "Knowledge" of a central unit which must take the decisions and collect all necessary information. However an alternative method is extensively used in nature: the method of collective intelligence governed by distributed intelligence. The study of socially living organisms shows how these groups solve many problems optimally. In such systems, which may be made of a large number of units, the problems are collectively self-solved in real time through the simple behaviour of the units, which interact with each other and with the environment. The units are mixed with the environment and the groups exhibit organizational structures that are functional, robust, and adaptive (Detrain and Deneubourg, 2002). In such systems, imperfect or local information and randomness are essential ingredients.

Our fundamental hypothesis is that universal rules, at the correct level of description, govern natural and man-made systems and similar "algorithms" govern the behaviour of artificial or natural agents performing similar tasks in similar contexts. Among the different organizational schemes found in natural systems, we will focus on those that involve self-organization(see explained definition below) and where behavioural positive feed-backs are the keystones of the organization (Deneubourg and Goss, 1989; Parrish and Edelstein-Keshet, 1999; Detrain et al., 1999; Camazine et al., 2001; Hemelrijk, 2002; Couzin and Krause, 2003). Distributing the team within the environment of the problem to be solved and introducing these positive feed-backs interactions between the units allows the amplification of localised information found by one or a few of the units. Thus thanks to this type of coordination, the team reaction to these local signals is the solution to the problem. While no individual is aware of all the possible alternatives, and no individual possess an explicitly programmed solution, all together they reach an "unconscious" decision. The relevance or the meaning of the behavioural pattern is not found at the individual level but at the collective level. We defined this process as *functional self-organization* (Aron et al., 1990).

In particular, we will discuss two research challenges: the study of generic rules governing such systems and the transition from solitary to cooperative problems solving.

6.2 Brief Presentation of Concepts and Definitions

To avoid the confusion that still surrounds "complex systems" and related fields, we list below the definitions of the basic concepts we use in our research projects. Besides complexity science still presents a rich variety of frameworks or approaches which favours different or subtle variations in the definitions or the concepts. Of course, in a scientific context, the word "complexity" is not a synonym of "complicated" or large system controlled by too many parameters or variables.

6.2.1 Emergent Behaviour and Self-organization

By emergent behaviour or self-organisation we mean a collective behaviour that is not explicitly programmed in each individual but emerge at the level of the group from the numerous interactions between these individuals that only follow local rules (no map, no global representation) (Camazine et al., 2001). By "not explicitly" we mean that the behavioural rules do not refer to the global pattern that will emerge. Like molecules do not "refer" to oscillations or a pacemaker when they react and produce a chemical clock. It does not necessarily imply a large number of individuals but rather a large number of interactions and actions between the individuals and the environment. In our experimental work, we found self-organized behavioural patterns performed by small groups of animals not larger than 10 individuals. Moreover, the understanding of the emergent pattern must be based on the detailed study of the individual properties AND their interactions not only between themselves but also with the environment.For us, the opposition between a so-called reductionism and holism is meaningless.

6.2.2 Randomness: The Essential Ingredient

Individual actions include a level of *intrinsic* randomness. Like moving randomly or behaving in a probabilistic way, an action is never certain but has an intrinsic probability of occurring. The behaviour of each individual becomes then unpredictable. Randomness and fluctuations (also called noise) play an important role in allowing the system to find optimal solutions. This optimality is largely due to a balance between the fluctuations leading to innovations and the accuracy of the communications or behaviours (Deneubourg et al., 1983; Pasteels et al., 1987; Nicolis et al., 2003).

6.2.3 Predictability

The global outcome of population presenting emergent behaviour is certain in well characterised systems and in a normal context i.e. in the absence of catastrophe like unexpected rapid and dramatic change in the environment. Ants do bring food home or they simply die! Because often the system presents multiple possible states coexisting for the same conditions, the specific solutions that accomplish the global behaviour at the level of the group are statistically predictable. For instance the optimal solution to solve a problem is chosen in 90% of the cases while a less optimal solution is selected in 10% of the cases. Nevertheless, the problem is solved in 100% of the cases! The discussion is then shifted towards knowing if 10% of suboptimal behaviour is acceptable and not if the global outcome is predictable. To a large extend, the unpredictability perceived in this context often comes from either the *lack of knowledge* of the system or its *uncontrolled evolvability* at lower time scale. If, in an IT context, engineers *want to design* systems presenting autonomous collective behaviour they can *control* the needed level of predictability.

6.2.4 Evolution and Emergent Behaviour

Emergent behaviour is not an equivalent of evolution or even a necessity for evolution to take place. Emergent behaviour does not produce, in itself, new individual behavioural or signals. There is no contradiction or even competition between self-organisation and natural selection in evolution. On the contrary, evolution makes use of the properties of emergent behaviour by evolving the local rules that will produce new behaviours at the level of the population.

The time scales of emergent behaviour and evolution are completely different: the first takes place in a short time while the latter require a much longer time. In other words, ants use emergent behaviour, for example to bring food home today, while evolution is changing these local rules of this specific emergent behaviour to produce new species of ants (Camazine et al., 2001). For an account on how artificial evolution can be used to develop embodied andcommunicating Agents (see chapter 3).

6.3 Self-organisation in Living and Social Systems

Actually, a limited number of simple generic rules are at work in biological systems (from the cellular level to animal societies) and produce optimal emergent collective patterns for resources and work allocation, social differentiation, synchronisation or de-synchronisation without external pacemaker, clustering and sorting. These simple rules become building blocks for higher collective complexity.

What is remarkable is the *simplicity and parsimony* of these rules that allows solving a great variety of the problems encountered by populations that cannot resort only to a centralised organisation. Nevertheless, the implementations of these rules are a real challenge. In biological systems animals are not simple machines and the physiology that produce such building block behaviours is highly sophisticated and a scientific challenge to identify. In artificial systems, even if the level of animal sophistication is *absolutely not necessary*, their implementation of behavioural rules still represents a technological challenge. Below we list important features of agent populations presenting emergent behaviour:

– Dynamical systems with a large number of events.

– Descriptions based on models with a limited number of parameters are possible. Models make use of different tools like differential equations, cellular or Boolean automata, stochastic simulations, etc.

– The size of the population plays an important role. For the same set of behavioural rules different collective responses are observed as a function of the population size (see e.g. Detrain et al., 1991; Beekman et al., 2001). In living systems, the organisation plan or the behavioural rules may change as a function of the population size. It implies that in artificial complex systems, scalability is an issue that has to be included into the design according to the size of the population.

– The characteristics of communication play an important role. The range of communication (i.e. all to all, next neighbours, etc.) changes the pattern. Privileged linked between some agents (network of interactions) are important features. The lifetime of the communication signals is also an important factor as regards flexibility and/or persistence of the emerging patterns (Detrain et al., 2001).

– The depletion of the resources suppress resources may play an important role. Usually the negative feedbacks that are involved in these dynamics results of the depletion of resources produced by the activity. It can be an ingredient that helps in finding optimal solutions.

– With local knowledge, these systems are capable of adaptive self-reconfiguration or of producing a diversity of responses that do not need to be explicitly coded at the individual level.

– It is not so much the internal structure of the agents but rather the agent/environment and agent/agent interactions that may produce useful results. Through the dynamics of the interactions and without modifying the behavioural algorithms, groups of agents may adopt very different adaptive responses. This is clearly an important property of embodied systems that illustrate the synergy between the physics of a problem and the cognitive capacity of the individuals.

– Randomness is a positive ingredient to find optimal solutions. Randomness or noise is an intrinsic component of real-world systems. It is modulated in living systems and should be tuned in artificial systems. In many systems the randomness is important to find the optimal solutions.

– Biological systems are not fully self-organised complex systems, they present a mix between centralised and distributed "management". There is a balance between treatment of information at the individual and at the global level. Specialisation may pre-exist in the systems and affect the emergent behaviour.

– Well-known experimental and theoretical examples are found in animal societies which are conceptually close to artificial systems.

6.4 When Do Animal Population Use Emergent Behaviour or Self-organisation?

Emergent behaviours appear most useful in real populations of individuals that have to cooperate. Although this point seems obvious, some applications of the so called "ant algorithms" do not really fall into this category. Indeed, the population is just momentarily and artificially created to solve the problem like when solving an optimisation problem with an "ant algorithm" (Dorigo and Stützle, 2004). The problem is solved *a priori* and then the solution is implemented in a centralised manner. Even if this approach is interesting, we think that it is somehow diverging from the core logic of emergent behaviour. To illustrate this point, below, we list some of the characteristics of the populations presenting autonomous organisation.

– Actions and decisions are *simultaneous and mixed*, actions and decisions are concomitant.

– Only *limited "cognitive" capabilities* of agents are needed to collect and process information.

– Tasks or resources allocation between agents is flexible and autonomous.

– Agents are *unpredictable* because they need to be stochastic in some behaviour.

– There is *no need for a perfect global knowledge* of the system by the agents.

– It is an alternative to predict all the needs of an agent population at anytime.

Emergent behaviour is very useful when the decision has to be taken while action takes place. In a natural context, in animal populations, these types of behavioural pattern are only used in real time. It means for example, that there is no possibility to stop the system, to make an optimisation computation and to start it again. There is a progressive build-up of the solution and these systems keep a flexibility to respond to environmental or social changes. When ants are looking for the best source of food with the shortest path they find the optimal solution by working out, by *walking the computation* (in the literal meaning of the word "walk"!) and not by stopping solving an optimisation problem and then implementing it. This implies that it takes some time before an optimal solution is found, during this time the colony explore and make use of all available possibilities and can be under-performing.

Moreover, in many situations, populations are influenced by the environment that becomes a kind of particular agent in the system (Detrain et al., 1999; Detrain and Deneubourg, 2002). Nevertheless, the properties of the environment do not need to be encoded explicitly in the individuals neither do the agents need a global view. In this section, we discuss some mechanisms that produce adaptability to perform certain task in an autonomous way by a collection of individuals with fixed rules on the time scale of the task to be performed. There is still the important question how and why such mechanisms have been selected by evolution and how they could be further evolved or stabilized. Somehow, we are dealing with mechanisms that produce adaptability to perform a task at "short" time scale and for an evolutionary approach (see the contribution by Floreano and Keller and by Nolfi) are dealing with adaptability at "longer" time scale by evolving those mechanisms. We define "short time scale" by the time it takes to the system to actually perform the task and "long" by the time it takes for the system to perform a large number of time the task.

6.5 The Study of Generic Rules Governing Self-organized Systems

6.5.1 Introduction

One important research challenge is that to identify the type of IT problems that can be solved at the collective level by relying on different type of communication systems. The emphasis of this sub-project is on the basic characteristics of the communications that are important in an IT context. Using the methods of complex systems science and computer science, we aim at studying what kind of performances at the collective level different communication systems make possible. This study must take account of the parameters characterizing the situation (e.g. the number of agents), the characteristics of the tasks and also the information processing at the individual level.

A large number of collective decisions in social species (mainly social insects and gregarious arthropods, see e.g. Pasteels et al., 1987; T. et al., 1991; Beckers et al., 1992; Camazine et al., 1999; Millor et al., 1999; Visscher and Camazine, 1999; Saffre et al., 1999; Portha et al., 2002; Sumpter and Pratt, 2003; Jeanson et al., 2004) are based on the competition (in space and/or time) of behavioural positive feed-back corresponding to situations where :

i) the individual probability of adopting a behaviour increases with the number of individuals exhibiting this behaviour;

ii) the individual probability of leaving a behaviour decreases with the number of individuals exhibiting this behaviour.

In these systems, the nature of the positive feedbacks, such as recruitment, often involve specific behaviours by individuals; in contrast the negative feedbacks often arise "automatically " as a result of the limits or constraints in the system (for example, the depletion of building materials or the exhaustion of the "resources", Dussutour et al., 2004). Moreover, the number of parameters needed to characterize the communications is lower that the number characterizing the individual capabilities to process information. The repertoire is based on a set of n different signals (e.g. n different pheromones in the case of the chemical signals). The intensity of the signal perceived by the individual controls its response. This intensity depends on the emission and physical parameters affecting its propagation and lifetime. Often, one signal corresponds to one behaviour. However, in many cases, different intensities of the same signal may induce different behaviour such as attraction or fleeing.

6.5.2 Taxonomy of The Structures and Characteristics of The Communications

Most often, the classification of the social activities is based on the classical biological functions: feeding, reproduction, anti-predator behaviour,... However this classification occults the existence of generic rules that are involved in many tasks. A new classification should be proposed to explore the link between the problem and the characteristics of the interactions needed to solve it. Among the tasks that must be performed, having their counterpart in IT, clustering and sorting, the coupling between exploration and exploitation and the synchronization are the most important.

> Our first challenge is to start building this taxonomy and to find the characteristics of the communications (range, mean life time, network vs. broadcasting, level of noise) leading to an efficient problem-solving.

6.5.3 Number of Agents and Randomness

The emergence of a pattern largely depends on many parameters and the relationships between these parameters such as the coupled number of agents and intensity of the communication that affects the level of noise. This level may also depend of individual characteristics.

It is well known that if the intensity of communication is high enough, small groups are able to self-organize. However, it does not mean that these patterns are efficient. The study of optimality in such systems needs to take account of the stochasticity of the phenomenon. An approach accounting for the fluctuations in the number of foragers as well as in the process of decision has been carried out in the case where an ant colony has to choose to follow trails leading to food sources of different quality (Nicolis et al., 2003).

The study shows that the efficiency of the collective decision depends on the intensity of communication. For a given colony size, it exists an optimal intensity of the signal, and hence an optimal amplification and an optimal level of noise. This optimal amplification is related to the colony size: for a large colony, many individuals lay small quantities of pheromone, instead of a small group of individuals laying higher trail amount. At least, the optimal response of a large colony is always greater that the optimal response of a small colony.

> The challenging question is that the results obtained in the specific biological context of trail recruitment can be generalized to other decision processes involving different competing options.

For instance, aggregation can be described by similar mathematical models when individuals of a colony have the choice between different relative attractive sites to aggregate themselves (Rivault and Cloarec, 1998; Lioni and Deneubourg, 2004; Ame et al., 2004; Jeanson et al., 2004, in press). It can therefore be expected that since the mechanisms, underlying this phenomenon (and more generally all phenomena implying competition between positive feed-back) are similar to recruitment, there exists an optimized value of amplification and interaction between agents and the same relation exists between the optimal amplification and the size of the group.

6.5.4 How Many Patterns Only One Signal is Able to Produce?

As we previously mentioned in many cases the individual information processing is characterized by many more parameters that the communication. The agents are characterized by different variables: active vs inactive, mobile vs immobile, the number of behavioural states and/or internal states. Their individual response (transition between different behavioural states, the emission of the signal, etc.) is a function of perception of the signal, the different stimuli from the environment and their internal states. In this context an important question is how many patterns may be produced or how many tasks may be performed with only one signal. This study is in part motivated by the observation that (i) in the same species the same pheromone or blend of pheromones is involved in many tasks and (ii) that the same type of signal (e.g. trail pheromone) leads to very different patterns in different species.

> The challenge is to identify the rule of thumb governing the individual behaviour and the characteristics of the communications to produce the different structures observed in bio-systems such as collective choice, external memory and multistationnarity, synchronization and the different spatial patterns (e.g. Turing-like patterns, excitable waves).
>
> Correlatively, it is essential to test the following hypothesis: for a given task and environment, to keep an efficient collective response, any simplification of the behavioural algorithms must be compensated by an increase in the number of different signals exchanged between agents.

6.5.5 Intelligent Decision Criteria

In many tasks, to produce an efficient response, the agents must integrate many parameters to modulate their different behaviours. One way for insect societies to cope with the complexity of their environment is the use of intelligent decision criteria at the individual level (Detrain and Deneubourg, 1997; Mailleux et al., 2000, 2003; Detrain and Deneubourg, 2002). This concept seems particularly important not only in a biological situation but also in IT situation. Intelligent decision criteria do not require the ant to make some complex and precise assessment of all environmental parameters; instead, they rely on cues that automatically integrate several variables (inside or outside the nest). Since potential cues vary in their value as indicators, one might expect that, through evolution, only very good cues—those with a high, reliable, and functional informative content—have been retained as decision criteria. In other words, the "intelligence" of a decision criterion results not simply from the use of cues that intrinsically catch a part of the environmental complexity, but also from the selection of the best cue—that is, the one most pertinent for the activity of the ants.

> The main challenge is to identify the context and the characteristics in a task, where a large number of parameters must be integrated by the individuals that lead to the use of efficient intelligent criteria.

6.5.6 Alternative Scripts

Animal societies offer a complete blend of individual capacities and collective levels of intelligence and complexity. The self-organized systems, that we review, are examples of rules of thumb needing a limited cognitive ability and a limited access to global information to produce social complexity. However, this type of behavioural rules is not specific to self-organization and is shared with alternative scripts, such as hierarchical organisation (Hemelrijk, 2002) that are often assumed to involve high individual cognitive capacities and individual recognition. However, from the point of view of interactions between individuals, hierarchical scripts may be seen as a self-organization script where the individuals can only differ by the frequency of emission and/or the response threshold to different environmental stimuli and the intensity of the communication. These differences are enough to produce asymmetrical interactions between the individual, this asymmetry being one characteristic of the hierarchy. In animal societies, hierarchy is too often discussed in term of reproductive success and rarely analyzed in term of problem-solving or decision-making.

> Currently, it seems essential to find the conditions where a hierarchy is more (or less) efficient that a classical self-organized system where all the individuals may be identical and the interactions between individuals are symmetrical.

6.6 Building up Artificial Ecagents

6.6.1 Introduction

The research strategy adopted for this second sub-project is the counterpart of the main objective of Ecagents.

> A second complementary important challenge is that to build in laboratory, artificial groups of natural agents able to cooperate and to perform tasks that their "ancestral wild types" in natural conditions are unable to do.

From an IT point of view, these new societies can be autonomous teams of sensors able to process information (the animal being equipped with sensors or being itself the sensor), or factories -fortresses able to solve problems or produce chemical substances, manage wastes, etc.

These new societies can be generated by three different ways: genetics manipulations, environmental manipulations and social manipulations affecting the social ontogenesis or the size of the group. Here we will discuss only how environmental and social manipulation will be theoretically studied. Among all the problems to be solved to synthesize such groups, we will focus on cooperation based on new communication and the emergence of division of labour.

Among many wrong hypothesises concerning group living organisms, we are mainly concerned with two. The first one, based on a lack of observation, assumes that the behavioural algorithms of solitary and social animals performing the same task are always deeply different. The second one, previously mentioned, assumes that self-organization only works when a large number of individuals interact. In fact it is the number of events that is essential and not the number of individuals. Having in mind, that : (i) there is no fundamental difference between the behaviour of solitary and social organisms; (ii) that self-organization may be involved in solitary activity and (iii) it exist a strong behavioural plasticity, we assume that a diversity of collective structure may be generated without genetic modification.

6.6.2 Emergence of Cooperation and Division of Labour

Different studies show that some mechanisms used by group living organisms also govern the behaviour of solitary organisms. Solitary species use amplification mechanisms based on the chemical marking of their resting site or trail orientation. Stigmergy, a stimulus–response mechanism involved in building behaviour is not only used by social species but also solitary individuals (see e.g. for social or solitary spiders: Saffre et al., 1997; Gunderman et al., 1993). The consequences of such a generic logic could then be one of the keys in understanding the transition between different forms of cooperativity, and therefore different degrees in sociality. In other words, the transition between solitary to social activities do not always need the selection of new behavioural algorithms.

Chemical communications provide a good material to study the synthesis of emerging cooperation such as the building of a common network or the selection of a site of clustering.

Every individual may use its own trail or chemical blend to build its foraging network or mark its shelter. The chemical marking acts as an external memory. Preliminary theoretical results show that a slight inter-attraction between the markings of different individuals may induce the formation of a cluster. This leads to the question of how a solitary species might be manipulated to lead to clustering. More precisely what must be the characteristics of the marking and the environment to lead to different patterns of aggregation or the use of a common network of trails?

In the context of self-organization and transition between different social organizations, aggregation and its resulting increase in density is a prerequisite for the emergence of higher forms of cooperation such as social specialization. This simple dependency of density could lead or be involved in the process of the social differentiation. The interplay between amplification mechanisms (e.g., growth or learning) and the competition in a cluster could be enough to produce the social differentiation that has been described for very different species (e.g., social spiders, Rypstra, 1993, sea-urchins, Grosjean et al., 1996, queens of ants, Fewell and Page, 1999, for a model see Bonabeau et al., 1998).

Similarly, a specialization of the members of the cluster would be generated. In social insects, some divisions of labour are the result of self-organized mechanisms (see e.g. Beshers and Fewell, 2001) leading to a strong correlation between the colony size and the level of individual specialization: the bigger the colony, the higher the specialization. To summarize, task allocation and individual specialization will be shaped by the dynamics of aggregation that itself can be induced manipulating the environmental characteristics.

Part II

Human-Like Communication

The Characteristics of Human Language

Domenico Parisi

Institute of Cognitive Sciences and Technologies

7.1 Introduction

ECAgents (Embodied Communicating Agents) are artificial entities that have a body, interact with a physical environment, and communicate with each other and, possibly, with human beings. ECAgents can either be simulated in a computer or they can be actual physical artefacts. Constructing and analyzing ECAgents has two goals. One goal is purely scientific. ECAgents can be created in order to better understand the communications systems of natural organisms. If one succeeds in constructing artificial agents that demonstrably communicate in the same, or similar, way as some particular real animal, it is reasonable to think that one has reached some understanding of the way in which that particular animal communicates. The second goal is technological and applied. ECAgents can generate suggestions on how to construct ECDevices (Embodied Communicating Devices), i.e., physical artefacts that communicate with each other and, possibly, with their users, and that have some practically useful function. The two goals are mutually beneficial. Not only ECAgents with purely scientific goals can suggest new types of ECDevices but designing ECDevices that have practical applications may suggest new interesting scientific questions and possible answers to these questions.

The communication system of ECAgents can resemble animal communication systems or it can also resemble human language. Human language has some properties and functions in common with animal communication systems but it also has many properties and functions that make it different from animal communication systems. As always when one compares human beings with other animals, there is no neat dividing line between humans and other animals and one can find simpler manifestations of typical human traits in this or that non-human animal. Furthermore, human language first arose in primates that only possessed animal communication and an important research question is how the transition occurred. But human language clearly has a number of properties and functions that distinguish it from animal communication and, even if this or that typical feature of human language can be found, at least in some embryonic form, in animal systems, the simultaneous presence of all the features appears to be unique to human language (Hauser, 1996). Therefore, one can construct ECAgents that possess animal-like communication systems or one can concentrate on those properties and functions that typically characterize human language. ECDevices, of course, can possess mixed communication systems, with properties and function of both animal communication systems and human language, if this turns out to be useful for practical applications.

Which properties and functions characterize human language and distinguish human language from animal communication systems? The following is a possible list (cf. also Hockett, 1960a). Human language:

1. has syntax and, more generally, has signals which are made up of smaller signals
2. is culturally transmitted and culturally evolved
3. is used to communicate with oneself and not only with others
4. is particularly sophisticated for communicating information about the external environment
5. uses displaced signals
6. is intentional
7. is the product of a complex nervous system
8. influences human cognition.

Let us briefly comment on these eight characteristics of human language.

7.1.1 Human language has syntax and, more generally, uses signals which are made up of smaller signals

Animal signals tend to be simple, i.e., they are not made of smaller signals that have meaning. Although some signals of birds and nonhuman primates may be analysed as combinations of recurring parts, the parts do not appear to have separate meaning. Linguistic signals are complex. They are made up of smaller signals that have their own separate meaning, and it is the particular way in which the smaller signals are combined in a larger signal that determines the meaning of the overall signal. This combinatorial or compositional character of human signals manifests itself at a hierarchy of levels: phonemes (that do not have separate meaning) are composed into morphemes, morphemes into words, words into phrases, phrases into sentences, sentences into discourses and dialogues.

7.1.2 Human language is culturally transmitted and culturally evolved

Animal communication systems are genetically transmitted and they are the result of a long process of biological evolution. Human language is culturally transmitted, i.e., learned from others. Human infants acquire language by interacting with other people who already possess the language, although language learning clearly is based on species-specific genetic predispositions. Historical languages, such as English or Italian, arise through a process of selective cultural transmission of linguistic signals and the constant addition of new variants.

7.1.3 Human language is used to communicate with oneself and not only with others

Animals only use their communication systems socially, i.e., they exchange signals to communicate information from one individual to another individual. Humans use language both to communicate with other individuals and to communicate with themselves. This intra-individual use of language underlies much of what is called mental life, i.e., thinking, remembering, reasoning, predicting, planning, deciding, etc.

7.1.4 Human language is particularly sophisticated for communicating information about the external environment

Most animal signals communicate information about the sender (I am here; I am angry) and not about the external environment, whereas human language may communicate information about both the sender (I am angry) and the external environment (the book is on the table), and is particularly sophisticated for communicating information about the external environment.

7.1.5 Human language uses displaced signals

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 about past and future states of the sender (I was angry, I will be angry) or of the environment (the book was on the table, the book will be on the table).

7.1.6 Human language is intentional

Most animal communication is unintentional or expressive. Animal signals are emitted without thinking or deciding to emit them. They tend to be direct mappings from the current state of the sender or of the external environment to the production of the signal. Human language is intentional. The mapping from meaning to signal in the nervous system of the sender is more indirect and complex. This may be related to the fact that, unlike animal signals, linguistic signals even when they are addressed to another individual tend to also be signals for the sender (see (3) above).

7.1.7 Human language is the product of a complex nervous system

Human beings have a more complex nervous system than other animals and this property of their nervous system can be both a necessary pre-condition and a consequence of a more complex communication system such as human language. Animal communication systems can be produced by simple nervous systems that directly map input into output (Figure 1). Human language requires a more complex nervous system with distinct modules, separate pathways, recurrent circuits.

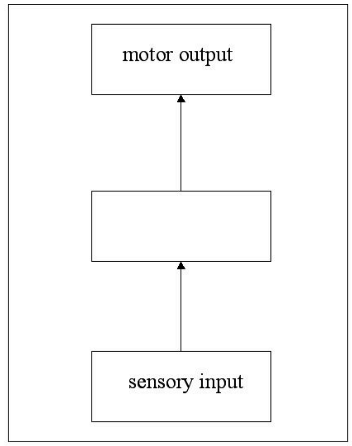

Fig. 1. Purely reactive network

7.1.8 Human language influences human cognition

The communication systems of nonhuman animals appear to be juxtaposed to their cognitive abilities and not to have any particular influence on these abilities. In contrast, human language seems to have an influence on the cognitive abilities of its users, for example on how human beings categorize the world, how they analyze the world, how they remember past experiences and prepare to future ones.

The goal of the present chapter is to explore how the ECAgents approach can be applied to the study of these eight characteristics of human language. As we have already said, we do not claim that animal communication systems entirely lack these characteristics. Human language has emerged in populations of organisms that originally lacked human language but possessed animal-like communication. Therefore, there can be no clearcut separation and difference between animal communication and human language. Furthermore, the aim of the ECAgents approach is not simply to conceptually identify and describe these eight characteristics but to actually construct artificial agents that display the eight characteristics. Since constructing ECAgents is not to design them but to let them evolve or learn whatever communication abilities they will eventually exhibit, one might start with simpler systems that lack human language and see how communication systems that have some of the properties of human language can gradually emerge.

7.2 ECAgents approaches to the eight characteristics of human language

In this section we sketch some research directions in an ECAgents approach to human language and we briefly mention some work which has already been done.

7.2.1 Human language has syntax and, more generally, has signals that are made up of smaller signals

Suppose you want to be able to communicate the following 8 facts:

1. The book is on the table
2. The pen on the table
3. The book is under the table
4. The pen is under the table
5. The book is on the chair
6. The pen is on the chair
7. The book is under the chair
8. The pen is under the chair

If your communication system is made of simple signals, you would need 8 different signals, one for each of the 8 different facts. If on the other hand you have human language 6 signals would be sufficient (book, pen, on, under, table, chair) and you would be able to communicate the 8 facts by combining together three simple signals (words) to form one complex signal (sentence). The compositional way to communication is very powerful. By adding one single simple signal, e.g., glass, you would be able to generate four more complex signals (the glass is on the table, the glass is under the table, the glass is on the chair, the glass is under the chair). With longer and longer complex signals, adding a limited number of further simple signals would allow you to generate an increasing and very large number of complex signals.

One critical challenge for ECAgents research is to be able to costruct ECAgents that start with an animal-like communication system with only simple, noncompositional signals and gradually develop a human-like communication system with complex, compositional signals. Once a communication system with complex signals exists, it has to be learned by new members of the community (children). Language is learned by noticing the systematic co-variation of specific signals with specific aspects of ones experience and incorporating these co-variations in ones nervous system (Chapter 10). If a learning agent is exposed to complex signals the agent has to notice the co-variation of specific sub-parts of a complex signal with specific sub-components of its current experience and to incorporate these partial co-variations in its nervous system, not the co-variation of the entire complex signal with the entire experience.

Human language is compositional all the way, from phonemes to morphemes to words, phrases, and sentences. But the critical aspect of human languages compositionality that we should be able to incorporate in ECAgents is syntax, which is the combining of words into phrases and phrases into sentences. In a sentence the meanings of the words are combined together to generate the meaning of the sentence. Since there may be many different ways to combine together the meaning of a set of words into the meaning of a sentence, a sentence provides a number of cues for combining together the meanings of the words in the way which is intended by the speaker. These cues are called grammar. Grammatical cues can consist in the order in which the words follow each other

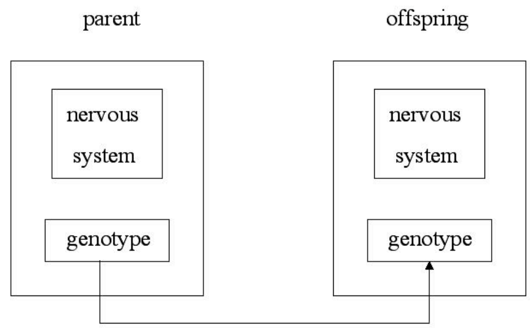

Fig. 2. Genetic transmission

in the sentence or in special signals which may be attached to words (bound morphemes) or free-standing (function words). (We return to compositionality in (6) below.)

7.2.2 Human language is culturally transmitted and culturally evolved

All communication systems are transmitted from one individual to another individual, either genetically (Figure 2) or culturally (Figure 3). Communication systems that are genetically inherited and evolve biologically, i.e., animal communication systems, are transmitted from parents to offspring. Communication systems which like human language are culturally inherited (learned from others) may be transmitted in a variety of manners. They are mostly inherited from parents (vertical cultural transmission; cf. Cavalli-Sforza and Feldman, 1981) and from other members of the preceding generation (oblique cultural transmission) but they are also learned from other members of the learners own generation (horizontal cultural transmission) and, in fact, from any individul with which the learner enters into contact and interacts. While biological transmission is entirely one-way, i.e., from parents to offspring, the cultural transmission of language is mostly one-way when the child first acquires the language but it becomes two-way in all the encounters that one individual has with other individuals during its life. Biologically inherited communication systems (animal communication systems) are transmitted at one single time, i.e., at birth, although the genotype can translate into the phenotype as a result of a temporal process of development. In contrast, human language is learned and modified during the entire course of an individuals life and is the result of the particular social network of interactions which exists in a collectivity of individuals. If a collection of individuals with a shared language divides up into two separate collections with no interacting links between the two new collections, different languages will tend to emerge in the two collections (Chapter 12).

Biologically and culturally transmitted systems also differ in the manner in which they change. Biological evolutionary change is mostly the result of random

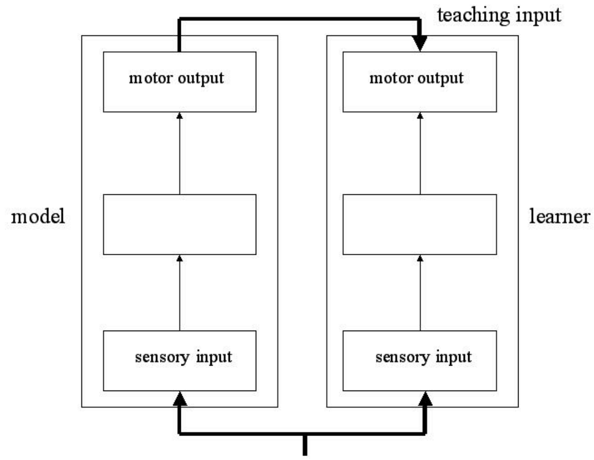

Fig. 3. Cultural transmission

mutations and the creation of new genotypes that are new combinations of parts of the genotype of one individual (mother) with parts of the genotype of another individual (father) - for sexually reproducing organisms. Cultural evolutionary change is the result of a more complicated set of factors. When an individual learns to speak in the same way as the individuals with which the individual interacts, there may occur random transmission errors that resemble random genetic mutations, but also the creation of new ways of speaking that combine aspects of the language of a large number of other individuals (not just two, mother and father, as in sexual recombination), in fact, as we have noted, of all the individuals with which the individual interacts.

Furthermore, the mechanism of selective reproduction is simpler in biological reproduction. Only the genotypes of the individuals that have offspring are found in the next generation. In selective cultural (linguistic) reproduction there may be a number of different factors that determine which individuals transmit their language to other individuals. Boyd and Richerson (1985) mention direct bias (the cultural/linguistic trait is seen as leading to success), indirect bias (the cultural/linguistic trait does not lead per se to success but is associated with traits that lead to success), prestige bias (the trait confers prestige to its possessor), frequency bias (the traits exhibited by a larger number of individuals tend to be copied more that those exhibited by a smaller number of individuals), and its contrary, minority bias (preferred traits are those exhibited by a minority of individuals).

Finally, unlike animal communication systems, human language, as all culturally transmitted traits, may change because of the intentional creation of new words and new ways of speaking.

If a communication system is genetically transmitted and evolved it is inevitable to ask whether it benefits the sender or the receiver of signals or both.

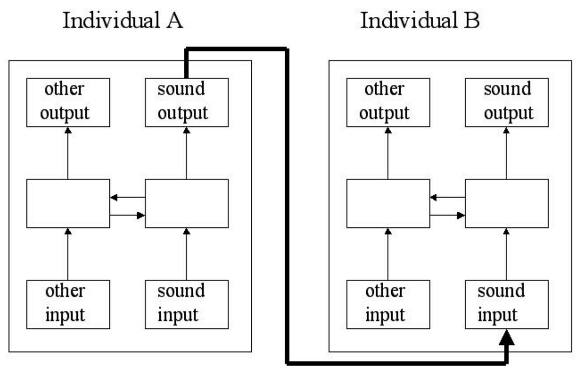

Fig. 4. Language for communication with others

Since human language is culturally rather than genetically transmitted these problems arise in a different manner in the case of language. Language appears to be learned because of a general, genetically inherited tendency to learn from others (Herbert Simons docility), which develops biologically because it is mostly, even if not always, beneficial for the individual. However, problems of who benefits from language arise not in the context of language transmission but in the context of language use. What are the advantages for the sender of saying one particular thing? What are the advantages for the receiver of behaving appropriately in response to what has been said?

7.2.3 Human language is used to communicate with oneself and not only with others

Animals use signals to communicate with other animals, mostly conspecifics but also members of other species. Human language is used to communicate with other individuals (conspecifics) (Figure 4) but it is also used to communicate with oneself. One individual generates a signal but this signal is not produced to communicate information to another individual but is produced to communicate information to oneself (private speech; Figure 5). In many cases the signal, typically, a sound, is not even externally emitted but it is only internally generated, so that other individuals cannot perceive the signal (inner speech; Figure parisi6).

An important objective of research on ECAgents is to construct artificial organisms that produce signals for themselves (Chapter 10). Humans produce signals both for other individuals and for themselves, and they appear to use mostly the same signals both for others and for themselves. (But according to Vygotsky, inner speech is somewhat different from external speech. Cf. Vygotsky, 1986) However, it is an open question whether it might exist real organisms - or whether it would be possible to construct ECAgents - that have a com-

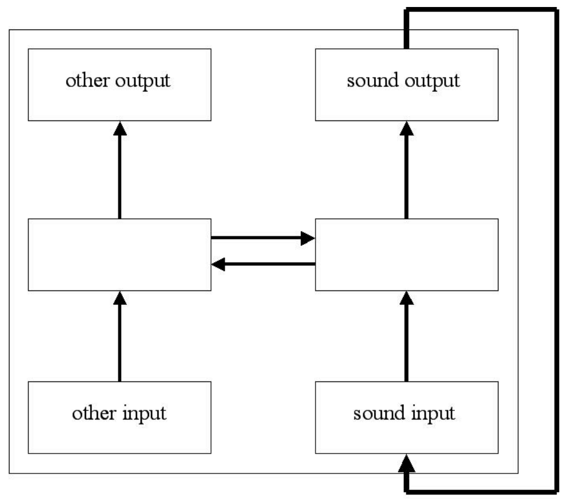

Fig. 5. Private speech

munication system which is only used to communicate with oneself, and not for communicating with other individuals (cf. Wittgensteins private language argument).

But, of course, the most interesting research question is what the function(s) of producing signals for oneself might be. A simple function might be a memory function. Information which arrives to the input units of an agents neural network might be better stored for future use if the information is mapped into linguistic signals and it is the linguistic signals that are retained in the neural networks memory rather than the raw information itself. Linguistic signals may occupy less space in memory than the raw information to which they refer or, if the preservation of information in memory requires recycling of the information, the recycling can be easier and more efficient if it is signals that are recycled, that is, repeated to oneself, instead of the raw information itself.

Storing information in the form of linguistic signals may take place in two different conditions. In one condition, one individual perceives some raw information as input and it produces a signal that describes the information as output. The signal is received by another individual, which stores the linguistic signal and, when it has to use the information, it maps back the signal into the information. In another situation, the individual is all alone, it perceives some information in the environment which it would be useful to keep in its memory, and the individual produces a signal and stores in its memory the signal rather than the information itself (Mirolli and Parisi, submitted).

Other adaptive uses of producing signals for oneself could be to linguistically articulate ones experience if this leads to better performance, to linguistically label ones predictions on the future if this makes it possible to generate chains of linguistically labeled predictions extending further into the future, to allow rea-

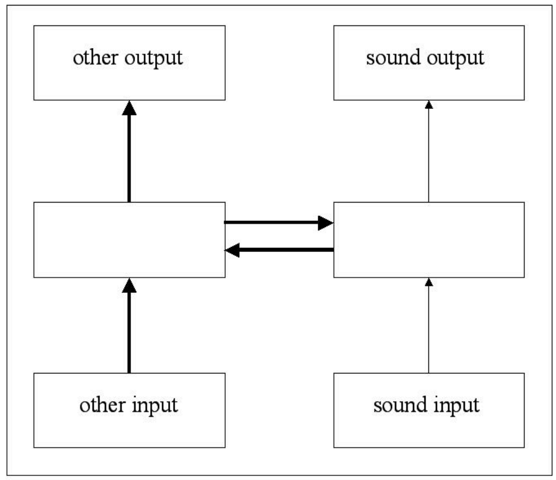

Fig. 6. Inner speech

soning as the deduction of linguistically described facts from other linguistically described facts.

7.2.4 Human language is particularly sophisticated for communicating information about the external environment

Animal signals mostly communicate information about the sender of the signal, its current location, its species, sexual, or individual identity, its current emotional state, its intentions and attitudes, etc (Hauser, 1996). There are exceptions such pheronome signals, food calls, alarm calls, but these signals communicate very restricted information about the external environment. In contrast, human language is very sophisticated for communicating information about the external environment and, more specifically, spatial information: where things are, how they can be reached with the hands or legs, what their spatial relations are, etc. One can even advance the hypothesis that the advantages of possessing a communication system so useful for communicating information about the external environment have been an important pressure for its biological/cultural emergence. In any case, language has a rich repertoire of signals for identifying objects and landmarks in the environment and for describing spatial relationships between objects and landmarks. These signals appear to be critical for ECAgents that tend to displace themselves, or be displaced by humans, in the environment, and that have to communicate to each other where things are in the environment and how they can be reached.

7.2.5 Human language uses displaced signals

Imagine an agent that discovers where some entity, say, a prey, is located in the environment and it wants to communicate this information to other individuals so that the other individuals can also find the prey. One way of doing this is to remain near to the prey and to emit a signal, say a loud sound, which can be received by the other individuals. The other individuals respond to the received signal by approaching the source of the signal, that is, the sender, and, therefore, the prey itself. This solution, however, has many limitations. One limitation is that the sender has to produce the signal while remaining near to the prey. The signal is useless if the sender moves away and then it produces the signal. Another limitation is that the other individuals must be close enough so that they can receive the signal, i.e., hear the sound. A third limitation is that hearing the signal may induce the prey to fly away.

A different solution is to produce a signal which co-varies with the location in which the prey has been discovered, where the location of the prey is identified with respect to some landmark. Imagine that the prey can be found either near to a river or near to a hill. The discoverer of the prey produces one signal when it finds the prey near to the river and a different signal when it finds the prey near to the hill. The other individuals respond to the first signal by going to the river and they respond to the second signal by going to the hill. This system of communicating information about the location of the prey has none of the limitations of the previous system. The sender of the signal can produce the signal whatever its current location in space. It can produce the signal in any place and at any time. The receivers of the signal must be near to the sender of the signal when the signal is emitted in order to be able to hear the signal but this may happen separately for each individual receiver of the signal. Furthermore, one receiver of the signal can communicate the signal to another individual, and so on in a chain, with no need that all the individuals be together at any given time and place. Finally, since the discoverer of the prey can produce the signal after it has moved away from the prey, the signal can be produced with no risk that the prey hears the signal and flies away (Caretti, Baldassarre, and Parisi, in preparation).

Signals whose meaning or function is independent of the current location of the sender of the signal and of the time in which they are produced, can be called displaced signals (Hockett, 1960b). Emitting a loud sound when one discovers the prey is to produce a non-displaced signal. Emitting a signal that co-varies with the location in which the prey has been discovered, is to produce a displaced signal. Animals signals tend to be non-displaced. Linguistic signals, except so called deictic signals (e.g., this, that, I, you, here, there, etc.), are displaced signals.

One interesting contrast between displaced vs deictic signals concerns pointing. Pointing, with the gaze or with a finger, is one way of communicating where things are. Notwithstanding its limitations as a deictic signal, pointing has advantages in comparison with the use of explicit linguistic signals, and one interesting research direction is to create ECAgents that can use pointing gestures (Chapter 10). Although it is deictic and can only be used for communicating the location of objects which are present in the space currently directly accessible to the senses of both the sender and the receivers of the pointing signal, pointing is not generally found in animals. This seems to indicate that pointing is a complex cognitive/communicative ability and this complexity extends to deictic linguistic signals such as this, that, here, there, to the left of, etc.

7.2.6 Human language is intentional

While the production of signals in animals is spontaneous, inevitable, and mechanic, linguistic communication tends to be intentional and subject to the judgment of the speaker. (Partial and limited exceptions can be strategic communicative episodes in nonhuman primates in which the sender may try to deceive the receiver and it appears to calculate the effects of its signals.)

The characterization and implementation in artificial agents of this sixth difference between animal communication and human linguistic communication poses special problems because the distinction between intentional and unintentional communication is difficult to capture in operational terms. Here are some possible ways of proceeding.

Intentional communication, and intentional behavior more generally, appears to be linked to the tendency/ability to predict the consequences of ones own actions. A purely reactive agent is an agent that receives some input from the external environment or from inside its own body and responds by producing some movement that changes either the physical relation of the agents body to the external environment (e.g., the agent displaces itself in the environment) or the external environment itself (the agent manipulates the environment). The neural network that underlies the behavior of a purely reactive agent may have a purely a feed-forward architecture: from sensory input to motor output. But consider a network architecture which includes a set of units encoding a prediction of the next sensory input. Given the current input which is encoded as some specific pattern of activation in the networks sensory units, the network is able to generate a pattern of activation in one subset of its internal units (prediction units) that matches the pattern of activation that will be observed in the sensory units at some future time. This pattern of activation is a prediction.

There are two types of predictions. An agent can generate a prediction of the next sensory input when the next sensory input is independent of the agents own behavior, e.g., predicting the weather. Or the agent can generate a prediction of the next sensory input when this input depends both on the current input and on the physical action with which the agent responds to the current input: moving ones eyes, arms or legs. These are the predictions that interest us here (Figure 7). To be able to predict the consequences of its own actions, the agent must be able to encode its motor response to the current sensory input as a pattern of activation in the motor output units but must generate a prediction of what its sensory consequences will be before the motor response is physically executed. If the prediction is generated after the motor response has been physically executed, there would be no need to predict the next sensory input because the executed motor response will actually produce its consequences and the next sensory input will be actually observed.

It has been shown that agents whose behavior is controlled by a neural network can learn to predict the consequences of their own behavior (Parisi et al., 1990; Nolfi et al., 1994). For example, an agent that displaces itself in the environment and tries to approach randomly distributed food elements can learn to predict how the position of food relative to itself changes as a function of its displacement movements. Or an agent that moves its arm to reach for an object, can predict how the proprioceptive input from its arm changes as a function of the movement of the arm. In these cases, the agent generates a prediction of the next sensory input (new position of food relative to the agent; new position of the arm) on the basis of the current input from food or arm and the planned movement with which the agent will respond to the current input. After the prediction has been generated, the planned movement is physically executed, the sensory input changes, and one can show that the predicted sensory input matches the actually observed input.

Why might be adaptive to develop an ability to predict the next sensory input that will result from ones own movements? In the simulations we have mentioned the predictions are not used by the neural network itself. This notwithstanding, it can be shown that learning to predict the consequences of ones own actions may cause changes in the neural networks connections weights which lead to an improved performance of the sensory-motor task. Agents that learn to predict the consequences of the displacements of their body or of their arm movements are better able to reach food with the entire body or objects with their arms than agents that do not learn to make these predictions.

However, the predictions about the consequences of ones own actions that are generated by a neural network can be used by the neural network itself if the internal units encoding the predictions are not dead ends but they feed back connections to the rest of the network (Figure 7). Using the predicted information the network can decide whether to physically execute a planned action or block the execution of the action, or it can decide which action to execute, action A or action B. For example, a network that can predict whether a stone of a given weight which is launched with a given force will actually reach a prey situated at a certain distance, can actually launch the stone with that force if the prediction is positive and refrain from doing so if the prediction is negative (Tria and Parisi, in preparation).

Let us return to communicative signals. Imagine an agent that is ready to respond to some input with the production of a signal. Consider two possibilities. In one case the agent is a purely reactive agent. The agent receives an input from the external environment (or from inside its own body) and it responds by producing a signal which is received by another agent. In the other case, the agent is not purely reactive. The agent responds to the input by formulating a signal in its signal producing output units but is able to delay the physical production of the signal until the agents neural network has generated a prediction concerning the consequences that the signal will cause in the receiver of the signal. If this prediction feeds back into the agents neural network because the prediction units send connections to the rest of the neural network, the agent can decide

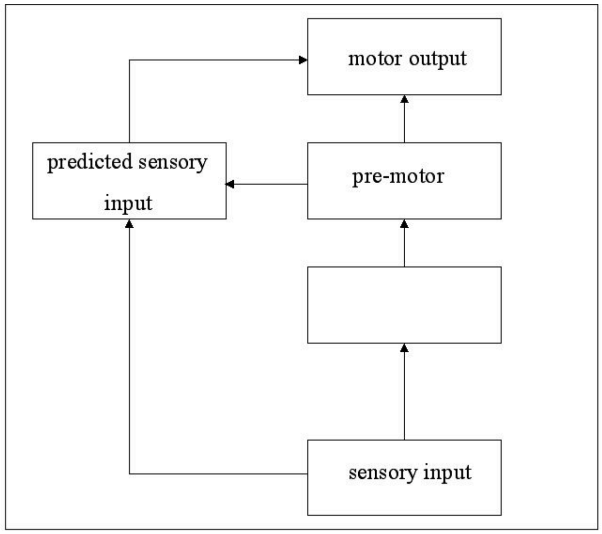

Fig. 7. Predicting neural network

whether to physically producing the planned signal or to refrain from doing so. In these circumstances we might begin to say that the sender has an intentional communicative behavior.

7.2.7 Human language is the product of a complex nervous system

Human beings have a more complex nervous system and a more complex communication system than other animals, especially insects. The two things clearly are related. Human language has only been possible given the complex nervous system possessed by humans and, at the same time, it is possibile that the development of a complex communication system such as language has been one of the evolutionary pressures for the emergence of a complex nervous system.

While constructing ECAgents with a simple, insect-like, communication systems may not require that any special attention be devoted to the architecture of the neural network controlling the ECAgents behavior, ECAgents with a human-like communication system should be endowed with a more complex and explicitly designed neural architecture.

Children from birth to 1 year do not have language. During their first year they develop from a sensory-motor point of view, acquiring various perceptual and manipulatory abilities such as looking at things and reaching and manipulating objects, and at the same time they acquire various acoustic/phono-articulatory abilities, such as repeating ones own sounds, babbling and, at least from 6 months to 1 year, incorporating in the sounds they produce some of the specific properties of the sounds of the particular language spoken in their environment. For ECAgents this implies that the connection weights linking input

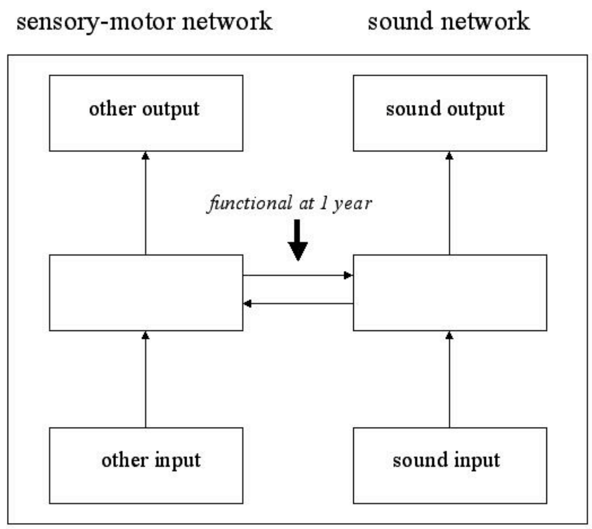

Fig. 8. Neural network underlying human language

to output inside the agents neural network are gradually modified so that an input causes the appropriate output. However, the two developments appear to be separated. It is as if the nervous system of the child before 1 year of age consisted of two separated sub-networks, a sensory-motor network with visual and tactile input and movement output (eyes, face, arms, hands, legs) and a sound network with sound input and sound (phono-articulatory) output.

At around 1 year the two sub-networks become functionally (and perhaps even anatomically) linked (Figure 8). The child begins to acquire language. The weights of the connections linking units in one network with units in the other network progressively change their value so that an input in one network causes an appropriate output in the other network, and vice versa. Language comprehension is to be able to generate the appropriate output in the sensory-motor network given some particular input in the sound network (Figure 9). Language production is to be able to generate the appropriate output in the sound network given some particular input in the sensory-motor network (Figure 10).

A neural network with an architecture made up of two initially separated and then interconnected sub-networks is more complex than a simple input/output reactive, insect-like, network. Consider that, to make both language comprehension and language production possible, there must be connections linking the two sub-networks both ways. There are connections from the sound network to the sensory-motor network (language comprehension) and connections from the sensory network to the sound network (language production). But given these reciprocal connections, other functions involving language are possible. For example, the agent can receive some non-linguistic input in its sensory-motor network, this input elicits an activation pattern in the internal units of the sensory-motor

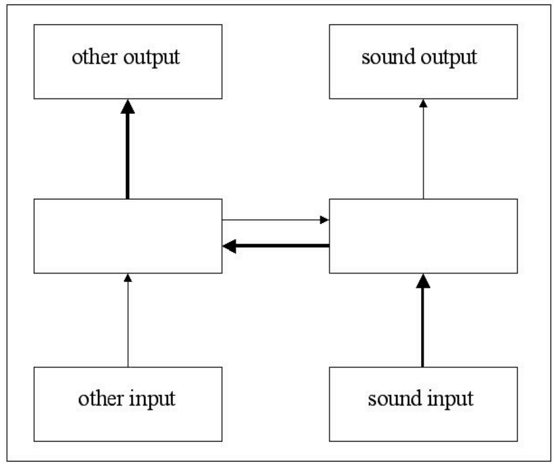

Fig. 9. Language comprehension

network, this activation pattern elicits in turn an activation pattern in the internal units of the sound network (via the connections from sensory-motor to sound network), and this activation pattern returns to the sensory-motor network (via the connections from sound to sensory-motor network). This implies that how the world is perceived and reacted to by the ECAgent is influenced by how the ECAgent linguistically labels and describes the world (Mirolli and Parisi, in press). The ECAgent lives in a linguistically commented world. Acting and thinking (talking to oneself) become intermingled.

Another dimension of architectural and functional complexity of the neural network of an ECAgent endowed with a human-like communication system derives from the multi-level compositionality of human language. Linguistic signals are made up of a succession of linguistic units: phonemes, morphemes, words, phrases, and sentences. How is this reflected in the structure and way of functioning of the ECAgents neural network?

Let us return to the sound sub-network, i.e., the sub-network which takes heard sounds as input and produces sounds via phono-articulatory movements as output. One can hypothesize that this network is made up of a succession of internal layers, one for each level of linguistic units (Figure 11). There is an internal layer for phonemes, just above the acoustic input units, followed by a layer for morphemes, then by one layer for words, one for phrases, and finally one for sentences. Each internal layer has an associated layer of memory units (Elman memory units) in which the activation pattern appearing in the corresponding internal units is copied at each cycle. These units function as a cumulative memory. For instance, given the word cats, first the sound /k/ is heard, it elicits an activation pattern in the phoneme internal layer, and this pattern is stored in the associated memory units. Then the sound /a/ is heard, this sound elicits an activation pattern in the phoneme internal units, and this

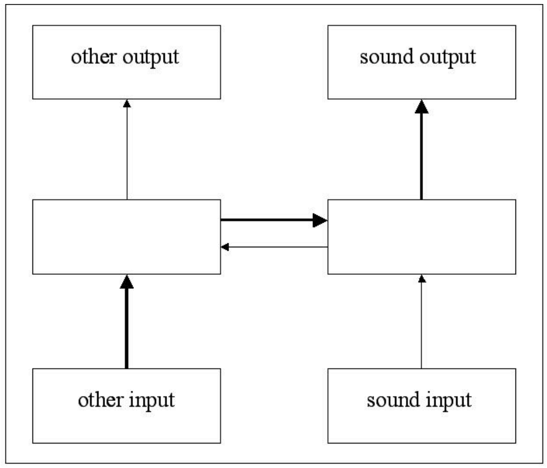

Fig. 10. Language production

pattern is also stored in the memory units together with the preceding pattern derived from the sound /k/. The same for the third sound of the word cat, i.e., the sound /t/. At this point the morpheme cat has been recognized, which means that the information which has accumulated in the memory units of the phoneme layer evokes an activation pattern in the next layer of internal units, the morpheme layer. This pattern is stored in the memory units associated with the morpheme layer. When the sound /s/ of cats is also processed, this sound is recognized as a new morpheme, its activation pattern is stored together with the activation pattern of the morpheme cat, and the two morphemes generate the word cats at the next higher level, the word layer of internal units.

How is a linguistic unit recognized? Aside from phonemes, which have no meaning, linguistic units, from morphemes to sentences, are recognized because of the connections linking the sound network to the sensory-motor network. A linguistic unit is recognized because an activation pattern in the sound network elicits an activation pattern in the sensory-motor network. The sequence of phonemes /k/ /a/ /t/ is recognized as the morpheme cat because the activation pattern elicited by the sequence of phonemes in the sound network elicit one specific activation pattern in the sensory-motor network (the meaning of cat). Notice that morphemes and words are different from phrases and sentences, though. Morphemes and words find their meanings already there in the sensory-motor network. Phrases and sentences obtain their meanings through a process of syntactic construction that can be characterized using formal rules (Chapter 10).

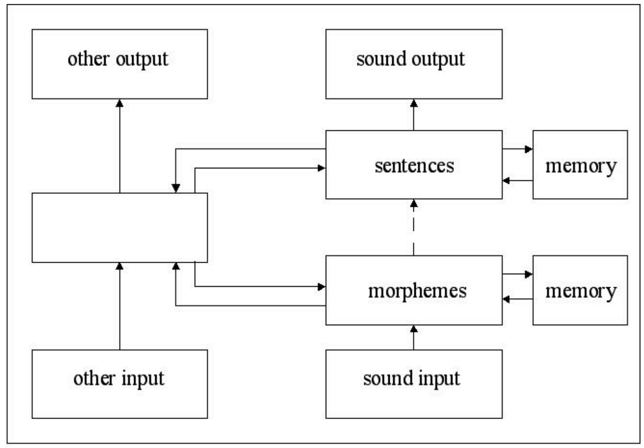

Fig. 11. Signals made of smaller signals

7.2.8 Human language influences human cognition

A final crucial difference between human language and animal communication systems is that animal communication systems do not appear to have any influence on how animals behave or think when they are not communicating whereas human language seems to lead to a rather global restructuring of the entire behavior and cognition of humans. The influence of language on human cognition is so deep and widespread that one can reasonably propose the hypothesis that language has emerged in humans not only because it is a very articulated and flexible social communication system but because it results in a much more articulated and powerful way of knowing and dealing with reality in the individual. The implication for ECAgents is that constructing ECAgents with human-like systems of communication will shape the entire behavior of ECAgents, not only the manner in which they communicate.

The influence of language on human behavior and cognition is to be linked to the fact that language is used by humans to talk to oneself and not only to communicate with others (see (3) above), and to the role that language plays in the mental life of humans, i.e., in their rememberings, thoughts, predictions, plans, etc. Humans live in a commented world, that is, in a world which is constantly linguistically labelled and described by them. They react to this commented world, not to the world as it is. However, the influence of language on human cognition may go beyond that. Language may influence cognition in humans even when humans are not speaking either to others or to themselves (thinking). The distinction can be captured by referring to the network architecture used in the preceding section to address language learning in the child. When an input is received by the sensory-motor sub-network and the input causes an activation pattern in the internal units of the sensory-motor network, two different things can happen. First, the activation pattern in the sensory-motor sub-network elicits an activation pattern in the internal units of the sound sub-network which in turn influences the activation pattern in the sensory-motor sub-network. In this way the agent is talking to itself and language can have an influence on the agents cognition. But it can also be that language has left a permanent trace in the sensory-motor sub-network itself, so that when an input arrives to the input units of the sensory-motor network, the way in which this input is internally elaborated (that is, the activation pattern it elicits in the sensory-motor sub-networks internal units) is influenced by language with no need to activate the sound sub-network.

How language can influence cognition in ECAgents and what are the consequences of having language for the behavior of ECAgents are very interesting research topics. Here are some examples of research directions that can be explored.

Categories in neural networks can be thought of as clouds of points in the abstract hyperspace that corresponds to a given layer of network units. The hyperspace has as many dimensions as are the units in the layer. One point in the hyperspace corresponds to one activation pattern that can be observed in the layers units. Each point belonging to the cloud is the activation pattern which appears in the layer of units when the agent experiences one instance of the category. Adopting an action-based view of cognition (Di Ferdinando and Parisi, 2004, in press), different experiences as put together to form a single category if the agent has to respond with the same action to all instances of the category. Learning is largely to adjust the networks connection weights so that these weights generate good clouds, that is, clouds that are as small as possible and as distant as possible from other clouds, i.e., from other categories that must be responded to with different actions. One role that language can have in cognition is that language can help the agent to have better clouds, i.e., clouds that are smaller and more distant from each other than the clouds of agents that do not have language.

Another influence that language can have on cognition is that language can allow the agent to articulate the way in which it perceives reality in ways which are suggested by language, for example isolating perceived objects that correspond to single words and separating different aspects of objects as these different aspects are separately articulated in a phrase or sentence, e.g., noun + adjective (Chapter 10).

A more general influence of language on cognition is the languages role in enlarging the agents temporal perspective on reality. Nonlinguistic agents can have both memory and prediction abilities that allow them to know and take into consideration in their behavior, at least to some limited extent, both the past and the future. However, it is clear that to maintain the past in the form of words that refer to past experiences and to articulate and to make explicit ones predictions about the future by putting these predictions in words, may greatly enlarge an agents temporal perspective on reality, amplify and articulate its overall knowledge of reality, and augment the effectiveness of its behavior.

Prerequisites related to Joint Attention

Frédéric Kaplan and Verena V. Hafner

Sony Computer Science Laboratory Paris

8.1 Introduction

Joint attention has recently received an increasing interest in the developmental robotics community. It became clearer that many of the difficulties encountered in human-robot interaction and communication between autonomous robots could be traced back to unsolved issues related to joint attention. Research in developmental psychology clearly states that skills for joint attention play a pivotal role for imitation, social cognition and the development of language. Building models for understanding the development of joint attention is certainly a crucial milestone on the road towards robots capable of some sort of social learning.

Despite an increasing number of works dealing with joint attention, existing computational and robotic models do not seem to agree on the central issues to be solved. For instance, in a recent paper, Nagai and colleagues describe "a constructive model that enables a robot to acquire the ability of joint attention" without a controlled environment nor external task evaluation (Nagai et al., 2003). Although this paper definitively makes an interesting contribution for understanding how a robot could learn to interpret human gaze in order to spot salient objects in its environment, it could be argued that it does not cover all the aspects of joint attention. Indeed, another model presented by Ikegami and Iizuka considers that the development of joint attention is closely related to the emergence of turn-taking behaviours, a rather different issue (Ikegami and Iizuka, 2003). The heterogeneity of these approaches gives a puzzling picture of this clearly important but ill-defined process.

We discuss in this chapter the concept of joint attention and the different skills underlying its development. In the line of Tomasello's views (Tomasello, 1995; Tomasello et al., 2004), we argue that joint attention implies viewing the behaviour of other agents as intentionally-driven. In that sense, joint attention is much more than gaze following or simultaneous looking. Summarising results from developmental psychology, the chapter presents a timeline showing at what age the different prerequisites for joint attention arise during the first two years in the life of a child. In relation with this developmental timeline, the chapter reviews the current state-of-the-art in robotic and computational models of joint attention and identifies which issues remain to be addressed.

8.2 What is Joint Attention?

8.2.1 Defining attention

Before discussing what joint attention is, the first step is to agree on a non-controversial definition of attention. Attention can be defined as the process whereby an agent concentrates on some features of the environment to the (relative) exclusion of others. This process can occur in two situations.

1. **Passive attention**: a salient event happens (e.g. loud noise) and automatically triggers the attention of the agent.
2. **Active attention**: the agent is involved in an intentionally directed process (e.g. climbing a mountain) and must actively select particular features of its environment.

The attentional behaviour is the externally perceivable behaviour that goes along with the attention process. To reach joint attention, agents must actively track and manipulate the attention of each other. Discussing the prerequisites of this coordination is the aim of this chapter. But before that, we must specify what we mean by joint attention.

8.2.2 Defining joint attention

*Joint attention is **not** simultaneous looking* Joint attention is often associated with a situation where two agents are looking at the same thing. We will now examine four cases of simultaneous looking which do not qualify for joint attention. For better illustration, we use examples of interaction between two robots (Figure 1).

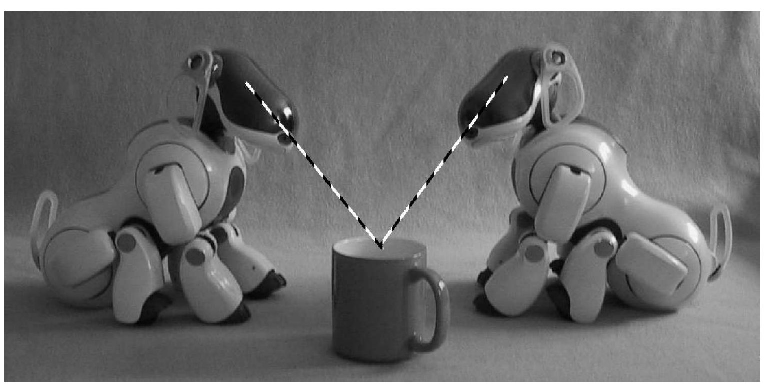

Fig. 1. Two Sony AIBO robots are looking simultaneously at a coffee cup. Is this already joint attention?

Case 1a: Simultaneous looking triggered by a salient event (passive attention). The two robots are sitting in a room. Suddenly, one of their toys makes a squeaking noise. They both turn and look at it immediately.

Case 1b: Simultaneous looking triggered by a "pop-out" effect (passive attention). The robots found a box filled with balls. All the balls are blue, apart from one which is pink. Both robots are attracted by the pink ball.

Case 2: Coincidental simultaneous looking. The robots are looking for a toy to play with. At the same moment, they both see a pink ball on the floor. They pay attention to it without noticing each other. Each other's attention is not monitored.

Case 3: Gaze following. One robot is looking at a new toy. The other less experienced robot follows his gaze since it has learned that by doing that, it will often see something interesting. But attention is not joint, as the first robot is not paying attention to the behaviour of the other one.

Case 4: Coordinated gaze on an object. Both robots are looking at a toy bunny, and are also aware that the other one is looking, too. From an outside observer's point of view, this situation looks like joint attention. However, one robot is attending to the bunny in order to play with it, the other one is purely attracted by its colour. They are therefore not attending to the same aspect of the object.

These different cases of simultaneous looking are summarised in table 1. For an outside observer, these cases might still seem like examples of joint attention when taken out of context, however they are not.

Joint attention as a shared intentional relation to the world Joint attention is an active bilateral process which involves attention alternation, but it can only be fully understood if we assume that it is realized by intentional agents. Active attention occurs when an agent is involved in an intentionally directed process. This means that the agent tries to achieve a particular desirable situation that constitutes its aim or **goal** (e.g. being on top of a mountain, reducing hunger, following someone, learning something). The **intention** is the plan of action that the agent chooses for realizing this particular goal. This plan includes both the means and the pursued goal (Tomasello et al., 2004).

To realize its aim, the agent focuses selectively on relevant perceptual features. In that sense, attention is intentionally directed perception (Tomasello, 1995). Its intentional behaviour is also associated with particular emotional responses corresponding to progress, successes and failures in the pursuit of the goal. The only way for an agent to read the intention of another agent is by watching its behaviour. Here are a few examples:

Example 1: Intention detection through general behaviour. One robot sees another robot walking towards the charging station. He infers that his battery is low and that he needs to recharge. In this case, the observer did not need to track the other one's attention to understand the underlying intention.

Example 2: Intention detection through attentional behaviour. One robot is looking attentively at the closed door. The other robot infers that it attends to the door because it wants to go outside. Here, tracking the attentional behaviour is relevant to understand what the other robot intends to do.

Example 3: Intention detection through emotional behaviour. One robot kicks a ball which then hits a toy. The robot emits a 'sad' sound, goes to the ball and kicks it again. Now the ball rolls through the door to another room. The robot emits a 'happy' sound. Based on these two signals the other robot can interpret that it wanted to kick the ball out of the room.

To reach joint attention an agent must understand, monitor and direct the intentions underlying the attentional behaviour of the other agent. Joint attention can only be reached if both agents are aware of this coordination of "perspectives" towards the world (Hobson, 2002). In the same way that attention cannot be reduced to visual orientation, joint attention is much more than a geometrical phenomenon. It needs to be understood as a crucial step in the development of social cognition.

8.2.3 The prerequisites of joint attention

Reaching joint attention implies at least four kinds of prerequisites.

- **Attention Detection.** An agent must be able to track the attentional behaviour of other agents. This may imply being able to follow the gaze of another agent.
- **Attention Manipulation.** Agents must be able to manipulate the attentional behaviour of other agents. The use of pointing gestures or words can be used in that respect.
- **Social coordination.** Agents must be able to engage in coordinated interaction with other agents. This implies mastering social techniques such as turn-taking, role-switching and ritualised games.
- **Intentional understanding.** Agents must view themselves and others as intentional agents. They must understand that others have intentions possibly different from their own. Agents capable of intentional understanding interpret and predict the behaviour of other agents in terms of means used to reach particular goals.

The rest of the chapter examines data drawn from developmental psychology on the development of these capabilities and discusses existing robotic and computational models for each of them. Distinguishing between these four kind of skills helps clarifying the developmental picture underlying the emergence of joint attention. However, we do not claim that these different prerequisites appear from independent developmental pathways. On the contrary, it could be argued that, at several stages of this developmental process, skills for attention detection, attention manipulation, social coordination and intentional understanding are intrinsically linked.

Table 1. Different cases of simultaneous looking

Case	Active / Passive	Attention detection	Unilateral / Bilateral
Case 1: Simultaneous Looking triggered by a salient event or a "pop-out" effect	Passive	No	-
Case 2: Coincidental simultaneous looking	Active	No	-
Case 3: Gaze following	Active	Yes	Unilateral
Case 4: Coordinated gaze on same object	Active	Yes	Bilateral

8.3 Developmental Timeline

We will now discuss at what age the different skills and prerequisites for joint attention arise in young children during their development. Table 2 presents these skills in the temporal order in which they occur first between three and 24 months when joint attention is fully developed. For better illustration, some situations on attention detection and attention manipulation are displayed in figure 2 using Sony AIBOs. Several of these developmental landmarks are subject to controversial arguments. Some of these controversies are shortly discussed in this section. But discussing the detailed experimental evidences underlying each milestones is beyond the scope of this review. This timeline is only intended to give a general overview of the parallel development of each prerequisite of joint attention.

8.3.1 Attention detection

In the first month of their lives, children progressively bootstrap the capability to pay attention to a growing number of things in their environment: their own body, external objects, animate beings, etc. During this developmental process, they start paying attention to the attentional behaviour of other agents.

T1.1 Mutual gaze. (Figure 2a) Mutual gaze between an adult and a child occurs first around the age of three months. At this age, the child shows a strong preference towards face-like patterns and is capable of recognising and maintaining eye contact. This sensibility of eye contact is also reported in the behaviour of many

Table 2. Developmental timelines of the prerequisites for joint attention

Age from:	**Attention detection**	**Attention manipulation**	**Social coordination**	**Intentional understanding**
0-3 m	**T1.1 Mutual gaze** - Eye contact detection		**T3.1 Protoconversations**: Simple rhythmic interaction including turn-taking mediated by the caregiver.	**T4.1 Early identification with other persons**
4 m			**T3.2** Possibility of breaking interactions	
6 m	**T1.2** Discrimination between **left and right position** of head and gaze		**T3.3 Shared games**: Conventional routines established between the child and the caregiver	**T4.2 Animate-inanimate distinction**: discrimination between physical and social causality
9 m	**T1.3 Gaze angle detection** - fixation on the first salient object encountered	**T2.1 Imperative Pointing**: Drawing attention as a request for reaching an object (attention not monitored)	**T3.4 Simple immediate imitation**: The child commonly imitates a movement performed by the caregiver. Evidence of capabilities for sequence learning.	**T4.3** First **goal-directed behaviours**. Evidence of domain-general inferential abilities
12 m	**T1.4 Gaze angle detection** - fixation on any salient object encountered - Accuracy increased in the presence of a pointing gesture	**T2.2 Declarative Pointing**: Drawing attention using gestures		**T4.4 Goal understanding**. Observed behaviour understood as goal-directed
13 m		**T2.3 Declarative/Referential words**: Drawing attention using a word		
18 m	**T1.5 Gaze following** toward object outside the field of view - Full object permanence	**T2.4 First predications**: Drawing attention using non-linguistic gesture for the topic and a word to specify which aspect of the object should be attended	**T3.5 Complex imitative games** Social exchanges using imitation including conventional routines and role-switching	**T4.5 Intentional understanding**. Children understand that different action plans can be associated with the same goal.
24 m		**T2.5 Conversations**: Both topic and aspect can now be specified linguistically		

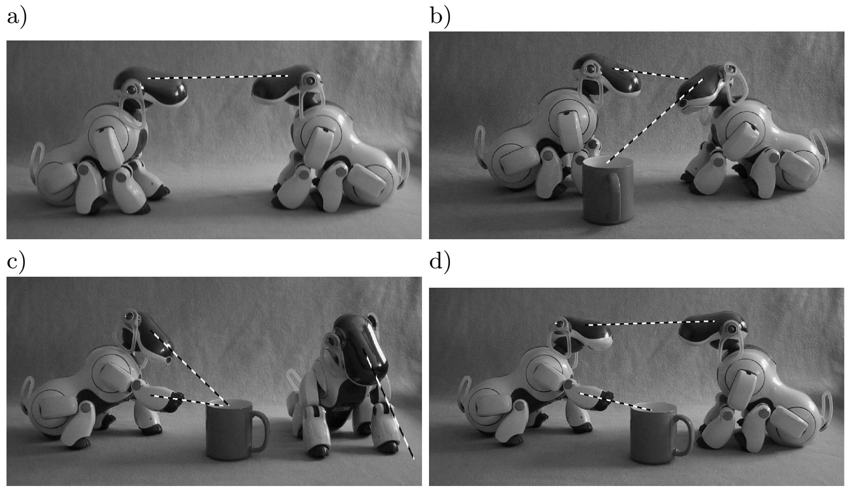

Fig. 2. Demonstration of different situations preceding joint attention during development. a) Mutual Gaze. Both robots are attending to each other's gaze simultaneously. b) Gaze Following. One of the robots is paying attention to an object, the other one watches its eyes in order to detect where it is looking. c) Imperative Pointing. Pointing to an object regardless whether another person or robot is attending. d) Declarative Pointing. Pointing to an object to create shared attention.

animals, in particular in primates (Cheney and Seyfarth, 1990). Mutual gaze is a special case of attentional behaviour since it does not involve any objects or persons apart from the two involved.

T1.2-5 Gaze following. (Figure 2b) At the age of six months, the first true incident of attention detection starts. The child is able to attend to an object in the correct side of the room depending on where the adult is looking at (T1.2). The angle error between the attended object of the adult and the attended object of the infant can be as large as 60 degrees (Butterworth, 1995). Only at the age of nine months can the gaze direction of the adult be accurately detected, however, always the first object within the line of sight is chosen (T1.3). The correct object can be attended to by the age of twelve months (Butterworth and Jarrett, 1991) taking into account vergence and probably context (T1.4). By this age, only objects which are in the field of view of the child are being considered, even though the child is already turning to sounds coming from behind (Butterworth, 1995; Butterworth and Cochran, 1980). Only at 18 months, children start following the gaze of an adult to objects outside their field of view (T1.5). If directing the gaze towards an object is supported by also pointing towards that object, the accuracy of attending to the correct object increases in infants older than twelve months (Butterworth, 1995). Before that age, pointing is not understood by the child and does not make any difference to the child's attention.

8.3.2 Attention manipulation

Skills which fall into the category of attention manipulation are the act of pointing at something and the use of language. We differentiate between drawing attention to oneself and to others or other objects since the first ability is already present in the first month of a child's life.

T2.1 Imperative pointing. (Figure 2c) The first occurrence of pointing, imperative pointing, starts around the age of nine months (Baron-Cohen, 1997). Imperative pointing is the request for a certain object, using a gesture. Imperative pointing might be an extension of grasping an object, and it also occurs when nobody who could pay attention is present in the room. This means that the attention is not monitored.

T2.2 Declarative pointing. (Figure 2d) At twelve months, shortly before the use of linguistic symbols, pointing starts to become declarative. It is used to draw someone's attention to something which might also be outside of reach for the adult, such as objects like the sun or an aeroplane. One could think that this pointing behaviour results from an imitation of the gestures of the adult. However, some studies with young children found no relation between the production of pointing and the comprehension of pointing (Desrochers et al., 1995). This would mean that attention directing skills emerge independently from capabilities in attention following. This issue is still under debate.

T2.3 Declarative/Referential words. After drawing attention using gestures, the child starts to use single words to draw attention to objects or persons around the age of 13 months.

T2.4-5 First predications. First predication follows at about 18 months, and already requires building of a simple context representation. At this age, the child specifies the subject of interaction by pointing and then adds a comment linguistically in order to draw the attention of the adult towards a particular aspect of it (e.g. "big") (T2.4). By the age of 24 months, both the topic and the comment start to be expressed linguistically (e.g. "big dog") (T2.5).

8.3.3 Social coordination

Social coordination is a crucial element for the development of social cognition. Starting from simple shared rhythmic patterns, children manage to engage in increasingly complex routines with their parents. In the first months, these "games" are usually initiated by the parent, but become more symmetrical later on. The structure of interactions becomes conventionalised through negotiation processes involving child and parents. Like good dancers, children learn to find the right equilibrium between following the rhythm and breaking it to keep the interaction entertaining.

The development of social coordination is not limited to behavioural patterns. Through interpersonal couplings, children and caregivers adapt to coordinate emotions, perspectives and goals. From early one-to-one interactions (dyadic), more complex coordination patterns gradually emerge involving external entities (triadic). This step is tightly linked with the development of new attention detection and attention manipulation skills, as well as new forms of behavioural understanding. Coordination extends in time, involving longer shared plans.

T3.1-2 Protoconversation. Six-week old children are already communicating extensively face-to-face with their caregiver. These first simple rhythmic interactions are crucial for the development of social know-how (Trevarthen, 1979). Newson argues that these early social responses are treated by the adult as normal social behaviour (Newson, 1979). For instance when the child does something that can be interpreted as role switching or change in the course of the “dialog”, the adult adapts in order to make it become meaningful. In such conditions, these proto-dialogs exhibit already simple turn-taking behaviours. As the adult scaffolds these interaction into structured dialogs, children learn to predict the social effects of their behaviour (Schaffer, 1977). By the age of four months, children are able to break their caregiver’s gaze in order to look at other things in the world (Siegel, 1999) (T.3.2). This opens to the possibility of more complex interactions.

T3.3 Shared games. Each caregiver develops his or her own set of conventional games. By the age of six months, a child manages to master an important number of them. These ritualised structures play a crucial role for defining roles and imposing consistency and predictability in social exchanges (Kaye, 1982). A key point is that games are not simply learned by the child in a passive way. Each conventional routine is the result of a negotiation, where both the child and the caregiver adapt in order to reach a common coordination pattern.

T3.4-5 Imitative games. A common interaction routine consists in the immediate imitation by the child of a movement produced by the caregiver (T3.4). This skill, already present in the very young infant, gradually develops and is used commonly around nine months. Nadel has emphasised the role of such immediate imitations for bootstrapping social exchanges in particular for turn-taking, role switching and in order to share topic (Nadel, 2002). Around 18 months, it starts to be used inside complex games as children manage to understand observed behaviour in terms of goals and means (T3.5). Other forms of complex social interaction appear in parallel. They involve the same components: coordination of action, attention and intention patterns.

8.3.4 Intentional understanding

Tomasello argues that a crucial behavioural transition occurs around twelve months (Tomasello, 1995). Before one year, children begin following and directing the attention of other persons, but do not view them as intentional agents. At the beginning of the second year of their life, they demonstrate a qualitative change in the nature of their behaviour. Complex social skills such as social referencing, imitative learning or symbolic communication with gestures appear almost simultaneously (see table 2). This synchrony suggests that a radical shift has occurred in children's awareness of their environment: they have developed intentional understanding.

There is a vast range of theories on how to interpret this shift from totally nativistic to totally cultural hypotheses. For instance, Trevarthen argues that children view other persons as intentional agents from birth, independently from any prior experience (Trevarthen, 1979). Similar views are supported by other authors who consider that humans are hardwired from birth to interpret autonomous behaviour as intentional (Asch, 1952; Premack, 1990). On the opposite side, other researchers like Kaye believe that children construct the notion of intentional agents totally from experience. During the first year of their life, an important part of children's experiences are mediated by the parents. The fact that parents treat children as intentional agents even before they are such may also play an important role for their development of intentional understanding ("parents create persons") (Kaye, 1982). These views are sometimes criticised on the ground of the important cultural differences that exist around the world about the way to nurture young children.

The kind of skills needed to achieve intentional understanding are less easy to identify than for the other prerequisites of joint attention, and the related developmental timelines are often controversial. Several authors have stressed that the intentional understanding involves at least two kinds of capabilities: **parsing skills** and processes for **making inferences and plans about hidden states** (Baird and Baldwin, 2001; Povinelli, 2001; Wellman and Phillips, 2001).

Parsing consists in discovering statistical regularities and segmenting observed behaviours into separated action-units. For each action-unit, relevant perceptual features must be spotted for anticipating the following sequences of actions. For instance, statistical regularities about attentional behaviour toward objects can inform about the target that an agent tries to reach.

Intentional understanding might also imply the development of prediction systems capable of handling hidden states (not directly perceivable) such as goals, emotions or tastes of others. Moreover, intention systems are typically structured in a hierarchical manner. Goals at one level are realized through subgoals and take part of higher action plans. Handling such embedded structures requires complex prediction systems.

These two kinds of processes are likely to work in close concert guiding rapid processing and interpretation of others. Their development may be closely coupled (Baird and Baldwin, 2001) but they may also result from independent developmental (or evolutionary) histories. Povinelli in particular argues that apes display some advanced form of behaviour parsing but are not capable of making complex inferences about mental states of others (Povinelli, 2001).

Other data suggest that at least some aspects of intentional action can be understood by apes (Tomasello et al., 2003).

Eventually, detecting cues of intentional behaviours and reasoning about mental states may not be sufficient in the absence of a process to match and discriminate one's own action with the ones of others. This **identification between self and others** is a necessary developmental step for the acquisition of intentional understanding. Let us now consider more precisely when these different skills in the first two years of a child's life arise.

T4.1 Early identification. Early identification with other persons, taking the form of simple imitative behaviours, has been observed in the first months of life. To explain these experiments, some totally or partially nativist theories have been put forward (Meltzoff and Gopnick, 1993; Moore and Corkum, 1994). Whatever their innate basis is, these neonatal forms of imitation make children exposed to situations in which their intention and the one of the adult happen to converge. They may play a role for the progressive distinction by the child of the first and third person perspectives.

T4.2 Animate/inanimate distinction. Distinction between animate and inanimate objects is thought to emerge gradually during the first six months of a child's life. Discrimination of moving objects is observed at birth. Early sensibilities to self-propelled movement and discrimination between mechanical and biological motions have been experimentally reported for two-months old children (Bertenthal, 1996). At six months, children have been shown to distinguish between physical causality (to pull, to push) and social causality (to pursue, to avoid) (Rochat et al., 1997). 7-month-old children recognise that humans can cause one another to move in the absence of physical contact but that inanimate objects like blocks cannot (Woodward et al., 1993). Other experimental evidence shows that by this age, some form of distinction between animate and inanimate entities is active (Poulin-Dubois, 1999) (Sperber et al., 1994). Children at this age may predict what animate actors will do in familiar situations, but not in novel ones. This suggests that although they understand animate action, they do not yet reason in terms of goals and intentions.

T4.3 Goal-directed behaviours. Piaget observes that children first start to display goal directed behaviour around nine months (Piaget, 1952). They may for instance remove an obstacle in order to reach a particular place. This means that they start to differentiate goals and means in their own behaviour and view their own behaviour as intentionally-driven. At the same age, children also show a beginning of awareness that some actions they observe are directed towards particular objects (Wellman and Phillips, 2001). This shows initial competencies in behaviour parsing. More generally, 9-month-old children have been shown to possess domain-general inferential abilities that may serve as the basis for making inferences about intentions (Baldwin et al., 1993).

T4.4 Goal understanding. Goal-directed behaviour becomes common around twelve months (Frye, 1991). Extensions of this discrimination for the interpretation of the behaviour of other agents may occur as a consequence of this first finding (Tomasello, 1995). At this age, children can infer the causal links between actions of others and detect behavioural regularities between gaze direction and goal-directed motor sequences. For instance they may be surprised if someone looks at one toy and then grabs another one (Wellman and Phillips, 2001). Children at this age may understand that observed actions are directed toward some particular target states, and recognise successes and failures in repeated attempts. However, Tomasello suggests that they do not yet understand that various plans (intentions) can be associated with the same goal (Tomasello et al., 2004)

T4.5 Intentional understanding. Experimental evidences that infants understand other's goals and intentions multiply at 18 months. At this age, children who watched an adult engage in an unsuccessful behaviour imitate the model by producing the intended action instead of the observed one ((Meltzoff, 1995), see also (Carpenter et al., 1998) for similar experimental evidences). In other experiments, 18-month-old children are shown to adapt to an unspecific request like 'give me some more' by taking into account information that the adult previously displayed about his tastes and desires (Repacholi and Gopnik, 1997). Several other experimental results show that at this age (and even a few months before), children start to be capable of linking the means used with the targeted goals and to analyse observed behaviour in those terms (Tomasello et al., 2004). This new understanding serves as a basis for efficient social learning.

8.4 Robotic and Computational Models

The precise developmental route that leads to mastering the necessary skills for joint attention is largely unknown. Robots are ideal tools to model the development of joint attention. Their embodiment in the real world allows for interactions between robots as well as interaction between humans and robots. Experiments are - in contrast to observing the behaviour of children - repeatable and different aspects can be easily separated. The idea is not to directly match data obtained in robotic experiments with quantitative results of the developmental psychology literature. Computational and robotic models are to be understood as a source of inspiration for psychology. By showing which qualitative behaviours emerge out of a particular software architecture, physical embodiment and environmental conditions, these models may shed new light on observations made during children experiments.

In this section, we review the state-of-the-art research in developmental robotics concerning joint attention and its various prerequisites. No system has yet achieved true joint attention between a robot and a human or between two robots in the sense we defined it in the previous sections. Several crucial steps have started to be investigated, but important parts of this developmental puzzle are still unexplored.

8.4.1 Models for attention detection and attention manipulation

Table 2 shows that the child manages to make progress in detecting and manipulating the attention of the adult through a series of steps of increasing complexity. Some of these skills have already been designed by hand on a robot. Imai et al.'s robot 'Robovie' (Imai et al., 2001) is able to attract a human's attention by pointing at an object and establishing mutual gaze. Kozima et al. (Kozima and Yano, 2001) have designed the robot called 'Infanoid' that can track human faces and objects with salient colour (T1.1), point to and reach for objects (T2.1), and gaze alternatively between faces and objects (T1.2-4).

Scassellati describes how he intends to accomplish joint attention between the robot and a human, but he mostly concentrates on issues related to attention detection (Scassellati, 1999). So far, only the eye contact has been implemented on the robot 'Cog'. Applied techniques are face detection using ratio templates (Sinha, 1996) and eye extraction (T1.1).

Some researchers tackle the development of attention detection, as opposed to simply designing a system capable of doing it. Carlson and Triesch (Carlson and Triesch, 2003) present a computational model of the emergence of gaze following based on reinforcement learning. They identify a basic set of mechanisms sufficient for the development of this skill. The model has been tested in a virtual environment by Jasso et al. (Jasso et al., 2004). Hafner and Kaplan demonstrate how four-legged robots can learn to interpret pointing gestures of one another. One of the robots takes the role of an adult and is pointing to an object, the other robot, the learner, has to interpret the pointing gesture correctly in order to find the object (Hafner and Kaplan, 2004). Nagai and colleagues describe a learning module that learns the correlation between the gaze of a human and an object in the visual field at a certain position. The robot progressively learns to use the human gaze in order to find objects more rapidly (Nagai et al., 2002, 2003). This corresponds to the acquisition of gaze following (T1.2-5).

Several issues concerning the development of attention detection and manipulation have not been addressed yet. How can pointing emerge from grasping behaviour (T2.1)? How does declarative pointing appear (T2.2)? By which process can words replace gestures for drawing attention (T2.3)? On which basis does predication appear (T2.4, T2.5)?

8.4.2 Models for the emergence of social coordination

Several robotic experiments have emphasised the importance of structured interactions (T3.3) for the development of higher social skills like language acquisition (Breazeal, 2002; Steels and Kaplan, 2000b; Steels et al., 2002b; Steels and Kaplan, 1999b), but a limited number of works has addressed the problem of how shared interaction routines necessary for coordinating behaviour in joint attention may develop.

Ikegami and Iizuka (Ikegami and Iizuka, 2003) use robots in a simulated environment to study turn-taking. Their experiment demonstrates the evolution of a turn-taking behaviour for two robots when a fitness function explicitly favours such a behaviour (T3.1). Andry et al. (Andry et al., 2001) report several experiments where a robot demonstrates immediate imitation for simple motor skills (T3.4) and discuss how simple architectures could account for the emergence of rhythmic interactions (T.3.1) including the possibility of breaking rhythm (T3.2). Ito and Tani present an experiment where a human and a humanoid robot engage in stable and unstable phases of interaction using particular entrainment dynamics (T3.2) (Ito and Tani, 2004). Imitation has recently been an important topic of investigation (Dautenhahn and Nehaniv, 2002) but only a few works investigate its role for social coordination.

Most of the work remains to be done for this aspect of joint attention. What kind of reward structure must be present so that interaction and entrainment spontaneously emerge (T3.1)? What dynamics lead to the formation of turns during the interaction (T3.1)? How is the structure of new games captured (T3.3)?

8.4.3 Models for the emergence of intentional understanding

How can a robot start to view the behaviour of another robot as intentional? Which techniques can it use to parse the behaviour of others in a meaningful way? How can it start making inferences about hidden states? Almost no prior art in the developmental robotics literature deals explicitly with these issues.

Goals and intentions are of course central issues for classical artificial intelligence. Research in this area has influenced the way we consider decision making or planning. More recently, research on agent architectures (Dignum and Conte, 1998) has put a major emphasis on the same issues. However their models do not give much insights on the developmental and cognitive mechanisms that lead to the notion of intentionally-directed behaviour.

Behaviour parsing has been indirectly addressed by a variety of experiments in research concerning the symbol grounding and anchoring problem (Harnad, 1990; Coradeschi and Saffiotti, 2003). Most works implement a set of perceptual primitives capable of extracting relevant features in action sequences (e.g.(Roy and Pentland, 2002; Siskind, 2001; Dominey, 2003; Steels and Baillie, 2003)). But these models do not address the issue on how such perceptual primitives may arise in a developmentally convincing way. Moreover, most of these works present experiments done in very carefully controlled environments in order to obtain satisfactory results with state-of-the-art artificial vision techniques. Indeed, object segmentation and recognition are very difficult to perform in real complex environments, especially when templates of the targeted objects are not known in advance. Behaviour parsing stays an open issue for robotics.

In research on imitation, some authors have investigated the problem of “what to imitate” in the observed behaviour of another agent (e.g. (Alissandrakis et al., 2000)). They address the issue on how to decompose and recreate an observed behaviour. These questions can be considered to be central for the emergence of behaviour understanding (T4.4-5). But they are only part of the picture.

Intentional matching remains also an underinvestigated issue. Taking inspiration from animal training techniques, Kaplan et al. showed how a robot could try to model its user's expectations and adapts in order to perform a particular desired behaviour while keeping its general behavioural autonomy (Kaplan et al., 2002). However the robot did not develop intentional understanding by itself.

The development of intentional understanding is probably the most challenging prerequisite that research on joint attention has to investigate. None of the milestones that we have identified in our timeline seem to have been already addressed in a satisfactory manner by computational or robotic models. What are the mechanisms or dynamics that enable an agent to identify itself with other agents of the same kind (T4.1)? How can it make the distinction between animate and inanimate entities (T4.2)? How can a robot discover the goal-means distinction if these notions are not already explicit in its internal architecture (T4.3-4)? How can it apply this insight to interpret the behaviour of other agents (T4.5)?

8.5 Conclusions

Table 3. Open questions and challenges for joint attention in robotics

Attention detection and manipulation	Social coordination	Intentional understanding
How can pointing emerge from grasping behaviour (T2.1)? How does declarative pointing appear (T2.2)? By which process can words replace gestures for drawing attention (T2.3)? On which basis does predication appear (T2.4, T2.5)?	What kind of reward structure must be present so that interaction and entrainment spontaneously emerge (T3.1)? What dynamics lead to the formation of turns during the interaction (T3.1)? How is the structure of new games captured (T3.3)?	What are the mechanisms or dynamics that enable an agent to identify itself with other agents of the same kind (T4.1)? How can it make the distinction between animate and inanimate entities (T4.2)? How can a robot discover the goal-means distinction if these notions are not already explicit in its internal architecture (T4.3-4)? How can it apply this insight to interpret the behaviour of other agents (T4.5)?

The development of joint attention between a human and a robot or between two robots depends on the successive appearance of a number of underlying skills. The aim of the present chapter is to **identify the challenges** and to pinpoint what kinds of results are still to be obtained in order to succeed in this goal. The overall picture that arises from this survey is a fragmented puzzle. Important research efforts currently focus on skills for attention detection, but most of the issues regarding the other prerequisites are only partially modelled (Table 3). The most underinvestigated aspects of this problem is the modelling of the mechanisms responsible for the emergence of intentional understanding. Understanding this crucial step in child development would open up the way to the creation of robots with a qualitatively different kind of awareness, making the problems of social learning easier and ultimately leading to the development of true joint attention.

The challenges of joint attention show tight similarities with the challenges of imitation, which currently receive much attention in the developmental robotics community (Dautenhahn and Nehaniv, 2002). Likewise, the emergence of imitative capabilities involves attention detection, social coordination and intentional understanding. Understanding the interplay between the development of these prerequisites is the core issue of these two problems.

The contribution of computational and robotic models can take two forms. Most of the models focus on a single developmental step (e.g. showing the emergence of gaze following when an adequate reward system is present). The increasing number of models permits a better understanding of what are the easy and hard parts of the problem. However, by studying the development of each prerequisite in a separated manner, these models may not capture synergetic dynamics linking their parallel development. Instead of designing different models to study independently attention detection, attention manipulation, social coordination or intentional understanding, one strategy could be to build architectures with generic developmental principles and to study which embodiment and environmental conditions lead to the simultaneous development of these skills. Current results obtained with a generic architecture for autonomous mental development may be considered too preliminary to deal with issues like joint attention. Nevertheless, such models may offer interesting new perspectives by explicitly addressing the links between the development of perception, action and interpersonal coupling.

To be understood properly, the development of joint attention must be understood as a whole. To account for the complete picture, models must reenact the coordinated development of skills like gaze following, declarative pointing, ritualised games, behavioural parsing, intentional inferences and matching. These interconnected challenges account for the pivotal role played by joint attention in the development of social cognition. We believe robots are ideal tools to make progress in the study of these complex issues.

Prerequisites related to the Communication Medium

Pierre-Yves Oudeyer

Sony CSL Paris

9.1 The physical medium of communication: the need for a system of labels

Communication relies on the existence of a physical system which can carry information. For example, humans can use speech sounds, visual signs or writing. Each of these systems is a conventional code which defines how humans can use the physical medium to build forms which can carry information. For example, the speech system is a code which defines for each linguistic community how the continuous space of sounds is broken up into perceptual categories. These categories are used to build a shared set of acoustic labels by which human can speak to each other. It is a necessity that embodied communicating agents possess this sort of shared systems of physical forms, which we will call “label systems”. It is not reasonable to think that a label system could be pre-programmed into artificial communicating agents, because it is both unpractical and not flexible: agents with different label systems can not communicate.

If one wants to build embodied communicating agents, then it is a major challenge to provide them with mechanisms which lead to the formation of shared label systems which can transmit information in a robust and adaptive manner in noisy and changing environments. A constraint that should be always applied is that we should not pre-suppose anything about the properties of the physical medium on which they will develop form systems. For examples, this might be sound or vision through real sensors and actuators, and not only electrical impulses through cables.

Yet, this objective of building embodied communicating agents can be reached if we take inspiration from existing label systems, such as the human speech code. As a consequence, the technological challenge calls for a scientific challenge: can we understand the principles of formation of human label systems? We will now detail the example of the human speech code, This will allow us to explain more precisely both these scientific and technological challenges.

9.2 The human speech code: an example of robust and adaptive label system

The human speech code is of extreme interest because it is a label system both very efficient and robust working on a high-dimensional continuous noisy inhomogeneous physical medium. Indeed, the vocalization of a word, i.e. a sequence of phonemes, is achieved through the physical movement of many articulators in the human vocal tract, combined with the vibration of the glottis, which produces an acoustic wave, which propagates in a noisy environment full of other uncorrelated sounds, and then activates the cochlear neurons of the receiver. Of course, the vocalization of the same word by the same speaker never produces exactly the same movements of the articulators, because they are sensitive to many external parameters like the speed of talking, the degree of arousal or the degree of humidity in the mouth. Moreover, each speaker has a different vocal tract and a different acoustic sensibility. Yet, humans can pronounce and recognize easily five words per seconds, which is about 20 phonemes per seconds. If we reformulate this in terms of information theory, the speech code is an error correcting code which is highly efficient. But to the difference of all error correcting codes which are used in the computer technologies to transmit "bits" over noisy channels, like hash-codes or turbo codes, the human speech code is not pre-wired and fixed: it is a culturally established adaptive code. Its strength is that is builds on a mechanism which allows agents who do not already share conventions to develop one that will allow them to communicate. On the contrary, a machine which has a pre-programmed code like hash-code to transmit labels to other machines will never be able to communicate with a machine which has a pre-programmed turbo code to transmit label to other machines.

This is why it is interesting to look more closely at some crucial properties of the human speech code:

- **Property 1 (discreteness and combinatoriality)**: speech sounds are phonemically coded as opposed to holistically coded. This implies two aspects: 1) in each language, the continuum of possible sounds is broken into discrete units; 2) these units are systematically re-used to build higher level structures of sounds, like syllables.
  For example, in articulatory phonology (Browman and Goldstein, 1986), a vocalization is viewed as multiple tracks in which gestures are performed in parallel (the set of tracks is called the gestural score). A gesture is the combination of several articulators (e.g. the jaw, the tongue) to perform a constriction somewhere in the mouth. The constriction is defined by the place of obstruction of the air as well as the manner. While for example, given a sub-set of organs, the space of possible places of constrictions is a continuum (for example the vowel continua from low to high, executed by the tongue body) each language uses only a few places to perform gestures. This is what we call discreteness. Furthermore, gestures and their combinations, that may be called "phonemes", are systematically re-used in the gestural scores who specify the syllables of each language. Some researchers call this "phonemic coding".
- **Property 2 (phonotactics and patterns)**: The way phonemes are combined is also very particular: 1) only certain phoneme sequences are allowed to form a syllable in each language, the set of which defines the phonotactics of the language (for example, "spink" is a possible syllable in English, but "npink" and "ptink" are not possible; in Berberian, "tgzmt" and "tKsmt"

are allowed, but impossible in French); 2) the set of allowed phoneme combinations is organized into patterns. This organization into patterns means that for example, one can summarize the allowed phonemes of Japanese by the patterns "CV/CVC/VC", where "CV" for example defines syllables composed of two slots, and in the first slot only the phonemes belonging to a group that we call "consonants" are allowed, while in the second slot, only the phonemes belonging to the group that we call "vowels" are allowed. The phonotactics of a language introduce constraints which help in speech recognition: it provides equivalents of error correcting codes when the hearer of an utterance is in a noisy environment.

– **Property 3 (sharing)**: the speakers of a particular language use the same phonemes and they categorize speech sounds in the same manner. Yet, they do not necessarily pronounce each of them exactly the same way. They also share the same phonotactics. This is what makes it a conventional code.
– **Property 4 (diversity)**: At the same time, each language categorizes speech sounds in its own way, and sometimes does it very differently from other languages. For example, Japanese speakers categorize the "l" of "lead" and the "r" or "read" as identical. Different languages may also have very different phonotactics. This shows the flexibility with which the acoustic physical medium can be used by different societies of speakers. This also shows that the speech code is really created culturally, and not pre-coded in the genes of speakers.

9.3 Challenges

9.3.1 Learning an existing code

A first challenge is to understand how human infants acquire the human speech code, and how one could build a machine which performs the same task. This involves several levels of skills:

– **Learning perceptual categories**: Very early on and even before they can produce any articulate speech sounds, infants are able to categorize appropriately the sounds of their environment language (Vihman, 1996). They learn the perceptual code of their parent's vocalizations: English infants learn for example that the [l] in "lead" is a different sound category than the [r] in "read" while Japanese infants learn that these are the same sound category. Chinese learn that [ma] with different pitches correspond to different categories while French infants learn that they are the same. This learning is largely unsupervised and happens well before infants understand the meaning of words. It is still a mystery how they achieve this, and no machine is currently able to perform as well as they do.
– **Learning to control the vocal tract**: The control system of the vocal tract of human adults is highly complex since it enables to move concurrently and in a coordinated manner dozens of muscles and articulators. Learning to perform non-trivial constrictions (obstructions of the air flow which modulate

the sound source produced by the glottis) is thus a motor task as difficult as for example the reaching of objects with the arm. Infants take a while before they can produce the complex movements which are needed to produce the speech code of adults. Understanding this process of motor exploration and learning is an open question and it is a challenge for Embodied Communicating Agents to manage to modulate in complex manners a physical medium such as the high dimensional vocal tract.

- **Learning the mapping between acoustic waves and motor commands**: Once perceptual categories and complex motor control are mastered, the main challenge remains. Indeed, the infant needs to be able to find the motor commands corresponding to a given speech acoustic wave that it perceives, and he needs to be able to associate a motor command with a perceptual category. In brief, he needs to learn the correspondences between the two spaces. This task has several complications: 1) the mapping is not one-to-one, because there are many articulatory configurations which produce the same sound; 2) the mapping needs to be learnt not only for the two spaces of the infant, but also between for example the acoustic speech waves produced by adults and its own motor space. This introduces a high difficulty since no two individuals have exactly the same vocal tract, and in particular the adult vocal tract is very different from the infant's vocal tract. For Embodied Communicating Agents, this case is typically very frequent, since two agents typically have different bodies and possibilities to modulate a physical medium such as acoustic waves. Another complication is, as explained earlier, that the same word is never pronounced exactly the same manner by the same speaker, and is subject to a high level of noise. One phoneme for example can be vocalized in a very different manner depending on the context, because of co-articulation. Yet, the infant is able to consistently categorize streams of acoustic waves into streams of phonemic categories. Embodied Communicating Agents should be able to do that.

9.3.2 The formation of a shared code

Learning an existing speech code is one thing. The creation of a new shared code in a population of agents is another one. The question of how a conventional speech code such as the French vocalization system or the Chinese vocalization system might have formed needs to be answered. And how similar codes can be formed by populations of artificial agents? Both questions can be declined in two versions, depending on the assumptions that we make about the initial level of the social and cognitive capacities of the agents. A first version concerns the case where agents already possess rich modes of social coordination, possibly including already some conventions such as interactional rituals or the capacity to play language games (Steels, 1997b). The question is then, given a set of conventions that they already share, how they can use them to create a new one. In this case, because we assume complex social and cognitive capabilities of the agents, we should also expect that the label system that they form has the complexity, the robustness, and the adaptivity shown by for example the human speech code. This means on the one hand that the label system should be able to use efficiently a high-dimensional continuous non-linear noisy channel such as the vocal tract/acoustic waves/ear medium. This also means that the label system possess all the properties that we listed in the previous section, possess hundreds of perceptual categories, and possess phonotactics. The code should be robust to a continuous flux within and out of the population of agents, and might possibly evolve with time.

A second version of the challenge concerns the case where agents already possess neither means to coordinate socially nor conventions (including those like language games). This is a scientific challenge as much as a technological challenge. Indeed, from a scientific point of view, this version of the question is crucial to the understanding of the origins of language. Indeed, a conventional physical system of labels, as a vehicle of information, is a pre-requisite for linguistic communication. Without such a label system, no linguistic communication is possible. Thus, it is crucial to understand how the bootstrapping of the first label system was made in a population of humans which did not already possess conventions such as interactional frameworks like language games (which pre-suppose already the existence of primitive conventional symbolic systems). From a technological point of view, this is also a very useful challenge since it would allow machines which do not already share means to coordinate socially to build an initial label systems, a building block necessary for their access to mutual communication.

9.3.3 Understanding the role of morpho-perceptual constraints

As a scientific challenge, it is also a necessity to understand which aspects of the human speech code are due to the constraints of the particular physical medium on which it relies, and which aspects are not. In turn, even if in general one should not pre-suppose particular properties of the physical medium for general methods of label systems formation in embodied communicating agents, this could be useful in some cases to take care of the physical design of this medium so that the information transfer is maximally robust, efficient and adaptive.

More particularly, this challenge includes the problem of knowing whether the discreteness and combinatoriality (the phonemic coding) of the human speech code is a consequence of the non-linearities between the articulatory and perceptual spaces. We should also understand what is the link between the statistical preferences of human phonemic repertoires and the morpho-physiological constraints. In the same line, we should understand if phonotactics is more the result of optimizing the error correction in information transmission or the consequence of energetical or motor constraints.

9.4 Existing approaches: scopes and limits

There is now a growing literature on the origins of human speech systems, which try to understand where the organization of the human speech code comes from.

We will now detail a number of representative approaches, each time explaining the scopes and the limits.

9.4.1 The reductionist approach

One of the approaches is "reductionist": it tries to reduce properties of the speech system to properties of some of its parts. In other words, this approach hopes to find a physiological or neural structure whose characteristics are sufficient to deduce the properties of speech.

For example, cognitive innatism (Chomsky and Halle, 1968; Pinker and Bloom, 1990) defends the idea that the brain features a neural device specific to language (the Language Acquisition Device) which knows at birth the properties of speech sounds. This knowledge is supposed to be pre-programmed in the genome. A limit of this approach is that its defenders have remained rather imprecise on what it means for a brain to know innately the properties of language. In other words, this hypothesis is not naturalized. Also, no precise account of the origins of these innate devices has ever been provided.

Other researchers focus on the vocal tract physics as well as on the cochlea electro-mechanics. For example, they claim that the categories that appear in speech systems reflect the non-linearities of the mapping from motor commands to percepts. Phonemes would correspond to articulatory configurations for which small changes lead to small changes in the produced sound. Stevens (1972) defends this idea. There is no doubt that the morpho-perceptual apparatus influences the shape of speech sounds. Yet, this reductionist approach has straightforward weaknesses. For example, it does not explain the large diversity of speech systems in the world's languages (Ladefoged and Maddieson, 1996). Also, there are many experiments which show that the zones of non-linearity in perception of some languages are not compatible with those of some other ones (e.g. Japanese do not make any perceptual difference between the "l" of "lead" and the "r" of "read").

Another example of this type of explanation is that of Studdert-Kennedy and Goldstein (2003) for the origins of discreteness, or "particulate speech" in his terms. Studdert-Kennedy and Goldstein remark that the vocal apparatus is physiologically composed of discrete independent articulators like the jaw, the tongue, the lips, the velum, etc. This implies that there is some discrete re-use in complex utterances due to the independent articulators that move. We completely agree with this remark. Yet, some other aspects of discreteness are not accounted. Indeed, for example, as Studdert-Kennedy and Goldstein (2003) note, once you have chosen to use a given set of articulators, there remains the problems of how the continuous space of possible constrictions or timings between gestures is discretized. Goldstein (2003) proposed a solution to this question that we will review later in the chapter (since it is not reductionist but is a mixture of self-organization and functionalism).

One has to note that this "reductionist" approach proposes answers to the questions concerning the presence of phonemic coding and statistical preferences in the human speech code, but they address neither the diversity of speech sounds nor the fact that they are shared across communities of agents. In fact, they are of limited help for the building of embodied communicating agents, since they do not address the question of the formation of speech codes with these properties (they only address "why" questions). But one has to say that this is certainly not their goal either.

9.4.2 The functionalist approach

The functionalist approach attempts to explain the properties of speech sounds by relating them to their function. Basically, it answers the "why" question by saying "the system has property N because it helps to achieve function F". It answers the "how" question by saying "systems with property N were formed through Darwinian evolution (genetic or cultural) under the pressure to achieve function F". This approach could also be called "adaptationist" : systems with property N were designed for ("ad") their current utility ("apt"). Note that most often the functionalist explanations take into account the constraints due to brain structure, perceptual and vocal systems.

Typically, in the case of the four properties of speech sounds we are interested in, this function is "communication". This means that the sounds of a speech code may be perceptually distinct enough so that they are not confused and communication can take place. The constraints which are involved typically include a cost of production, which evaluates how much energy is to be spent to produce the sounds. So, in this view, speech sounds are a reservoir of forms which is quasi-optimal in terms of perceptual distinctiveness and energy production.

For example, Lindblom (1998) showed that if we search for vowel systems which are a good compromise between perceptual distinctiveness and energy cost of articulation, then we find the most frequent vowel systems in human languages. Lindblom (1998) also showed similar results concerning the re-use of units to form syllables.

Operational scenarios describing how cultural Darwinian evolution formed these systems have also been described. For example, de Boer (2001) and Oudeyer (2001) built computer simulations which showed how cultural evolution might have worked, through processes of imitations among agents. In these simulations, the same mechanism explains both the acquisition of vowels (de Boer, 2001) or syllables (Oudeyer, 2001) and their formation; this mechanism is imitation. As a consequence, these works also propose an answer to the question: "How are vowel systems acquired by speakers?".

One has to note that the models of de Boer (2001) and Oudeyer (2001) do not deal with questions concerning discreteness (which is built in) and systematic re-use. Yet, these models are interesting since they show a process of formation of a convention, i.e. a vowel systems or syllable systems, within a population of agents. This really adds value to the work of Lindblom for example, since it provides a mechanism of (implicit) optimisation which Lindblom (1998) assumed.

Yet, one has also to remark that the imitation game that agents play is quite complex and requires a lot of assumptions about the capabilities of agents. Each of the agents maintains a repertoire of prototypes, which were associations between a motor program and its acoustic image. In a round of the game, one agent, called the speaker, chose an item of its repertoire, and uttered it to another agent, called the hearer. Then the hearer would search in its repertoire for the closest prototype to the speaker's sound, and produce it (he imitates). Then the speaker categorizes the utterance of the hearer and checks if the closest prototype in its repertoire is the one he used to produce its initial sound. Then he tells the hearer whether it was "good" or "bad". Each item in the repertoires has a score which is used to promote items which lead to successful imitations and prune the others. In case of bad imitations, depending on the scores of the prototype used by the hearer, either this prototype is modified so as to better match the sound of the speaker, or a new prototype is created, as close as possible to the sound of the speaker.

From the description of the game, it is clear that to perform this kind of imitation game, a lot of computational/cognitive power is needed. First of all, agents need to be able to play a game, involving successive turn-taking and asymmetric changing roles. Second, they need to be able to voluntarily try to copy the sound production of others, and be able to evaluate this copy. Finally, when they are speakers, they need to recognize that they are being imitated intentionally, and give feed-back/re-inforcement to the hearer about the success or not. The hearer has to be able to understand the feedback, i.e. that from the point of view of the other, he did or did not manage to imitate successfully.

The system developed by de Boer (2001) and extended by Oudeyer (2001) addresses the first version of the questions proposed in the section entitled "The formation of a shared code". Indeed, the level of complexity needed to form speech sound systems in this model is characteristic of a society of agents which has already some complex ways of interacting socially, and has already a system of communication (which allows them for example to know who is the speaker and who is the hearer, and which signal means "good" and which signal means "bad"). The imitation game is itself a system of conventions (the rules of the game!), and agents communicate while playing it. It requires the transfer of information from one agent to another, and so requires that this information be carried by some shared "forms". So it pre-supposes that there is already a shared system of forms. The vowel systems that appear do not really appear "from scratch". This does not mean at all that there is a flaw in de Boer's model, but rather that it deals with the evolution of language rather than with the origins (or, in other terms it deals with the formation of languageS - "les langues" in French - rather than with the formation of language - "le langage" in French). Indeed, de Boer presented interesting results about sound change, provoked by stochasticity and learning by successive generations of agents. But the model does not address the bootstrapping question: how the first shared repertoire of forms appeared, in a society with no communication and language-like interaction patterns? In particular, the question of why agents imitate each other in the context of de Boer's model (this is programmed in) is open.

Another model in the same spirit was proposed by Browman and Goldstein (2000) and Goldstein (2003). This model is very interesting since it is the only one we know, except the work presented in the present chapter, which tries to approach the question of the origins of the discretization of the continuum of gestures (they call this "emergence of discrete gestures").They built a simulation in which two agents could produce two gestures, each parameterised by a constriction parameter taken in a continuous one-dimensional space (this space is typically the space of possible places of constrictions, or the continuous temporal interval between two gestures). Agents interacted following the rules of the "attunement game". In one round of the game, both agents produced their two gestures, using for each of them a parameter taken in the continuum with a certain probability. This probability was uniform for both gestures at the beginning of the simulation: this meant that a whole continuum of parameters was used. Then, agents recovered the parameter of the other agent's first gesture, and compared it to the parameter they used themselves. If this matched, then two things occurred: the probability to use this parameter for the first gesture was increased, and the probability to use the same value for the second gesture is decreased. This simulated the idea that agents are attempting to produce both of their gestures differently (so that they are contrasted and can be differentiated), and the idea that they try to produce each of them similarly to the corresponding one of the other agent (so that a convention is established). At the end of the simulations, agents converged to a state in which they used only one value for each gesture, so the space was discretized, and these pairs of values were the same for the two agents in the same simulation and different in different simulations. Goldstein made simulations using and not using non-linearities of the articulatory to acoustic mapping. Not using it led to the uniform use of all parameters across all simulations, while using it led to the statistical preference of parameters falling in the stable zones of the mapping.

Like the simulations in (de Boer, 2001; Oudeyer, 2001), in this model agents have coordinated interactions: they follow the rules of a game. Indeed, they both need to produce their gestures together in one round of the game. Secondly, as in the "imitation game", a pressure for differentiating sounds is programmed in, as well as a pressure to copy the parameters of the other agent. This means that it is supposed that agents already live in a community in which complex communication exists. Thus, it remains to be seen how discrete speech, which has been argued to be crucial for the rise of language (Studdert-Kennedy and Goldstein, 2003), may have been there without supposing that complex communication has already risen. More precisely, how discrete speech may appear without a pressure to contrast sounds? This is one of the issues that this approach does not solve. Also, in the model of Goldstein, one assumption is that agents directly exchange the targets that they used to produce gestures (there is noise, but they are still given targets). Yet, the vocalizations of humans are continuous trajectories, first in the acoustic space, and then in the organ relation space. So what a human gets from the gesture of another is not the target, but the realization of this target which is a continuous trajectory from the start position to the target. And because targets are sequenced, vocalizations do not stop at targets but continue their "road" towards the next target. The task of recovering the targets from the continuous trajectory is very difficult, and at least has not been solved by human speech engineers. Maybe the human brain is equipped with an innate ability to detect events corresponding to targets in the stream, but this is a strong speculation and so incorporating it in a model is a strong (but yet interesting) assumption.

9.4.3 The "blind snow-flake maker" approach

There is another track of research, which we think has been left nearly unexplored in the field of the origins of language, and speech in particular. This is what we may call the blind snow-flake maker approach (by analogy with the "blind watch-maker" of Dawkins (1986) which illustrated the functionalist approach). This approach is typically adapted to answering the second version of the questions proposed in the section "The formation of a shared code". It is represented by works like (Oudeyer, 2005).

When investigating the origins of a system like the human speech code, two types of answers must be provided (Oudeyer, 2003). The first type is a functional answer: it establishes the function of sound systems, and shows that human sound systems have an organization which makes them efficient for achieving this function. This kind of answer can be proposed using the functional approach that we described earlier. For example Lindblom (1998) showed that statistical regularities of vowel systems could be predicted by searching for the vowel systems with quasi-optimal perceptual distinctiveness. This type of answer is necessary, but not sufficient: it does not explain how evolution (genetic or cultural) may have found these optimal structures, and how a community may choose a particular solution among the many good ones. Works such as the system presented in (de Boer, 2001; Oudeyer, 2001) provide a first element of answer, at a cultural level. But as we have explained, they assume capabilities for the agents which are evolutionary complex and in fact pre-suppose some forms of primitive linguistic capacities, such as the existence of primitive convention systems. The complexity of these assumed capabilities does not allow to understand easily how evolution could have formed them, even if we assume an explicit pressure for communication. In particular, it is possible that "naive" Darwinian search with random variations is not efficient enough for finding complex structures like those of speech: the search space is too big (Ball, 1999). This is why we need yet another answer: we have to account for how natural selection may have found these structures. A possible way to do that is to show how self-organization can constrain the search space and help natural selection. This may be done by showing how a much simpler system can self-organize spontaneously and form the structure we want to explain.

For illustration, it is useful to observe that the structure of the argumentation this approach about the origins of speech is the same as the one of Thomson (1961) about the explanation of hexagonal cells in honey-bees nests (see Figure 1). The cells in the honey-bees nests have a perfect hexagonal shape. How did bees came to build such structures? A first element of answer appears if one remarks that the hexagon is the shape which necessitates the minimum amount

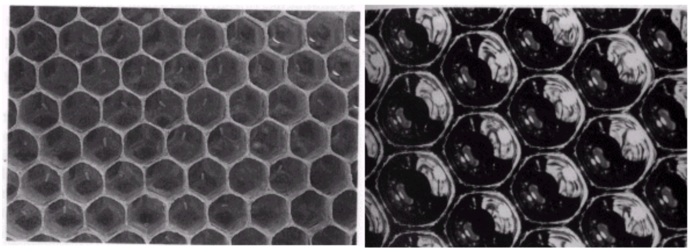

Fig. 1. The cells in the honey-bees nests (figure on the left) have a perfect hexagonal shape. Packed water bubbles take spontaneously this shape under the laws of physics (figure on the right). This lead D'Arcy Thompson to think that these same laws of physics might be of great help in the building of their hexagonal wax cells.

of wax in order to cover a plane with cells of a given surface. So, the hexagon makes the bees spend less metabolic energy, and so they are more efficient for survival and reproduction than if they would build other shapes. One can then propose the classical neo-Darwinian explanation: the bees must have begun by constructing random shapes, then with random mutations and selections, more efficient shapes were progressively found, until one day the perfect hexagon was found. Now, a genome which would lead a bee to build exactly hexagons must be rather complex and is really a needle in a haystack. And it seems that the classical version of the neo-Darwinian mechanism with random mutations is not efficient enough for natural selection to have found such a genome. So the explanation is not sufficient. D'Arcy Thompson completed it. He remarked that when wax cells, with a shape not too twisted, were heated as they actually are by the working bees, then they have approximately the same physical properties as water droplets packed one over the other. And it happens that when droplets are packed, they spontaneously take the shape of hexagons. So, D'Arcy Thompson shows that natural selection did not have to find genomes which pre-program precisely the construction of hexagons, but only genomes who made bees pack cells whose shape should not be too twisted, and then physics would do the rest[1]. He showed how self-organized mechanisms (even if the term did not exist at the time) could constrain the space of shapes and facilitate the action of natural selection.

[1] This does not mean that nowadays honey bees have not a precise innate hard wired neural structure which allows them to build precisely hexagonal shapes, as has been suggested in further studies such as those of von Frisch (1974). The argument of D'Arcy Thompson just says that initially the honey bees might have just relied on the self-organization of heated packed wax cells, which would have lead them to "find" the hexagon, but later on in their evolutionary history, they might have incorporated in their genome schemata for building directly those hexagons, in a process similar to the Baldwin effect (Baldwin, 1896) in which cultural evolution is replaced here by the self-organization of coupled neural maps.

Oudeyer (2005) uses a similar approach to find explanations of the origins of human speech codes. (Oudeyer, 2005) presents a system which shows how a speech code may form in a society of agents which do not already possess means to communicate and coordinate in a language-like manner (as opposed to the agents described in (de Boer, 2001; Kaplan, 2001; Oudeyer, 2001)) and which do not already possess a convention and complex cognitive skills for linguistic processing (as opposed to the agents in (Kirby, 1999b) for example). The agents in this system have in fact no social skills at all. This shows how one crucial prerequisite of language, which is the existence of an organized medium which can carry information in a conventional code shared by a population, may appear without linguistic features being already there.

The self-organized mechanism of this system appears as a necessary complement to the classical neo-Darwinian account of the origins of speech sounds. It is compatible with the classical neo-Darwinian scenario in which the environment favours the replication of individuals capable of speech. In this scenario, the artificial system plays the same role as the laws of the physics of droplets in the explanation of the hexagonal shape of wax cells: it shows how self-organized mechanisms can facilitate the work of natural selection by constraining the shape space. Indeed, the system shows that natural selection did not necessarily have to find genomes which pre-programmed the brain in precise and specific ways so as to be able to create and learn discrete speech systems. The capacity of coordinated social interactions and the behaviour of imitation are also examples of mechanisms which are not necessarily pre-required for the creation of the first discrete speech systems, as this system demonstrates. This draws the contours of a convincing classical neo-Darwinian scenario, by filling the conceptual gaps that made it stay an idea rather than a real working mechanism.

Furthermore, in this system the same mechanism accounts for properties of the speech code like discreteness, compositionality, universal tendencies, sharing and diversity. This account is original because: 1) only one mechanism is used to account for all these properties and 2) we need neither a pressure for efficient communication nor innate neural devices specific to speech (the same neural devices used in this chapter can be used to learn hand-eye coordination for example). In particular, having made simulations both with and without non-linearities in the articulatory/perceptual mapping allows us to say that in principle, whereas the particular phonemes which appear in human languages are under the influence of the properties of this mapping [2], their mere existence, which means the phenomenon of phonemic coding, does not require non-linearities in this mapping but can be due to the sensory-motor coupling dynamics. This contrasts with the existing views that the existence of phonemic coding necessarily need either non-linearities, as defended by Stevens (1972) and Mrayati et al. (1988), or an explicit functional pressure for efficient communication, as defended by Lindblom (1998).

[2] the system can predict the most frequent vowel systems in human languages when using a realistic model of the production and perception of vowels

Oudeyer (2005) also studies an extension of this system allowing to study the formation of rules of syntax for sounds combinations. It shows how similar assumptions can also lead to the formation of primitive shared phonotactics as well as phonological patterns. It also studies theoretically how the addition of constraints such as non-linearities due to the articulatory/acoustic mapping or such as the energetic cost of vocalizations could influence the statistical preferences of populations of agents for certain kinds of phonotactics. This shows that if one wants to be able to predict the actual phonotactics preferences in the human languages, then it is crucial to take into account all the constraints as well as their interactions.

Stages and Computational Challenges

Luc Steels

SONY Computer Science Laboratory Paris

10.1 Introduction

This document focuses on how we could achieve self-organised *human-like* language communication in artificial embodied communicating agents. By self-organised, we will mean that the agents should be capable to autonomously develop and negotiate a communication system, which is adapted to their communicative needs and the environments in which they operate. No human intervention is required, neither for supplying the initial language nor for adapting the language or training the agents. By human-like language, we mean that the communication system has the properties defined in [section by Parisi], in particular that the communication system is established and propagates in a cultural rather than genetic fashion, and that it is representational, i.e. that a rich conceptualisation of the world is being expressed (Talmy, 2000).

Human-like communication rests on many highly advanced prerequisites such as routinised turn-taking, joint attention, speech recognition and production, and grounded world modeling [see section by Hafner and Kaplan]. Here it is assumed that such prerequisites are in place, either by some developmental process or by prior scaffolding of the agents.

10.1.1 Motivation for Stages

From a research point of view, it is obviously extremely useful to identify a set of stages or milestones with respect to human language, because then simpler forms of language could be studied before tackling more difficult ones and we could investigate transitions between stages, similar to the way this is done in biology (Maynard-Smith and Szathmary, 1988). From an engineering point of view, a division into stages is also very useful because we could then already build useful applications without having to tackle the full complexity of human language.

On the other hand, it is not so obvious to identify different stages in human language, and there is definitely no consensus about it. All human languages are known to have a similar level of complexity and expressive power, so different stages are not so easily observable as in biological species. Bickerton (Bickerton, 1992) has proposed that there are two stages: a lexical stage, which he calls proto-language and is similar to the language of 2-year olds or pidgins spoken in trading contexts, and a grammatical stage, which has the complexity of full human language. Jackendoff (Jackendoff, 1999) paints a more complex picture with a dozen different stages. He includes for example a step in which a system of grammatical relationships conveys a set of semantic relations, a stage in which there is the emergence of a system of inflections, and so on. Jackendoff's stages are motivated by fossils that he identifies in existing human languages.

This paper proposes an alternative set of six stages, motivated by research into the self-organisation of communication systems in artificial Embodied Communicating Agents. Each stage is characterised by a particular level of complexity at the language side and a particular level of expressive power at the meaning side. Each stage is assumed to build further on the capacities achieved atearlier stages and requires a major 'breakthrough'. The stages proposed here do not necessarily correspond to historical stages in human language evolution or child language acquisition but are considered to be useful plateaus for structuring the investigation and the engineering of embodied communicating Agents.

Apart from a precise identification of the expressive power and linguistic complexity characteristic for each stage, we need to find answers to the following four questions:

1. What are the cognitive functions and interaction protocols that an agent needs in order to see the emergence of a communication system (including a conceptualisation system) at the level of the group?
2. What kind of computational theory can causally explain these cognitive functions and behaviors, i.e. what information structures and information processes are necessary and sufficient to generate them?
3. What are plausible neurobiological embodiments of this computational theory?[1]
4. How and why might a transition between two stages come about? A transition may either be achieved through genetic evolution introducing new cognitive mechanisms, or by the recruitment from existing cognitive strategies of the best strategy for a particular level of communication.

For experiments and practical applications, (1) and (2) are the most important issues because they must be solved to allow the construction of operational systems. Most progress so far has been on these questions. (3) and (4) are particularly relevant if we want to understand an evolutionary or developmental scenario for how humans could have evolved the capacity for language and how artificial agents might emulate this.

10.1.2 Capacities and Inventories

Before introducing the different stages, it is useful to make a distinction between:

1. The *capacity for language* (also known as the architecture of the language faculty), which is the set of cognitive mechanisms that agents need in order to enter and participate in a language community. A mechanism could for example be a bi-directional associative memory to be used for storing and retrieving words making up a lexicon.

[1] The distinction between (1), (2) and (3) was first introduced by David Marr in the domain of vision (Marr, 1982)

2. The *language inventory of an individual agent* or *ideolect*, which is the body of knowledge (individual lexicon and grammar) that an agent uses to map form to meaning and meaning to form. It contains for example a specific lexicon.
3. The *communal language*, which is the consensus that has arisen in a particular population on how to express meanings. This is obviously emergent from the activities of the individual agents and is not physically stored anywhere.

All of these have undergone change or are undergoing change in the case of humans and human languages. When we speak about research in the evolution or development of human language we must therefore make a distinction between (1) evolution of the capacity for language (ontogenetically and phylogenetically), (2) development and change of an individual's inventory throughout life, and (3) emergence and continuous change of the communal language in a population. Also in the case of artificial agents, we must accept that there is constant change at the level of individual agents as they develop and acquire, invent or adapt their interventories to remain adapted to the tasks and environments in which they have to operate. This then impacts the change in the communal language of the group.

With respect to the language inventory and communal language, we will furthermore make a distinction between the concrete elements routinely used in a language: the lexical items, syntactic and semantic categories, grammatical rules and constructions, and the meta-level structure of the language (further called the meta-grammar) which constrains how the language can be expanded. This meta-grammar captures the sytematicity of the language. For example, a language might use cases like nominative, accusative, dative and case markings (as in German or Latin) for the expression of event-argument structure of verbs, but it might also use grammatical relations (subject, direct object, indirect object) with word order and prepositions (as in English), or post-nominal particles (as in Japanese) for the same purpose. If a certain language has chosen one approach and a new event needs to be expressed, the grammar should be expanded in the same 'style' as is already dominant in the language.

By analogy with ideolect and communal language inventory, we make a distinction between:

1. The *individual meta-grammar* which is the meta-grammar of a single individual.
2. The *communal meta-grammar* which is the meta-grammar underlying a communal language.

Some linguists believe that there is a single meta-grammar with a set of basic principles and parameters that is universally underlying all languages. It is called Universal Grammar (UG) and assumed to be innate (Chomsky, 1995). The individual meta-grammar then consists of a particular parameter setting constraining the general principles of UG in a way compatible with a particular grammar (Niyogi and Berwick, 1995). Others argue that there are no absolute universal properties but rather universal tendencies in languages, and that meta-grammars take shape and evolve in a cultural fashion, just as the grammars themselves, constrained by human cognitive embodiment and the nature of the task. In that case, there is no universal meta-grammar in which parameters are to be set for defining the meta-grammar of a particular language.

In addition to linguistic capacities there are also the conceptual capabilities to produce or interpret the semantic structures expressed by language. These conceptual capabilities also undergo change and co-evolve with the increased complexity of the language. In the case of Embodied Communicating Agents we focus on grounded meanings, i.e. meanings that are anchored in the world through the sensori-motor embodiment of the agent, and so we have to worry about how this grounding will take place and how concepts remain adaptive if the needs of agents or the structure of the environment changes.

Again it is useful to make a distinction between three aspects:

1. The *conceptual capacity* of the agent, which is the set of cognitive mechanisms and behaviors that agents need in order to conceptualise semantic structures, interpret them, or acquire new perceptually grounded categories that can be used as building blocks for conceptualisation. Such mechanisms could include, for example, a neural network that performs categorisation based on a nearest-neighbor comparison against a set of prototypes.
2. The *conceptual repertoire of individual agents* also known as the *agent's ontology*[2]. This is the set of categories and conceptualisation strategies that agents have at their disposal at a particular point in time. It could consist for example of a set of prototypes implemented as the weights in a network.
3. The *conceptual repertoire of the population as a whole*. This is the set of categories that is shared between the different agents in a population and acts therefore as their common ground. Sharing is a prerequisite for successful communication although the sharing does not have to be perfect.

. For human beings, each of these aspects undergoes change as well, and not just during childhood. As new tasks and domains are tackled, new categories or conceptalisations become relevant and others become obsolete. There is a strong co-evolution of conceptualisation and language, in the sense that expansion of the conceptual repertoires pushes the language to lexicalise or grammaticalise new conceptualisations and at the same time the conventionalised expression of a conceptualisation helps to spread and maintain it in the population and helps individuals to align their conceptual repertoires so that they become more similar (Garrod and Anderson, 1987). These human abilities for forming and coordinating new conceptualisations is precisely what we must understand and synthesise in artificial embodied agents, if we want to achieve self-organised communication systems.

The different aspects to be studied are summarised in the following table:

[2] Unfortunately the term ontology as currently used in computer science - and as used here - is different from the same term in philosophy where it only refers to what counts as objects (onto is from the Greek 'to be').

	language	*meaning*
architecture	linguistic capacity	conceptual capacity
repertoire of individual	ideolect	ontology
repertoire of group	language	shared ontology

The remainder of the paper first introduces two explanation structures for explaining how linguistic or conceptual complexity may arise in a population. Then the notion of language games is introduced as a useful framework in which to define and study stages in human-like language communication. Next the proposed stages are discussed in some more detail and pointers are given to the current state of the art.

10.2 Explaining the origins of Linguistic or Conceptual Structure

Broadly speaking, the field of language evolution is exploring two approaches for explaining the origins and evolution of complexity for each of these aspects: a sociobiological and a sociocultural approach. Each approach is applicable for explaining the origins of specific conventions of a language, the linguistic strategies that agents use for establishing conventions (the meta-grammar), the ontologies that are available for conceptualising meanings expressable in the language, and the transitions from one stage to another. Even for the architecture of the language and conceptual faculty, both approaches can in principle be applied. Of course various mixed approaches are available, where some parts are assumed to evolve in a sociobiological way and others in a sociocultural way.

10.2.1 Sociobiological explanations

The first approach has been pioneered by Hurford (Hurford, 1989), who introduced the term sociobiological. In line with other nativist trends in linguistics (Pinker, 1994), a sociobiological explanation relies on genetic coding and natural selection, and is hence closely related to approaches advocated by evolutionary psychologists (Barkow et al., 1992). We paraphrase Hurford (o.c., p.194):

1. Individuals who are more successful communicators enjoy a selective advantage and are more likely to reproduce than individuals who are worse communicators.
2. If an innate strategy X for communication is superior to other conceivable strategies, its possessors tend to enjoy a reproductive advantage over others, thereby increasing the prevalence of this strategy in the next generation.
3. Therefore over an evolutionary timespan, strategy X displaces all rivals, and ends up being *the* strategy by which communication systems are naturally acquired.

The term strategy has to be construed broadly. It can be a strategy for acquiring the lexicon of a language, the strategy of using a particular word for expressing a certain meaning, a strategy for using a certain grammatical construction, a meta-strategy for expanding the grammar of a language in a particular way, etc.

Hurford first applied this line of thinking to investigate the optimal strategy for the most basic stage of language (later called stage I) in which agents name individual objects or situations. He shows (further confirmed by (Smith, 2004)) that a Saussurean strategy in which the associations between names and objects is bi-directional is optimal and hence that this strategy could have become genetically innate by natural selection.

The same line of argument can and has been applied for other stages in language evolution. For example, to explain that a compositional coding becomes dominant over a holistic coding strategy (later called the transition from Stage II to Stage III), there could be competition between agents who use a compositional coding strategy and agents who don't. Assuming that a compositional coding strategy results in better communication systems and a reproductive advantage, the genes coding for such a strategy would then progressively spread in the population. This argument has been made by (Nowak et al., 2000) (although note that the authors equate syntactic with compositional in this paper). Similar arguments have also been made for the linguistic meta-conventions or Universal Grammar (e.g. the available parts of speech or SVO word orders used in languages) which is assumed to strongly bias the structure of specific human natural languages (Briscoe, 2005). There are a few researchers who have used a sociobiological approach for deriving specific lexicons and grammars (Parisi and Cangelosi, 2002), although there is a wide consensus that, in the case of human-like communication, these are not genetically coded.

10.2.2 Sociocultural explanations

The second approach is sociocultural (see e.g. (Mufwene, 2001), (Steels et al., 2002a)). It does not rely on genetic evolution but on two other mechanisms: (1) cultural evolution, and (2) self-organisation. The explanation structure for cultural evolution is as follows:

1. Given a population where individuals have several possible strategies for negotiating and using communication systems.
2. Some strategies might be more effective than others and this impacts the success that agents have in communication, and/or the effort they require.
3. Agents try to optimize communicative success and effort by choosing the best strategy, therefore, over a cultural timespan, the more effective strategy displaces all rivals, and ends up being *the* strategy by which communication systems are established and maintained in that specific population.

The main difference between the sociobiological and sociocultural approach to language evolution is that selection does not go through fitness, reproductive success, and genetic coding, but through cultural choice and direct feedback on success or effort in communication.

Again, we can apply this explanation structure at all levels: for specific lexicons and grammars, for ontologies, as well as meta-grammars. It can also be applied to the architecture of the linguistic or conceptual faculties. As defined earlier, this architecture is the subset of neural mechanisms that are actually used by speakers and hearers of a language. In a sociobiological explanation, this subset is assumed to be unique and specialised for language (as in Chomsky's Innate Language Acquisition Device). In a sociocultural explanation, individuals come to the language communication task with a large set of physiological and neurally embodied cognitive mechanisms, but these mechanisms are not specific to language. They include the ability to recognise or reproduce hierarchical structures, the ability to store, retrieve, and learn bi-directional associations, the ability to compute analogies between structures, etc. Which ones of these abilities is part of the language faculty is determined by the kind of language structures the group has decided to settle on. It is perfectly possible that one group may make other choices compared to another one. To take an easily observable example: Some languages (like English) only use sounds produced by exhaling, whereas others (like the Masai language in Kenya) use not only exhaled but also inhaled sounds, which requires rather different control skills and is very difficult if you are not used to it. It is not that English speakers are in principle uncapable of producing sounds while inhaling (so they have the required physiology and neural machinery for control) rather they do not exploit it for their language. Similarly, many australian aboriginal languages (see (Dixon, 1979)) do not use word order for syntactic purposes relying instead on an elaborate system of morphological markers and agreement, which implies that the cognitive mechanisms required for parsing or production pay less attention to sequencing compared to those required for parsing English, which relies almost exclusively on word order.

But according to the sociocultural approach, cultural selection is only part of the story because there are many cases in which there are still several possible choices left, so agents have a true choice, just like a population has a choice to drive on the left side or on the right side of the road. There is no point in looking for selection criteria as both solutions would work equally well. Thus, it is a pure cultural choice whether a pen will be called "pen" or "ori" (Japanese for pen), whether word order is heavily used by the grammar (as in English) or hardly (as in Australian Aboriginal languages), whether the case system of the language will use a nominative-accusative distinction (as in German) or an ergative-absolutive one (as in Basque), etc. And even if there are selection criteria at stake, languages and their ontologies do not have a unique optimal solution but are trying to solve multiple constraints which are to some extent in conflict. For example, a holistic coding is in some sense more efficient because a complex meaning is conveyed with a single word, however holistic coding requires a bigger lexicon. So there is a trade-off between introducing the kind of more complex processing that is required for a compositional coding versus using a larger memory. So languages, must make choices which criteria to optimise and these choices could (and have) evolved over time.

Even though there are multiple choices possible on how to set up and maintain an effective communication system, it is nevertheless crucial that everybody makes the same choice. So a mechanism is needed to explain how a distributed population can arrive at a consensus without a central coordinator, telepathy, or innate specification. In the sociobiological approach, consensus is reached because all agents share the same genes. In a sociocultural approach, consensus must emerge by a collective decision process. One of the key results from recent research in the sociocultural evolution of language is that this emergence can be explained with a mechanism familiar from complex systems science, namely self-organisation (in the sense of Prigogine (Prigogine and Stengers, 1984)), i.e. a positive feedback loop is established by coupling use to success so that successful elements get used more and thereby have more success (see (Steels, 1996b) for one of the first examples). This kind of dynamics is closely related to the opinion dynamics currently studied in economics (Weisbuch et al., 1994) and its thorough investigation is a natural target for complex systems methods.

10.2.3 Interactions

There are obvious parallels and interactions between sociobiological and sociocultural approaches. The selection criteria that are used in a sociobiological explanation are relevant to a sociocultural explanation as well, and vice-versa. It is only the way these criteria are assumed to impact the evolving communication system that is different. Thus, if a particular style of grammar (say one exploiting word order instead of morphological markers) is experienced as more effective by the population, sociobiologists would argue that this style becomes innate due to the reproductive success agents enjoy, whereas socioculturalists would argue that it becomes adopted based on a cultural consensus.

Note that in a sociocultural explanation, transitions from one stage of complexity to another have to work without any central agency computing global properties. The agents have only local information based on their own one-to-one interactions, and strategy switching has to happen based on such information. Moreover the different strategies must be able to exist alongside each other and the winning strategy has to be 'evolutionary stable' in the sense that if new agents enter into the population with other strategies they will not be able to turn back the clock but progressively adopt the strategy already present in the group. At the same time, strategies may keep shifting in the population under influence of population change, errors in parsing, production or cognition (Steels and Kaplan, 1998), or the desire of the agents to explore other parts of the space of possible languages in order to optimise certain aspects of their language.

Another form of interaction between sociobiological and sociocultural approaches could arise when there are certain strategies which evolve in a sociocultural fashion but then become genetically engrained due to the Baldwin effect (Munroe and Cangelosi, 2002).

In the ECAgents project, we are emphasising the sociocultural approach because this is the most relevant for building artificial Embodied Communicating Agents with practical utility, but the sociobiological approach is also being explored. Simply stated, the sociobiological approach has the disadvantage of requiring many generations of agents before successful shared communication systems are established, whereas for concrete applications it is much more desirable that a particular population (even if it is changing) is able to self-organise and maintain a communication system in rapid, cultural time, and that this communication system can adapted quickly to the environments of the agents (see (Steels and Belpaeme, 2005) for a more extensive argumentation). This does not imply any claim about the role of sociobiological evolution in the case of human language. Moreover the sociobiological approach will also be explored.

10.3 The Language Game Framework

Research in other domains of complex systems science has shown the utility of defining a basic minimal model in which most of the dynamics that we need to study appear. For example, the prisoner's dilemma and its iterated variants have proven to be highly valuable in studying many issues in the evolution of cooperation. We need to do something similar for the study of language evolution. Earlier research in the synthesis of human-language like communication (reviewed in (Steels, 1997c)) has shown that language games played by a population of agents can play this role.

A language game is a situated interaction between two agents playing the role of speaker and hearer. It is a cooperative game with no winner or loser. A language game takes place in a particular environmental setting that acts as context. The participating agents are randomly drawn from the population and they take turns being speaker and hearer. There can be a change in the population, with new agents coming in and others leaving, but this is optional. It should be possible to see a communication system emerge even in the absence of population change.

There are obviously many different language games that humans play. The game that has proven very fruitful and is particularly interesting for artificial agents is concerned with reference, and is called the *Guessing Game* (Steels, 1996b). The speaker identifies an object or aspect of the world (this could also be an action or a state of affairs), and the hearer has to guess which object the speaker has chosen on the basis of verbal information. When the game fails, speaker and hearer can still try to repair the communication based on non-verbal interaction such as through pointing, and this can then be the basis for expanding or aligning parts of their linguistic and conceptual inventories. Other games that have been studied are the *Description Game*, in which one agent simply describes a scene to another agent, or an *Action Game* in which one agent asks another agent to do something in the world (Steels et al., 2002b). The rest of this paper focuses exclusively on the Guessing Game, which has so far been studied the most. Other language games typically involve some kind of reference so the Guessing Game in in some sense primary.

10.3.1 Components of Guessing Games

The many investigations that have already been carried out over the past decade (beginning with (Steels, 1996c) have shown that the generation and adaptation of a repertoire of categories for conceptualisation can fruitfully be studied through *discrimination games*, in which an agent selects an object as topic and tries to come up with a discriminating description that applies to the object but not for anyone of the other objects in the context, and *interpretation games*, where an agent applies a discriminating description to filter the objects in the context in order to make an educated guess about the topic chosen by the speaker. When the ontology of the agent, i.e. the repertoire of available distinctions, is insufficient, agents use a '*meaning pump*' to generate new distinctions. For example, a new prototype is introduced and committed to memory.

The production and recognition of sentences is studied through *coding games*, in which the speaker attempts to code meaning into a sentence using a language inventory made up of words or grammatical constructions, and *decoding games*, in which the hearer decodes a sentence using his own language inventory in order to reconstruct the intended meaning. When the inventory of available form-meaning pairs is insufficient or uncoordinated, the agents use invention or learning to expand their language inventories or adapt them to that of others.

Guessing games combine these two types of games (see figure 1): The speaker chooses a topic from the set of objects in the context and then plays a discrimination game to find a possible distinctive description that can act as the meaning of the sentence. The speaker then plays a coding game to construct a sentence. The hearer plays a decoding game to reconstruct the intended meaning from the sentence and then an interpretation game to make a guesss about the possible topic. A language game succeeds if the topic guessed by the hearer is equal to the topic chosen by the speaker.

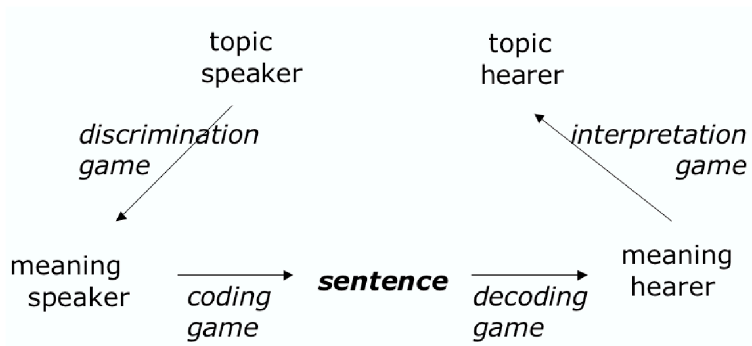

Fig. 1. A guessing game combines discrimination and interpretation games with coding and decoding games

All these steps have been called games (following a Wittgensteinian tradition) because there is clear success and failure and agents try to get better in the game. On the other hand, they are not adversary but cooperative games and so the techniques of game theory may be not so relevant.

10.3.2 Formal Definitions

Let there be a population of agents $\mathcal{A} = \{a_1, ...a_n\}$. The state of an agent a at time t is defined as $a_t = \langle \mathcal{I}_{a,t}, \mathcal{K}_{a,t} \rangle$ where $\mathcal{I}_{a,t}$ is the language inventory known by the agent at time t and $\mathcal{K}_{a,t}$ is the conceptual repertoire (ontology) of the agent. $\mathcal{S}_{a,t}$ is the set of all possible sentences which can be constructed using the agent's language inventory $\mathcal{I}_{a,t}$ and $\mathcal{M}_{a,t}$ is the set of possible meanings that can be constructed from $\mathcal{K}_{a,t}$. Let M be the expressed or parsed meaning and S a sentence that expresses M.

Assume a set of objects in the domain of discourse $\mathcal{O} = \{o_1, ..., o_n\}$ and a context $\mathcal{C} \subseteq \mathcal{O}$. Let the topic t_a be an object or a set of objects chosen by agent a among the objects in the context.

A *discrimination game DG* played by an agent a at time t is defined as a tuple:

$$DG_{a,t} = \langle t_a, \mathcal{C}, M_a \rangle$$

where

- $t_a \in \mathcal{C}$ is the topic,
- $\mathcal{C}$ is the context,
- $M_a \in \mathcal{M}_{a,t+1}$ is a meaning M which is distinctive for t_a, or $\emptyset$ if no distinctive meaning can be found. A meaning M is distinctive for a topic t_a in a context $\mathcal{C}$ iff M, possibly instantiated with a unique set of variable bindings, is true for t_a and for no other object in C.

The ontology of the agent, and hence the set of meanings, may change as a side effect of the game: $\mathcal{K}_{a,t}$ is the ontology available at the beginning of the game, and $\mathcal{K}_{a,t+1}$ the ontology at time $t + 1$.

The score of a discrimination game $score(DG_{a,t})$ is 1 (success) iff $M_a \neq \emptyset$, otherwise it is 0.

An *interpretation game IG* played by an agent a at time t is defined as a tuple:

$$IG_{a,t} = \langle M_a, \mathcal{C}, t_a \rangle$$

where

- $M_a \in \mathcal{M}_{a,t}$ is a possible meaning,
- $\mathcal{C}$ is a context,
- $t_a \in \mathcal{C}$ is the unique topic t_a wuch that M_a is distinctive for this topic in the context $\mathcal{C}$. If there is no such topic the result of the game is $\emptyset$, the empty object.

The ontology of the agent, and hence the set of meanings, may change as a side effect of the game: $\mathcal{K}_{a,t}$ is the ontology available at the beginning of the game, and $\mathcal{K}_{a,t+1}$ the ontology at time $t + 1$.

The score of an interpretation game $score(IG_{a,t})$ is 1 (success) iff $t_a \neq \emptyset$, otherwise it is 0. Note that the game is also a failure if there is more than one possible topic for the given context.

A *Coding Game COG* played by an agent a at time t is defined as:

$COG_{a,t} = \langle M_a, \sigma \rangle$

where

- $M_a \in \mathcal{M}_{a,t}$ is a possible meaning and
- $\sigma \in \mathcal{S}_{a,t+1}$ is a possible sentence in the case of success, or $\emptyset$ (the empty sentence) in the case of failure.

The language inventory of the agent may change as a side effect of the game: $\mathcal{I}_{a,t}$ is the language inventory available to the agent at the start of the game, and $\mathcal{I}_{a,t+1}$ the language inventory of a at time $t + 1$.

The score obtained in a coding game $code(COG_{a,t})$ is 1 (success), iff $\sigma \neq \emptyset$, otherwise it is 0.

A *Decoding Game DEG* played by an agent a at time t is defined as:

$DEG_{a,t} = \langle \sigma, M_a \rangle$

where

- $\sigma \in \mathcal{S}_{a,t+1}$ is a sentence constructable from the agent's language inventory,
- $M_a \in \mathcal{M}_{a,t}$ is a possible meaning in the case of success, or $\emptyset$ (the empty description) in the case of failure.

The language inventory of the agent may change as a side effect of the game: $\mathcal{I}_{a,t}$ is the language inventory available to the agent at the start of the game, and $\mathcal{I}_{a,t+1}$ the language inventory of a at time $t + 1$.

The score obtained in a decoding game $score(DEG_{a,t})$ is 1 (success), iff $o \neq \emptyset$, otherwise it is 0.

A *Guessing Game* (GG) is a combination of all games defined in this section. Assume two agents randomly chosen from the population: a speaker $s \in \mathcal{A}$ and a hearer $h \in \mathcal{A}$ where $s \neq h$. Then a Guessing Game GG played at time t between s and h is defined as

$GG_{s,h,t} = \langle \mathcal{C}, t_s, \sigma, t_h \rangle,$

where

- $\mathcal{C} \subseteq \mathcal{O}$ is a context,
- $t_s \in \mathcal{C}$ is a topic chosen by the speaker,
- σ is a sentence,
- $t_h \in \mathcal{C}$ is a topic guessed by the hearer.

A *Guessing Game* $GG_{s,h,t}$ decomposes into the following games as defined in this section:

$DG_{s,t} = \langle t_s, \mathcal{C}, M_s \rangle$, $COG_{s,t} = \langle M_s, \sigma \rangle$
$DEG_{h,t} = \langle \sigma, M_h \rangle$, $IG_{h,t} = \langle M_h, \mathcal{C}, t_h \rangle$

The score of a Guessing Game $score(GG)$ is 1 (success) if $t_s = t_h$, otherwise it is 0. Note that the meanings used by the speaker and the hearer, do not necessarily have to be equal in order to be successful in the game. At the end of the game, the ontologies and inventories of both participating agents may be further updated based on the outcome of the guessing game.

The minimal goal of the agents is to maximise their cumulative score in consecutive guessing games but there could be additional goals to optimise various aspects of this system, for example, the size of a sentence, the size of the inventory, the amount of polysemy or synonymy, etc.

Initially the agents have no inventory at all: $\forall a, \mathcal{I}_a = \emptyset$ and inventories must be expanded as a side effect of the coding and decoding game to maximise communicative success. Moreover the agents have no ontology either: $\forall a, \mathcal{K}_a = \emptyset$ and they must develop their ontologies as a side effect of discrimination and interpretation games. Of course it is already useful to do experiments where the ontology is fixed and given, but then the issue of the co-evolution between language and meaning cannot be addressed.

10.3.3 A Remark on Objects

The present formalisation makes a short-cut with respect to the objects in the domain. It would be more accurate to make a distinction between the set of objects in reality $\mathcal{O}_r$ and the set of objects as perceived by the speaker $\mathcal{O}_s$ or the hearer $\mathcal{O}_h$. It is well known in real world robotics that object or event recognition is extraordinarily difficult and may not yield the same results due to different spatial positions of speaker and hearer, differences in light conditions, different expectations, etc. Moreover feedback on success in the game (whether $t_s = t_h$) is not so obvious either and usually has to go through real world actions. For example, the speaker may name an object in order to obtain it and the hearer then gives the desired object or not. Often feedback is of course much less direct and success or failure may only become apparent much later. These problems are not considered here, but are of extreme importance for embodied agents playing grounded language games. This paper assumes that $\mathcal{O}_s = \mathcal{O}_h$ and that agents are able to signal extra-linguistically that $t_s = t_h$.

10.3.4 Interaction Protocols

To play a guessing game, agents need to engage in a particular turn-taking action. The implementation of these actions on real world embodied agents is far from trivial (see [section by hafner and kaplan]) but here we assume that is developed earlier or programmed directly as explicit scaffolding.

The following steps are undertaken by the agents in the case of a successful Guessing Game. The context is set by the environment. The left column contains actions by the speaker and the right column actions by the hearer. The 'success' and pointing signals are based on non-verbal communication.

	context $\mathcal{C}$	
1. Speaker chooses $t_s \in \mathcal{C}$		
2. Speaker conceptualises t_s as M_s		
3. Speaker codes M_s as σ		
	$\longrightarrow \quad \sigma \quad \longrightarrow$	
		1. Hearer decodes σ as M_h
		2. Hearer interprets M_h as t_h
		3. Hearer points to t_h
	$\leftarrow$ pointing $\leftarrow$	
4. Speaker compares to t_s		
	$\rightarrow$success$\rightarrow$	
5. Speaker updates $\mathcal{K}_{s,t+1}$		4. Hearer updates $\mathcal{K}_{h,t+1}$
6. Speaker updates $\mathcal{I}_{s,t+1}$		5. Hearer updates $\mathcal{I}_{h,t+1}$

In the case of failure, the speaker can point to the topic to give the opportunity to the hearer to reconstruct a distinctive description that might have been the possible meaning of the sentence:

	context $\mathcal{C}$	
1. Speaker chooses $t_s \in \mathcal{C}$		
2. Speaker conceptualises t_s as M_s		
3. Speaker codes M_s as σ		
	$\longrightarrow \quad \sigma \quad \longrightarrow$	
		1. Hearer decodes σ as M_h
		2. Hearer interprets M_h as t_h
		3. Hearer points to t_h
	$\leftarrow$ pointing $\leftarrow$	
4. Speaker compares to t_s		
	$\rightarrow$ failure $\rightarrow$	
5. Speaker points to t_s		
	$\rightarrow$ pointing $\rightarrow$	
6. Speaker updates $\mathcal{K}_{s,t+1}$		4. Hearer conceptualises t_s as M_h (which may expand $\mathcal{K}_{h,t+1}$)
7. Speaker updates $\mathcal{I}_{s,t+1}$		5. Hearer updates $\mathcal{I}_{h,t+1}$

10.4 Overview

We are now ready to briefly discuss the proposed stages. Before discussing each stage, here is a table presenting for each stage the nature of the meaning to be expressed, the type of form used, and the major issue whose origins needs to be explained.

Stage	*Meaning*	*Form*	*Issue*
I	Individuals	(proper) Names	Convergence on convention
II	Single Categories	Single Words	Co-evo language/meaning
III	Multiple Categories	Multiple Words	Compositionality
IV	Multiple Objects + Predicates	Grammar	Origins of grammar
IVa	id.	Syntax	Exploitation of syntax
IVb	id.	Grammar	Intermediary layers
IVc	id.	Recursion	Hierarchical re-use
IVd	id.	Meta-grammar	Two-level evolution
V	Second Order predicates	Grammar	Second order
VI	Meta-level	Grammar	Level formation

10.5 Stage I: Names

The first stage is obviously the simplest one. It associates a name with an individual object. This is still very close to animal communication systems, such as alarm calls of vervet monkeys, which also associate a signal with a particular situation. Here we are of course looking at open-ended culturally negotiated as opposed to genetically evolved systems, where the set of objects and hence the set of names may always be extended. The big issue at this stage is how agents can establish a shared set of linguistic conventions without a central supervisor or without telepathy.

> **Definition:** In Stage I, M consists of an individual object and S a single word naming this object. This corresponds to proper names in natural language, like "John". The language game at this stage is also known as the Naming Game.

To reach this stage, agents must first of all be able to identify individual objects or situations. Although object recognition in itself very difficult, it can be achieved fairly reliably in restricted conditions so that concrete experiments can be made. Next, agents must be able to maintain a two-way associative memory, associating objects with names, so that they can lookup the name corresponding to an object (in coding), or the object corresponding to a name (in decoding). Because different names for the same object may be floating around in the population (synonymy) and different objects for the same name (homonymy), the memory of each agent will have to store several name-object associations each with a particular strength, reflecting the success of a particular association. Agents should then choose the association with the highest strength both for coding and decoding.

There are three types of functions that agents need to build up and maintain these associative memories:

invention. When a new name is needed, the speaker generates a new word from scratch and associates that in his memory with the object he needs to name.

Adoption. When the hearer encounters a new name, he also associates it in his memory with the object pointed at by the speaker.

Alignment. Agents update the strength of their association based on success in the game. When a particular association was successful, its strength is increased and that of competitors decreased. When an association was not successful its strength is decreased. This implements a reinforcement learning approach with lateral inhibition.

These three types of functions reoccur as fundamental primitives for establishing communication systems at every stage.

There is already a well-established literature on the Naming Game ((Steels, 1996b), (Oliphant, 1997)) and many computer simulations have shown beyond doubt that the lateral inhibition dynamics is effective (see for example figure 2 from (Steels and McIntyre, 1999)). Although the cognitive architecture of agents at this stage appears relatively straightforward, it is far from trivial to show in a theoretical way that a shared vocabulary emerges in the population given a particular set of microscopic behaviors. Progress is being made using techniques from complex systems science but theoretical proofs are still forthcoming. The question is similar to multi-agent decision problems in economics (for example the opinion dynamics discussed in (Weisbuch et al., 1994)) or more generally, the emergence of global coordinated state based on local interactions as studied in statistical physics.

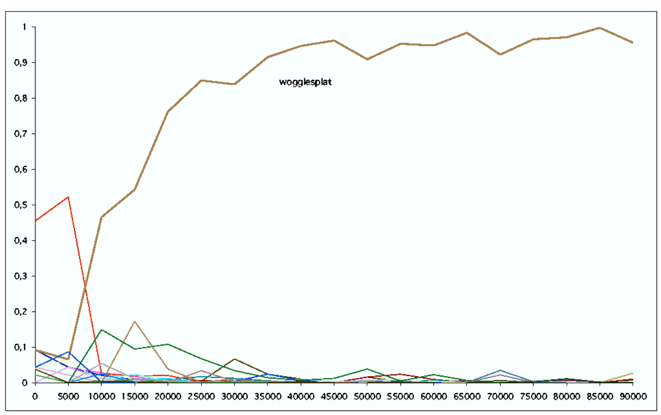

Fig. 2. The graph plots the frequence with which a certain word is used for expressing a particular meaning. There is a winner-take-all effect due to self-organisation.

Many of the computational components listed above can be easily translated into neural networks. For example, agents could use a classical feed forward neural network with a winner-take-all strategy for object recognition (for the discrimination game). New objects could be learned by adapting the weights in such a network, possibly based on supervised learning with feedback signals coming from other cognitive processes. the weights interpretation game is more difficult to implement in a neural network style. It requires a solution to the binding problem (von der Malsburg, 2003), because the network must be able to compare different objects in the context, each having their own features. The system must also test whether the object that has the highest activation based on object identification is the same as the object that was named by the hearer.

The associative memory for the lexicon is also relatively easy to implement in neural networks, for example using a bidirectional associative memory ((Kosko, 1988)) which is itself an extension of the Hopfield network. The strength of associations is then implemented as weights of relations in the network and a winner-take-all process decides which name to use for a given object and which object to use for a given name. Instead of using only Hebbian learning, lateral inhibition has to be added as pioneered in Kohonen networks. It is less obvious how a neural network implementation could invent new names when needed. Moreover the network will have to grow new nodes for new names and new objects, otherwise we have to artificially limit the number of possible names and possible objects. Another question is how the network could handle very large lexicons (at least thousands if not tens of thousands of words).

To build a complete agent in a neural network style requires furthermore that the networks for identification of individual objects are coupled to those that associate names and vice-versa, and that the turn-taking behavior is also implemented in a neural way. Even though such a complete agent has never been built, it is presumably within the state of the art in neural network research to do so.

The next interesting question is under what circumstances a communication system as required for the Naming Game could arise, and particularly how the agents would zoom in on the most efficient way to play such a game. Important work in this area was first done by (Hurford, 1989) who showed that agents endowed with a 'Saussurian strategy', i.e. with a bidirectional mapping between name and object as opposed to separate mappings, gives better results in communication and hence such agents will dominate after a while the population. More recent work along the same line has been carried out by (Smith, 2004), including experiments in which there is an attempt (so far not completely successful) to genetically evolve a population with a Saussurean strategy. Hurford, et.al. assumed a sociobiological framework but similar results could in principle be obtained in a cultural framework in which agents choose among a variety of possible strategies strategies which ones optimises success in communication while minimising effort, but concrete experiments remain to be done. Some speculation how the brain might have evolved to achieve communication systems at this stage are discussed in (Deacon, 1997a).

To summarise, we can say that (1) the required cognitive functions at stage I are well identified, (2) computational theories exist how to achieve them, (3) a plausible neural embodiment of this computational theory appears feasible, although it has not yet been studied very much, and (4) there has been some work on how a particular strategy for playing the naming game can become dominant in a population.

10.6 Stage II. Single Category

In the next stage, the coding/decoding behaviors remain the same but instead of naming individual objects, the agents express categories. At Stage II only single categories and single words are assumed. Some researchers have argued that these categories are innate, i.e. evolved through sociobiological evolution, however they have also been shown to be evolvable through sociocultural evolution as a side effect of playing discrimination and interpretation game. Perceptually grounded categorisations are constrained by embodiment and environment but there is usually still room for choice so that consensus which categorisations to use for language becomes again a cultural issue. This stage raises therefore the important issue of the co-evolution of a categorial repertoire (ontology) and a lexicon: How can the emerging lexicon have an influence on what categories arise in the group and how can the categories help shape the lexicon.

> **Definition:** In Stage II, M consists of a single category that identifies an individual object in a particular context. S names this category. This corresponds to words like "table" or "red", if used in a context where there is only one table or one red object.

To reach this stage, agents must be able to categorise the objects in the context to find a discriminative category, to identify an object in a context based on a distinctive category, and to build up a repertoire of such categories, adequate for playing discrimination and interpretation games. Next, agents must be able to maintain a two-way associative memory similar to Stage I. Instead of associating names to objects, agents now associate words to categories. There must also be an appropriate coupling between the games such that there is not only a progressive alignment of the lexicons of the agents but also of their ontologies.

Computational mechanisms capable to achieve these functions have been demonstrated and studied since the mid-nineties. Many possible ways exist to implement category formation for discrimination and interpretation, ranging from neural network (e.g. using radial basis function networks (Steels and Belpaeme, 2005)) to symbolic approaches (e.g. discrimination trees (Steels, 1996a) or prototypes (Steels and Vogt, 1997a)). It has also been demonstrated that output of discrimination can be fed into coding processes of the same sort as used in the Naming Game (Stage I) and that the same sort of decoding processes as in Stage I are adequate to interact well with interpretation. Moreover by the proper coupling of the two games (which is done by using the result of the total game to align both the linguistic and conceptual inventories of the agents), it has been shown that agents are able to reach a coordinated repertoire of perceptually grounded categories even if these categories were not given innately nor centrally coordinated. The only coordination has taken place through language (see figure 3 taken from (Steels and Belpaeme, 2005)). So implementing complete Guessing Games for Stage II using standard computational techniques is at the moment a well mastered technology and can already be the basis of non-trivial applications.

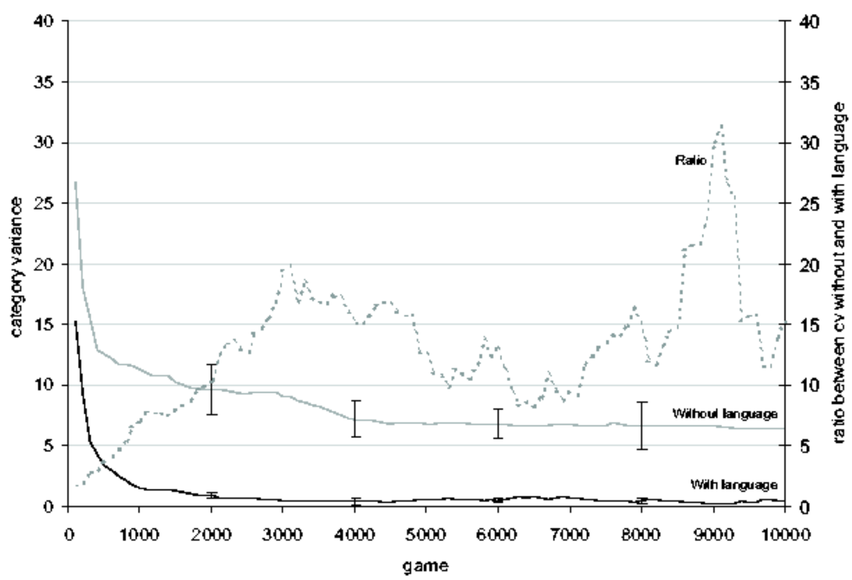

Fig. 3. The graph plots the cumulative category variance between the ontologies in a population of agents with and without language. Agents play discrimination and interpretation games and construct an ontology as a side effect. When there is no coordination through language, the ontologies do not converge, otherwise they do.

Although nobody has already built a complete neural network based agent, this is presumable feasible in principle given the current state of the art in neural network research. As for Stage I, certain parts of the required mechanisms fall easily within the domain of known neural network architectures. For example, categorisation can be done with simple feed-forward networks and the same bi-directional associative memory could be used as for Stage I. On the other hand implementing supporting processes (e.g. for the construction of new names for interpreting a category to pick out an object from the context, etc.) is more difficult and requires an adequate solution to the binding problem. Less work has so far been done on how an optimal set of strategies for playing language games at this stage could evolve or be culturally chosen by agents.

To summarise, we can say that (1) the required cognitive functions at stage II are well identified, (2) computational theories exist how to achieve them, (3) a plausible neural embodiment of this computational theory appears feasible, although it has not yet been studied very much, and (4) little work has been done on the evolution of the language faculty needed to play guessing games at this stage.

10.7 Stage III. Multiple Categories

The next stage is triggered by a need to express multiple categories. Multiple categories arise when the domain has sufficient structure so that categorisation in multiple conceptual spaces is more efficient than categorisation in a single space. For example, rather than having a single conceptual space where colour, size, shape, etc. are the possible dimensions, there are now different spaces so that there are potentially different categories like red/green, small/large, box/ball, etc. The need to express multiple categories introduces the potential for using multiple words, and thus a compositional coding, but that is not necessarily so. It is also possible to use the same strategy as in stage II and use a holistic coding. For example, a single word to express the category combination 'red large box' could be chosen as opposed to three different words. The big issue at this stage is therefore how compositional coding could become the dominating strategy.

> The**Definition:** In Stage III, M consists of multiple categories that conjunctively identify an object, as in "red box". S may consist of multiple words (but still without syntax).

The need for expressing multiple categories has triggered research into the multi-word guessing game (Looveren, 2001) which is in fact only a minor variation of the single-word guessing game at stage II. The cognitive functions and computational mechanisms that are required are well understood. A number of researchers have tried out sociobiological approaches to explain how a compositional coding might become genetically dominant in a group based on a direct coupling of certain communication criteria and reproductive success (Nowak and Krakauer, 1999), (Nowak et al., 2000) (although they unfortunately equate compositional coding with syntax). Some research has also been conducted in how the compositional strategy may become preferred over a holistic strategy in a cultural selection process. There has been further research on the impact of the learning bottleneck on the emergence and cultural transmission of a communication system that uses compositional coding (Kirby, 1999a).

10.8 Stage IV. Multiple Objects.

Stage IV is triggered as soon as multiple objects or relations between objects need to be expressed. It then becomes necessary to introduce some form of syntax or grammar. For example, “red ball under small box”, translates to the following predicate-calculus expression (with question marks before the variables):

red(?x), ball(?y), under(?a,?b), small(?c), box(?d)

Semantic interpretation means to find bindings for the variables involved. A lexical language does not communicate the equalities between the variables, and so the hearer cannot know that “red” and “ball” are about the same object (i.e. that ?x = ?y), that “small” and “box” about another one (i.e. that ?c = ?d), and that the ball is under the box and not the box under the ball (i.e. that ?x = ?a and ?b = ?c. By conveying this additional information through syntax, the speaker not only avoids misinterpretations but also reduces the computational complexity of the semantic interpretation process which is exponential w.r.t. the number of variables.

> **Definition:** M consists of multiple categories and multiple objects which conjunctively identify an object. We now need the power of logic-style predicate-argument structure for M. S necessarily contains syntactic elements to specify constraints beyond individual words.

It is useful to identify four substages which each lead to increased grammatical systematicity.

10.8.1 Substage IVa. Syntax.

The first way to link different lexical items and their meanings is by introducing some form of syntactic structure such as word order.

> **Definition:** In substage IVa, syntactic patterns, such as word order, are used for the linking of individual words, but these patterns are completely ad hoc for expressing constraints on a particular meaning configuration. For example, the combination of ”red (?x), ball(?x)” could be ”red ball” but that of ”blue” and ”ball” could be ”ball blue”.

This stage is comparable to programming languages which use syntactic structure to specify the arguments of procedures and functions but do not have any further systematicity. For example, a procedure

DrawWindow(Window,lefttopX,lefttopY,rightbottomX,rightbottomY,Color)

can be called by giving specific arguments, as in:

DrawWindow(¡window-5146¿,5, 10, 7, 8, ’red’)

But another procedure, say for MoveWindow might be written with its arguments in a quite different ordering:

MoveWindow(lefttopY,Window,lefttopX)

There is no reason to use the same sort of argument structure (for example the window is always the first argument, the X-coordinate is given before the Y-coordinate, X and Y of the same point are always next to each other, etc.) except if the programmer so desires.

The introduction of syntax already is a very significant step beyond the lexical languages studied in stage I to III. In order to achieve this level, agents need considerably more complex cognitive functions: They need to be able to detect the need for the introduction of syntax or be able to interpret a sentence even if no syntactic information is available or it is unknown what this information indicates. They need to be able to recognise and reproduce the syntactic structures that are used in a particular language, such as word order or intonation patterns. They need to be able to invent syntactic rules or adopt these rules from others. The same dynamics as in stage I-III is required to align the different conventions that agents invent, so that the population converges on a shared set of syntactic conventions. At the moment we have already many computational models that are able to handle these capabilities and they rely extensively on the state of the art in computational linguistics. However, implementing these computational mechanisms in a realistic neural embodiment is so far beyond the current state of the art in research. There are various types of networks (such as recurrent neural networks (Elman, 1990)) which are able to learn syntactic patterns and could therefore be used as foundation.

10.8.2 Stage IVb. Grammatical constructions.

At the next stage, a real natural language like grammar emerges. By this, we mean that there is now an intermediary layer of syntactic and semantic categories intervening between the form and the meaning, so that the mapping can take place at a more abstract level and that the resulting language becomes much more regular in structure.

> **Definition:** In substage IVb, a layer of abstract syntactic and semantic categories arises between the meanings M and the sentence S, so that the form-meaning mappings can become more abstract and hence more systemic.

To reach this stage, agents need first of all to build up and maintain a large set of internal linguistic categories, both syntactic and semantic categories. Examples of syntactic categories are noun, verb, nominative, masculine, past perfect. They may be syntactically marked with word order, affixes, intonation, etc. And are used to define or recognise syntactic patterns like SVO. Examples of semantic categories are agent, beneficiary, source, cause-transfer, state, etc. They are used to define or recognise semantic frames which reconceptualise meanings in a way that they fit within a particular language. Grammatical constructions then link semantic frames to syntactic patterns. A typical example of a construction is the Cause-transfer which links a semantic frame involving an agent, patient and target to a syntactic SVOtoO pattern (subject-verb-directobject-to-object,) (Goldberg, 1995) (see figure 4). A specific verb, like slides, is an instantiation of this construction.

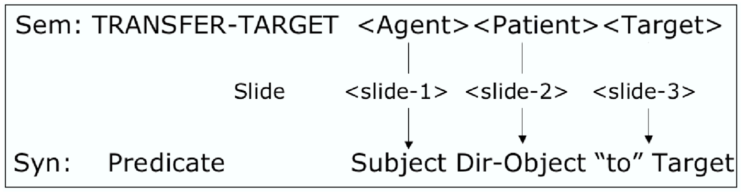

Fig. 4. A construction relates a syntactic pattern such as Subject+Predicate+DirectObject+PrepObject with a semantic frame such as TRANSFER-TO-TARGET+Agent+Patient+Target.

At the moment there is growing research into the computational implementation of construction grammars (ber), (Bryant), (Steels, 2004) based on the state of the art in computational linguistics. Unification-based grammars appear to be the most appropriate formalism but there are many technical questions to make it as flexible as possible, allowing bi-directionality and partial parsing/production. Also the literature on the acquisition and invention of constructions is growing (Chang and Maia, 2001) (Steels, 2004) strongly inspired by research on child language acquisition (Tomasello, 2000) and research in grammaticalisation phenomena (Heine, 1997). The same dynamics as discussed in earlier stages must again be implemented, i.e. the lateral inhibition dynamics so that conventions that are successful increase their spread in the population and those that are not culturally accepted die out.

To implement the necessary computational functions in a neurally realistic way appears at the moment quite beyond the state of the art. For example, it is necessary to implement processes that are capable to match complex structures against each other, to realise partial parsing and production using tens of thousands of cascading rules, to introduce new internal syntactic and semantic categories, etc. At this moment there is still an enormous amount of work left before large-scale computational simulations of the (cultural) evolution of grammar can be set up and this appears a prerequisite before we can turn to neural models. The computational work is also a prerequisite for investigating how there might be a transition from a stage where there is only syntax to one where grammar gets progressively introduced and maintained in the population.

10.8.3 Stage IVc. Recursion.

Natural languages exhibit a further powerful property, namely recursion, and it is useful to postulate a substage in which the grammatical language starts exploiting recursion.

> **Definition:** In substage IVc, the building blocks that have emerged within the intermediary layer may become themselves embedded in units of the same type. For example, a noun phrase may itself contain a noun phrase.

This stage requires that the computational mechanisms used for parsing and production become more complex and also the invention, adoption and alignment processes can handle recursive structures. This appears to require only minor extensions to the formalisms required at stage IVa and IVb but research in this area is still wide open at the moment.

10.8.4 Stage IVd. Meta-grammar.

We have earlier on made a distinction between the ideolect and communal language, i.e. the specific repertoires of words and grammatical constructions used by the individual or the population respectively, and the meta-grammar which enables individuals to expand their own ideolects and hence the communal language. It is clear that this meta-grammar also can undergo change in the individual and the group and so we need to postulate a level where this meta-grammar arises and changes.

> **Definition:** The principles that are used for inventing or acquiring grammar become themselves subject to convention, which requires that agents not only have representations of grammars but also of meta-grammars. Hence the systematicity in the language can be dramatically increased.

This stage requires that there is evolution at two different levels: the level of the grammar and the level of meta-grammar which specifies how the grammar has to be expanded. It has obviously significant implications for the computational mechanisms that need to be available as well as the invention and adoption operators. Although significant work has already been done within the framework of sociobiological evolution based on the notion of principles and parameters (see e.g. (Niyogi and Berwick, 1995)) much work remains to be done, particularly in the context of sociocultural evolution. It is to be expected that from a computational point of view we need a meta-level architecture and computational reflection.

10.9 Stage V. Second order predicates.

Stage V is triggered when the need arises to include predicates that modify other predicates, i.e. second order (or even higher order) predicates. For example the adverb “very” in “very big ball” is modifying the category expressed by the adjective “big”. It is not a predicate that ranges over objects in the domain of discourse directly. The known arsenal of logical operators, quantifiers and connectives now becomes part of the expressive power of possible semantic structures. This higher order usage of predicates clearly needs to be communicated explicitly and is a second reason why grammar becomes crucial and necessary in natural languages.

> **Definition:** M contains categories which operate over other categories. S requires additional grammatical devices to express the higher order use of categories.

Although there is significant amount of research in (logical) semantics and linguistics beyond first order logic (see e.g. research into Montague Grammar (Partee, 2003)) there are so far no convincing computational theories how such second order predicates could arise nor how the grammar could arise to express them. There are also no neural network models nor any scenarios on how specific cognitive strategies could be selected for.

10.10 Stage VI. Meta-level.

The final stage is triggered when it becomes possible that the objects of discourse are themselves elements of the language, for example, when it becomes possible to express "red is the name of a colour". At this point the language becomes its own meta-language and conventions can be explicitly negotiated as opposed to implicitly learned. This can then lead to a very rapid increase in the transfer and coordination of language inventories as well as the conceptual repertoires underlying language.

> **Definition:** The language becomes its own metalanguage: The objects of the domain include elements of the language itself. Additional grammatical structure is introduced to make this possible.

There has already been quite a bit of research on computational and logical meta-level representations and reflection (mostly in the seventies) so from a meaning point of view, there are many ideas on how conceptualisations processes could make use of a meta-level. On the other hand, there has so far been no research, as far as we know, within the context of language evolution research on how grammatical structures could arise that support meta-level expressions. So all the challenges remain open for this stage.

10.11 Conclusions

This paper introduced a number of stages for studying the evolution of language and meaning in artificial Embodied Communicating Agents. Each stage is characterised by a particular level of complexity at the meaning side and a particular level of complexity at the form side. Each stage also requires an increased set of more complex strategies for playing discrimination and interpretation games, including the formation of the ontologies required to play those games, as well as more powerful strategies for communicating these meanings in coding and decoding. The transition between stages is possible based on sociobiological or sociocultural processes or a combination of the two.

The table below lists the state of the art for each stage, with respect to the four questions listed earlier: (1) identification of required cognitive functions, (2) computational theory to achieve these functions, (3) neural implementation, and (4) understanding of the transitions between stages. 1 means that adequate proposals exist. 0 means that no significant results are known at this point, and - means that there have been encouraging ideas and experiments but no firm conclusive results yet.

Stage	Functions	Computational	Neural	Transition
Stage I	1	1	-	-
Stage II	1	1	-	-
Stage III	1	1	-	-
Stage IV				
Stage IVa	1	1	0	0
Stage IVb	1	1	0	0
Stage IVc	0	0	0	0
Stage IVd	0	0	0	0
Stage V	0	0	0	0
Stage VI	0	0	0	0

The lack of neural models, even if there is a known computational theory, is not surprising. Language processing requires a number of basic processes that are beyond the current state of the art in neural network research. We will need efficient working solutions to the binding problem and to the matching of complex symbolic structures before realistic models of language processing can be tackled. Fortunately for building real world applications of embodied communicating agents neural network implementations are not required.

Neurobiological Challenges

Eors Szathmary, Gergo Orban, Mate Lengyel, Zoltan Szatmary, and Peter Ittzes

Collegium Budapest

11.1 Introduction

Natural language is a unique communication and cultural inheritance system. In its practically unlimited hereditary potential it is similar to the genetic and the immune systems. The underlying principle is also similar in that all these systems are generative: they achieve unlimited capacity by the combination of limited primitives.

The origin of natural language is the last of the major evolutionary transitions (Maynard-Smith and Szathmary, 1997). Although later in society important transitions did happen in the way of storing, transmitting and using inherited information, they were not made possible or accompanied by relevant genetic changes in the biology of our species. In contrast, language has a genetic background, but it is an open question how a set of genes affect our language faculty. It is fair to say that with respect to their capacity to deal with the complexity of language, even so-called 'linguistically trained' animals are very far from us.

Language has certain design features, such as symbolic reference, compositionality and recursion, and cultural transmission (Hockett, 1960b). Understanding language origins and change is difficult because it involves three interwoven timescales and processes: individual learning, cultural transmission and biological evolution. These cannot be neatly separated from one another (Christiansen and Kirby, 2003). The fact that a population uses some useful language in a culturally transmitted way changes the fitness landscape of the population genetic processes.

The origin of language is an unsolved problem; some have even called it the 'hardest problem of science'. It is very hard because physiological and genetic experimentation on humans and even apes is very limited. The uniqueness of language prohibits, strictly speaking, application of the comparative method, so infinitely useful in other branches of biology.

This limitation of the approaches calls for other types of investigation. We believe that simulations of various kinds are indispensable elements of a successful research programme. Yet a vast range of computational approaches have also brought about relatively modest success (Elman et al., 1996). This is attributable, we believe, to the utterly artificial nature of many of the systems involved, such as connectionist networks using back-propagation, for example (see Marcus, 1998 for a detailed criticism).

Faculties	Human	Chimpanzee	Monkey
Voluntary control of the voice, face, hands	+	Hands	Hands
Imitation, level 2	+	Only with human training	–
Teaching	+	–	–
Theory of mind recursive/nonrecursive	Recursive	Nonrecursive	–
Capacity to acquire recursive and/or nonrecursive grammar	Both	Nonrecursive*	Nonrecursive

*Fitch-Hauser test needs to be repeated with chimpanzees (*4*).

Fig. 1. Faculties that underlie the evolution of language.

> The only system we know that has apparently solved the 'language problem' is biological evolution. Therefore, our research project is deliberately biomimetic as strongly as it can be, given the limitations prevailing in computability and basic knowledge.

The question then is whether a biomimetic strategy could thus be feasible which would expose simulated agents to selective challenges in such a way that communication should possibly arise as a means to enhance the performance of the agents in synergistic collaborative tasks. The agents must therefore have simulated evolvable nervous systems (a possible aim is a 'toy' cortex, see below), which are under partial (again biomimetic) genetic control (cf. Chapter 3). If this research line turns out to be successful, it leads to the evolutionary emergence of communicating agents, possibly endowed with a faculty to master key components of language such as recursion, symbolicism, compositionality, and cultural transmission. We now give justification for such an approach based on a brief survey of our current understanding of the biology of natural language.

Language needs certain prerequisites (Premack, 2004). There are some obvious prerequisites of language that are not especially relevant to our approach. For example, apes do not have a descended larynx or cortical control of their vocalisations. Undoubtedly, these traits must have evolved in the human linage, but we do not think that they are indispensable for language as such. One could have a functional language without a smaller number of phonemes, and sign language (Senghas et al., 2004) does not need either vocalisation or auditory analysis. Thus, the biomimetic approach shoud be mostly concerned with the neuronal implementation of linguistic operations, irrespective of the modality.

It seems difficult to imagine the origin of language without capacities for teaching (which differs from learning), imitation, and some theory of mind (Premack, 2004). Apes are very limited in all these capacities (Figure 1). It is fair to assume that these traits have undergone significant evolution because they were evolving together with language in the hominine lineage.

We conclude that in any selective scenario, capacities for teaching, imitation and some theory of mind must be rewarded, because an innate capacity for these renders language emergence more likely.

On the neurobiological side we must call attention to the fact that some textbooks (Kandel et al., 2000) still give a simplified image of the neurobiological basis of language. It would be very simple to have the Wernicke and Broca areas of the left hemisphere for semantics and syntax, respectively. But the localisation of language components in the brain is extremely plastic, both between and within individuals (Neville and Bavelier, 1998; Müller et al.). Surprisingly, if a removal of the left hemisphere happens early enough, the patient can nearly completely maintain his/her capacity to acquire language. This is of course in sharp contrast to the idea of anatomical modularity. It also puts severe limitation on the idea that it is only the afferent channels that changed in the evolution of the human brain: modality independence and the enormous brain plasticity in the localisation of language favour the idea that whatever has changed in the brain that has rendered it capable of linguistic processing must be a very widespread pattern of the neuronal networks in the brain (Szathmary, 2001). Components of language get localised somewhere in any particular brain in the most 'convenient' parts available. Language is just a certain activity pattern of the brain that finds its habitat like an amoeba in a medium. The metaphor 'language amoeba' expresses the plasticity of language but it also calls attention to the fact that a large part of the human brain is apparently a potential habitat for it, but no such habitat seems to exist in the ape brains (Szathmary, 2001). The biomimetic approach offers a test of these ideas.

A dogma about the histological uniformity of homologous brain areas in different primate species has also been around for some time. Recent investigations do not support such a claim (DeFelipe et al., 2002). In fact the primary visual cortex shows marked cytoarchitectonic variation (Preuss, 2000), even between chimps and man (Fig. 1). It is therefore not at all excluded that some of the species-specific differences in brain networks are genetically determined, and that some of them are crucial for our language capacity. But, as discussed above, these language-critical features must be a rather widespread network property.

A key feature of brain development is that there is something akin to selection in populations going on (Changeux, 1983): there is a vast overproduction of synapses and neurons, out of which at least half are eliminated under sensory influence (Fig. 2). We suspect that it is practically impossible to obtain a linguistically proficient neuronal network without pruning (Jeffery and Reid, 1997; Johnston, 2001). This is a testable prediction of the biomimetic approach. Pruning does not merely mean that some of the connection weights are set to zero: it also means that they remain zero for the rest of the lifetime of the network.

Genes affect language through the development of the brain. One could thus say that the origin of language is to a large extent an exercise in the linguistically relevant developmental genetics of the human brain (Szathmary, 2001).

The close genetic similarity between humans and chimps strongly suggests that the majority of changes relevant to the human condition are likely to have

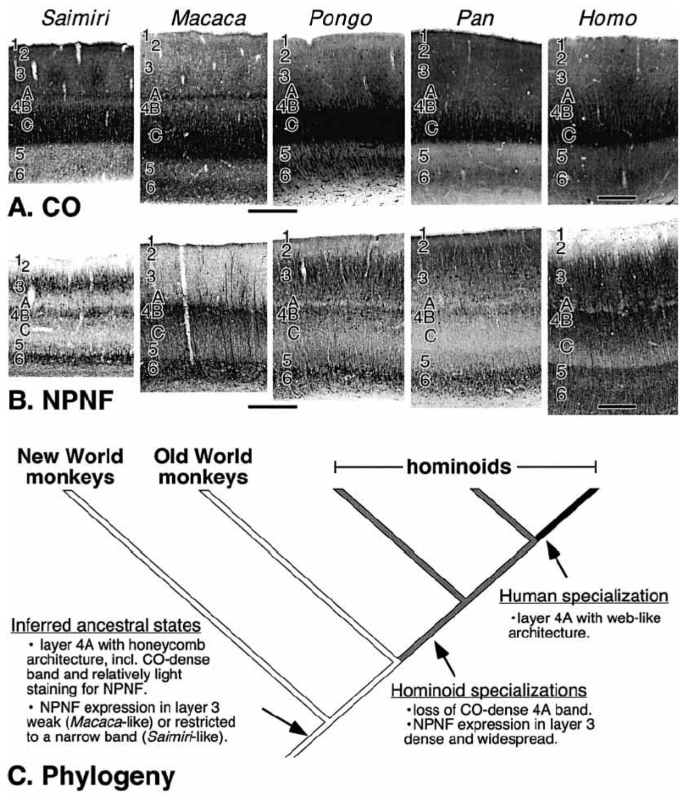

Fig. 2.

resulted from changes in gene regulation rather than from widespread changes of downstream structural genes. Recent genetic and genomic evidence corroborates this view. In contrast to other organs, genes expressed in the human brain seem almost always up-regulated relative to the homologous genes in chimp brains (Caceres et al., 2003). The functional consequences of this consistent pattern await further analysis.

We know something about genetic changes more directly relevant to language. The FOXP2 gene was discovered to have mutated in an English-speaking family (Gopnik, 1990, 1999). It has a pleiotropic effect: it causes orofacial dyspraxia, but it also affects the morphology of language: affected patients must learn or form the past tense of verbs or the plurals of nouns case by case, and even after practice they do so differently from unaffected humans (see Marcus and Fisher, 2003 for review). The gene has been under positive selection (Enard et al., 2002) in the past (Fig. 3), which shows that there are genetically influenced important traits of language other than recursion (Pinker and Jackendoff, 2005), contrary to some opinions (Hauser et al., 2002). We mention that there is

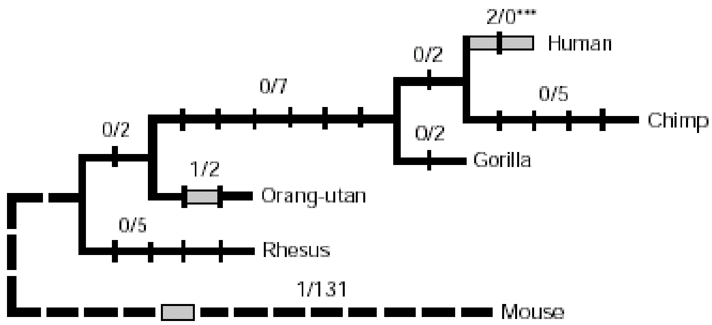

Fig. 3.

a single known human language apparently with no recursion (cited by Pinker and Jackendoff (2005)). It would be good to know how these particular people (speaking the Piraha language in the Amazon) manage recursion in other domains, such as object manipulation. Apes are very bad at recursion both in the theory of mind or 'action grammar' (Greenfield, 1991).

It does seem that the capacity to handle recursion is indeed different from species to species. Although the relevant experiment must be conducted with chimps as well, it has been demonstrated that tamarin monkeys are insensitive to auditory patterns defined by phrase structure grammar, whereas they discover violations of input conforming to finite state grammar (Fitch and Hauser, 2004). Human infants are sensitive to both before they can talk (Fig. 4).

It will be interesting to see what kind of experiment can produce consistent patterns in such a capacity in evolving neuronal networks, and then reverse engineer the proficient networks for this capacity.

We mentioned before that the fact that language changes while the genetic background also changes (which must have been true especially for the initial phases of language evolution), the processes and timescales are interwoven. This opens up the possibility for genetic assimilation (the Baldwin effect). Some changes that each individual must learn at first can become hard-wired in the brain later. Some have endorsed (Pinker and Bloom, 1990), while others have doubted (Deacon, 1997a) the importance of this mechanism in language evolution. Deacon's argument against it was that linguistic structures change so fast that there is no chance for the genetic system to assimilate any grammatical rule. This is likely to be true but not very important. There are linguistic operations, performed by neuronal computations, related to compositionality and recursion that must have appeared sometime in evolution. Whatever the explicit grammatical rules are, such operations must be executed.

Hence a much more likely scenario for the importance of genetic assimilation proposes that many operations must have first been learned, and those individuals whose brain was genetically preconditioned to a better (faster, more accurate)

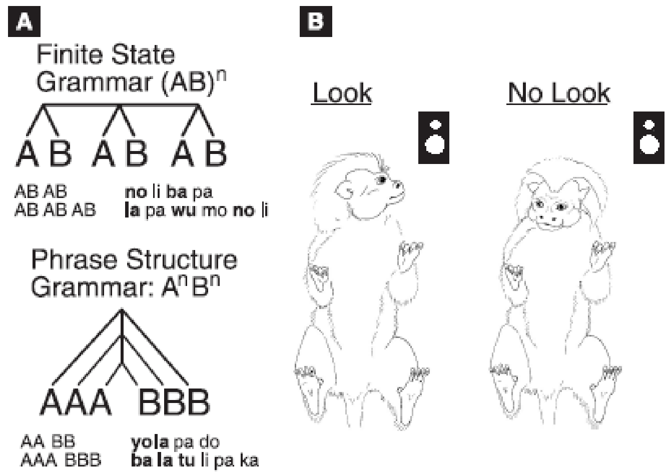

Fig. 4.

performance of these operations had a selective advantage (Szathmary, 2001). Learning was important in rendering the fitness landscape more climbable (Hinton and Nowlan, 1987). This view is consonant with Rapoport's (1999) view of brain evolution. This thesis is also open for experimental test.

An intriguing possible example of gene-culture coevolution has recently been raised by Bufill and Carbonell (2004). They call attention to a number of facts. First, human brain size did not increase in the past 150,000 years, and it did even decrease somewhat in the last 35,000 years. Second, a new allele, ?3 of the gene for apolipoprotein E originated sometime between 220,000 and 150,000 years ago. This allele improves synaptic repair (Teter et al., 2002). The original form, ?4 entails a greater risk of Alzheimer disease and a more rapid, age-related decline in general (Raber et al., 2000). More importantly, ApoE4 impairs hippocampal plasticity and interferes with environmental stimulation of synaptogenesis and memory in transgenic mice (Levi et al., 2003). Interestingly, the ancestral allele decreases fertility in men (Gerdes et al., 1996). The facts taken together indicate, but do not prove, a role in enhanced synaptogenesis in a period when syntactically complex language is thought to have originated. More evidence like this would be welcome in the future, since one such case can at best be suggestive.

In this chapter we focus on three issues in more detail. (1) What kinds of neurobiological platforms can contribute to the realization of ECAgents, embedded in simulated and robotic environments? This is important for information technology as well for basic biology (the evolution of the language faculty). (2) As explained in Chapter 10, concept formation must go hand in hand with language development. Although it seems to be true that some animals lacking language are able to form concepts, this problem has to be addressed in its own right within ECAgents as well. The reason is that if language is useful, then agents must talk about something. Hence during the course of language evolution/emergence semantics can hardly be separated from syntax/grammar, even if such separation is somewhat misleadingly emphasized for the current state of the human language faculty. Furthermore, language use feeds back onto concept formation (what is a unicorn?). Therefore, we must tackle the issue of concept formation as embedded in neuronal systems. (3) The biomimetic approach works only if the appropriate selective environments can be identified. Is it possible to draw suggestive conclusions from comparison of language origin scenarios in biology/linguistics? This final part of our chapter deals with this issue, and emphasizes connection to the stages suggested in the previous Chapter (10).

11.2 Evolvable Neuro-Genetic Systems for Communication

11.2.1 Introduction

It is widely accepted that human language is a product of an evolutionary process. A lot of efforts are devoted to the understanding of this process. Once we understand it, we may exploit that knowledge to create agents that use complex forms of communication.

Understanding an evolutionary process is relatively easy, if fossils of intermediate stages are available. Unfortunately, this is not the case with the human language faculty. What possible ways remain then to investigate the above-mentioned evolution? Simulation is a tempting possibility to replace reality, provided initial conditions and governing rules can be determined appropriately.

What needs to be simulated? Much depends on where we want to start. One possibility is to keep the language faculty fixed throughout the simulation so that all agents are born with full capacity to acquire a language. This approach approximates the cultural evolution of language. Or we might want to allow language faculty itself to evolve and eventually investigate the coevolution of language and language faculty. In fact, this same coevolution has taken place in the case of human language and brain.

With regard to our aim of creating a language faculty somewhat comparable to the human one, we can benefit from the above-mentioned two approaches differently. The first approach provides a method to test the performance of putative language devices under various circumstances. The results of the tests can in turn be used to refine language devices through further design and intuition. On the other hand – via automatic refinement of the language device through evolution – the second approach may save us efforts to redesign.

Evolving language faculties can be done under many theoretical frameworks. We turn to biology for inspiration (Fullmer and Miikkulainen, 1992; Sporns et al., 2004). If we are interested in what makes the difference between language use by humans and apes, we cannot avoid representing the language faculty in terms of biology, such as neuronal systems. Moreover, up to now these are the only known systems that seem capable of handling language proper.

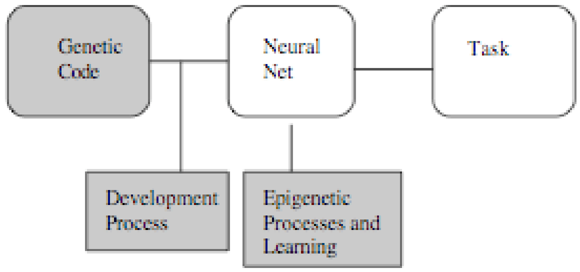

Fig. 5. The main units of a neuro-genetic system. (We don't mention the Genetic Algorithm here.) The shaded boxes are the most sensitive parts because of the required preconceptions.

11.2.2 Background and Questions

These directions (to mimic evolution and the neural system) seem to be correct and promising, but why do the results of these systems only work in simple cases? One problem is the computational background. Currently available computer capacity for parallel computing is really poor compared to the nervous system. Simulating massively parallel systems, such as neural nets on a computer system that is more or less designed on the basis of the Neumann principles is inefficient and the advantage of parallelism is lost. If we want to see agents with neural capabilities solving complex real world problems we have to wait for proper hardware implementation.

The other and even greater problem is that these model systems are not scalable. The development of communication requires increasingly complex computational background; therefore the background genetic and neural systems should inevitably keep pace with the problem that needs to be resolved (Zhang and Mühlenbein, 1993). For example, some previous neuro-genetic models used a direct coding of neurons and connections in the genome. This structure works properly in small neural networks, but if the dimension of the network is growing, the genome size is going to explode; therefore a failure is inherent.

What are the properties of a neuro-genetic system that is scalable? And what are the requirements of the various communication levels? If we think in the above-mentioned biomimetic way, then we have to conclude that the genetic system behind the artificial nervous system has to be simple; however, the presence of complex non-deterministic interactions is unavoidable.

One aim could then be to develop a neuro-genetic system that is supposed to allow the emergence of communicating agents. As mentioned above, this could be a biomimetic way to use genetic algorithms to evolve neuronal nets that are able to generate and decode language, which is required to solve a given task. It is obvious, too, that the structure of the network and the type of the genetic algorithm and coding are crucial for reaching our goal (Fig. 5).

The neural net has to be complex enough to solve a given task at a given complexity. The genetic coding, the development process and the epigenetic processes determine the structure of the neural net. These parts should support the required structure in the neural net in order to solve the given task. We have to be aware that the design phase of a similar complex system is the summarization of our knowledge on developmental and genetic processes. If our knowledge is not sufficient, the computational ability of the produced neural networks will remain stuck under our assumptions. The genetic code and the developmental process in the living systems are based on a fairly complex chemical world. One cannot simulate the whole chemical process under the genetic system; therefore we have to create hypotheses of the subject and create a model of these hypotheses. The greatest challenge is to build the simplest model, which can create a scalable and evolvable network (Stanley and Miikkulainen, 2003).

A possible task is to check the different ways of estimating the complexity of neural networks (Hoffmann, 1993). Numerous research projects are being conducted on this topic and we see that complexity is a very important, but not sufficient term to describe the success of a given network. Complexity measures may be useful in determingin the lowest degree of complication below which a network cannot possibly handle a given task. A burning open question is this: Are we able to find a minimal complexity level of the underlying artificial genetic code that is proven under the failure of the generated neural network, if we know the resolvable task? (Hoffmann, 1997).

This question is really difficult and even if one is able to find an answer, it does not mean that all the required properties of the code are included in the implementation.

In the next subsection we provide a brief enumeration of the supposedly most important properties of neuronal networks, genetic codes and development processes. The assumption that these are the (most) important elements for a working neuronal implementation are open to experimental investigation.

11.2.3 Genotype Properties and the Genetic Algorithm

Genome Structure and Genetic Operators: The structure of the genome is one of the most important issues. If we design a coding system, which is not open ended, evolution will stop at the boundary of the genetic space. One way to create an open-ended coding is to use a marker-based encoding scheme that was inspired by the structure of the real genetic code (Fullmer and Miikkulainen, 1992). The other promising way is to encode information of the network in a tree structure (Zhang and Mühlenbein, 1993). The genetic operators (point mutation, deletion, duplication, recombination) work on this tree structure and mutate the values of the nodes or change the structure of the tree. Duplication ensures the open-ended behaviour of the system, and is able to scale up the search space, if it is required in a new situation. To reduce the negative effect of genetic linkage (Calabretta et al., 2003), the use of recombination is necessary. The optimal combination of these operators is an open question and it depends on the coding system and on the fitness landscape of the given task.

11.2.4 Neuronal Networks (phenotype)

Complex Network Properties: Recent research revealed that large-scale brain networks have interesting complex network properties. Scale-free and small-world properties, functional segregation (modularity) and functional integration might be necessary to achieve complex cognitive functions (Sporns et al., 2004). We want to call attention to a useful distinction between neural and neuronal networks. Whereas the fomer take the nervous system as a loose analogy, the latter prefers to mimic the known biological properties at some model-dependent level. Emphatically, there is conscious exlusion of properties in the latter that are known to crucially contradict facts about patterns and processes in real nervous systems.

Degeneracy:A degenerate system, unlike a fully redundant one, is extremely adaptable to unpredictable changes (Tononi et al., 1998).

Lamination: It was shown that lamination might be advantageous to support more complex information processing in the cortex (Treves, 2003).

Topography:Topographical information in a neural network can play an important role. The interpretation of the structure can be easier. The other even more important reason to use topographical coding of neurons is the modeling of the developmental process. Models, which acknowledge spatial information in biological systems, yield various types of scale-free and small-world network attributes, as the ones that are common in brain structures (Sporns et al., 2004).

Neuron Morphology: It was supposed that more structured neuron morphology is required for more complex cortical functions (Elston et al., 2001). Morphology has meaning only in those cases that are represented in a topographical space. A more complex surface can facilitate more complex connection structure and if different neuron morphologies are encoded in the genome, the search process can change the network structure by changing morphology.

11.2.5 Genotype-Phenotype Mapping and Development

Synapse Development: If the scaling of a network causes problems, week or indirect mapping has to be applied (Zhang and Mühlenbein, 1993). It can be a graph generating algorithm or a partly stochastic connection building process (Balakrishnan and Honavar, 1995). There is even a new sub-discipline of evolutionary computation that engages in the field of genotype-phenotype mapping in artificial systems called artificial embryogeny (AE). (Stanley and Miikkulainen, 2003)

Learning:A learning phase in the development process can change the fitness landscape, and it can be as effective as a Lamarckian strategy at improving search. (Gruau and Whitley, 1993) Because synapse development is a non-deterministic process, learning plays a key role in synapse selection, too.

Synapse Selection, Pruning: One hypothesis for the importance of pruning is that if the synapses at first are overgrown and later pruned, the memory performance of the network is maximized in certain situations. (Chechik et al., 1998) Another possible explanation for the importance of pruning arises from the developmental process. If the synapses join another neuron in a stochastic way, optimal wiring should emerge after learning based pruning. Redundancy is a precondition to this process (Kerszberg et al., 1992).

11.2.6 Summary of Questions

- *What are the properties of a neuro-genetic system that is scalable? And what are the requirements of the various communication levels?*
- *The greatest challenge is to build the simplest model, which can create a scalable and evolvable network.*
- *Are we able to find a minimal complexity level of the underlying artificial genetic code that is proven under the failure of the generated neural network, if we know the resolvable task?*

11.3 Belief networks: a framework for language learning and concept formation

11.3.1 Introduction

How language is acquired, and how concepts about the environment are formed are the focus of two distinct disciplines in cognitive science (see Chapter 10 for a related discussion). The ultimate goal of language learning can be phrased as learning to recognize concepts expressed in utterances, that is 'to peep into the head' of the other speaker, and similarly, to express concepts in utterances, that is 'to drip concepts into the head' of the listener. Concept learning, on the other hand, is the process by which an agent discovers concepts underlying its perceptions that is derived from sensory input. As there are apparent parallels in the cognitive processes involved in language learning and concept learning, and context-dependent language strongly depends on the learning of concepts, we suggest a framework that can capture multiple aspects common in these two types of learning. However, we acknowledge the fact that whereas several animal species seem capable of concept formation, natural language is unique to humans. In the next subsection we provide a mathematically well-founded framework which accommodates both language learning and concept learning in the same computational architecture. Furthermore, as neurobiological plausibility requires that the computations necessary for these cognitive processes are implemented by neuronal networks, we also discuss the possible neuronal architectures capable of performing these tasks.

There exist two markedly different approaches for language production (Seidenberg, 1997). While the innateness hypothesis argues that grammatical structures are innate (Chomsky, 1965; Marcus et al.), the statistical approach emphasizes that language is acquired on a statistical basis by learning transitional probabilities of speech segments (Saffran et al., 1996; Pena et al., 2002). Besides the arguments of incomplete experience and fast learning of language, the success of formal learning theories gives a strong support for the innateness theory. However, among the drawbacks one can enumerate the learned nature of language, the incredible plasticity of linguistic structures and the fact that this theory is at odds with accounting for errors committed during language production (Yang, 2004). Also, there is accumulating evidence that statistical cues are used both at comprehending and producing utterances (Seidenberg et al., 2002), making a statistical approach perhaps more appealing. Statistical learning, too, has been criticized. Most importantly, learning transitional probabilities or joint distributions of high dimensionality would need an inordinate amount of time and data and require unachievable capacities (Yang, 2004).

A probabilistic approach to concept formation is also becoming acknowledged (Gopnik and Schulz, 2004). Causal learning in children has been described by Bayesian combination of likelihoods and prior knowledge (Gopnik et al., 2004). Various computational studies have shown examples of learning concepts based on a limited set of positive examples (Tenenbaum, 1999) and learning concepts using a semi-supervised learning algorithm (Kemp et al., 2004).

A possible approach combines language learning and concept formation in a unified framework by treating both as special cases of unsupervised learning of concepts from different sources of input. The input consists of utterances in the case of language learning, and high level perceptual primitives in the case of concept formation. This approach is encouraged by recognizing that in both cases the goal is to extract concepts from data, and thus the same mathematical tools may be applicable to modelling these processes.

An open question is whether the neural architecture that can perform computations necessary for learning concepts and linguistic structures using unsupervised learning paradigm is possible.

A Bayesian description of language learning accommodates the naive statistical learning and the innatist approaches as extremes of a continuum by using a specific and learnable parameterization of the joint distribution. Agents using perceptual learning rely solely on sensory information during learning, teaching signals, as required in a supervised learning paradigm, are not available. Unsupervised learning aims at discovering aspects of the statistical regularities of the data. This structure is provided by the laws of nature (physics, biology, etc.) and logic in the case of concept learning, and the rules of syntax in the case of language learning. Therefore, concept learning is successful if it develops its own internal model, which reflects how concepts (for instance the objects of the environment) cause perceptions, and language learning is successful if the model developed reflects how concepts lead to utterances.

11.3.2 Bayesian computation: An introduction

The final goal of an unsupervised learning process is to characterize the joint distribution of observed variables. As discussed above, describing a whole joint distribution of many variables relying solely on the occurrence frequencies of different values of the observed variables is practically unfeasible. In order to parameterize the joint distribution a number of constraints are introduced that describe how the variables of the system interact (or rather do not interact), i.e. how the occurrence of one value of a given variable affects the occurrence of the other variable. The mathematical mean of this dependency is given in the conditional probability. These conditional dependencies define a directed graph, called a Bayes network, of the joint distribution. For example, the colour, softness, and taste of a peach depend on its ripeness, and these dependencies are probabilistic due to natural variations among peaches. Thus knowing the ripeness of a peach allows one to make very strong prediction about its colour, softness, and taste. Moreover, if one wants to learn the distribution of peach properties, one may not need to represent a four-dimensional joint distribution, only a one-dimensional distribution of peach ripeness, and three two-dimensional conditional distributions that describe the dependency of colour, softness and taste on ripeness. However, unlike softness and colour, ripeness is not a directly observable quantity. In a graphical model there are both observed variables that take well defined values during observation, and hidden variables, whose values are not directly observable. This way, after an observation event the observer is faced with the challenge to estimate the occurrence of a given value of the hidden variable knowing only the conditional probability of the observed variable given the value of the hidden variable(s). The best estimate of this value is given by the Bayes rule:

$$\mathrm{P(Hidden \mid Observed) = P(Observed \mid Hidden)\ P(Hidden)\ /\ P(Observed)}.$$

As hidden variables by definition are not subject to observation, one has to make inference over the hidden variables in order to calculate posterior distributions.

Learning in a graphical model aims at estimating the parameters using training experience. Given the probabilistic nature of inference, the relationship between the likelihood of training experience given a set of model parameters and the posterior probability of model parameters is not straightforward. According to Bayes rule, the posterior probability of model parameters (P(Par|Stim)) can be calculated by combining the likelihood (P(Stim|Par)) with our prior expectations (P(Par)):

$$\mathrm{P(Par|Stim) = P(Stim|Par)\ P(Par)\ /\ P(Stim)}.$$

Bayes rule not only tells us how to calculate the posterior, but gives us the optimal way to calculate it, in other words, provides the formula to extract the most information that is available using the input and prior expectations. After making inference over hidden variables, the second step in learning the parameterized distribution is making inference over model parameters. Finally, if the graphical model itself is not known a further inference is needed that is performed over the graphical models.

11.3.3 Application of belief networks to language learning and concept formation

The formalism provides a transparent and intuitive representation of the not directly observed concepts as hidden variables of a graphical model, and a sound representation of the relationships between concepts and observed environmental variables or utterances, in the case of language. There are mathematically well founded recipes to learn distributions necessary for recognizing concepts and grammatical regularities. Additionally, the unsupervised learning paradigm does not necessitate the presence of a teacher, which would be unrealistic in most of the real world learning situations.

During learning of concepts and linguistic structures, the learning system has to cope with challenges of various origins. On one hand the system has to learn from a finite sample of data (e.g. a given linguistic structure in different contexts). On the other hand, both learning takes place in a noisy environment: statistical variations might arise in the input data, inherent stochasticity might be present, fidelity of transmission in communication channels might be low, communication partners might communicate ungrammatical sentences; and inherent ambiguity is present in the system (multiple relevant causes can explain the input). As a result, an agent has to be aware of the ambiguity in the input data in order to be successful in acquiring concepts and extracting syntactic rules. The best way to do so is representing this ambiguity in the model of the external world, in other words, the agent not only has to store the mean of model parameters, but also a distribution of model parameters weighted by their corresponding probability values. This claim is directly provided by the framework, as belief networks give a parameterization of the probability distribution. In this framework, during perception the model parameters have to be estimated relying on the distribution of training experience. A particularly effective way of extracting abstract rules from the input can be achieved if, besides representing observed variables, characterizing particular features of the input, hidden variables, possible causes, are distinguished. This way, one particular model of the input data is parameterized by the dependencies of hidden and observed variables.

Importantly, a system dealing with these issues has two critical roles: it has to function, both as a recognition model (i.e. recognizing concepts from perceptions and utterances), and as a generative model (predicting perceptions from available information and generating utterances from concepts). This feature is inherent to graphical models, as Bayes rule provides a way to operate in both ways. When learning concepts, various contingencies of possibly different modalities of sensory input have to be tackled; therefore the graphical model itself is subject to change, depending on the available information. Therefore inference over model structures (Belief networks) is also a key component of the learning mechanism.

Belief networks have already been applied to various biological learning phenomena, classical conditioning in animal experiments (Courville et al., 2004, 2005), causal learning in human infants (Gopnik and Schulz, 2004; Gopnik et al., 2004), and concept formation in humans (Tenenbaum, 1999; Kemp et al., 2004).

Also, statistical approaches have been applied to language acquisition problems (see above).

> The main question that remains open is the class of graphical models that are appropriate for evaluating the joint distribution of relevant variables in concept and language learning.
>
> A fundamental challenge is to find the optimal balance between openness (so that many different specific model structures can be accommodated within the chosen class of models) and learnability (so that model structure can be efficiently inferred from limited data). This amounts to putting the hoary nature-nurture dilemma in this case into a concrete, well-defined, quantitative framework.
>
> Note, that the previous question may be studied for the two problems (language learning and concept formation) separately. However, another open issue concerns the interaction of the two learning systems: how the two systems could use the same set of hidden variables (the concepts), what is the structure of dependencies between hidden variables that is appropriate for both systems, and how learning in one system can potentially bootstrap learning in the other system.

A hint to this question is given by looking at psychophysical experiments that describe what basic chunks of the visual scene are used in infants and adults in order to characterize the statistical regularities of the environment (Fiser and Aslin, 2002). In a series of studies these researchers explored how more complex regularities and how irrelevant environmental variables affect the effective performance of human subjects.

11.3.4 Neurobiological implementation

While a statistical approach characterizes the computations necessary for concept and language learning, the question of how neuronal networks implement these computations remains open. Basic operations that a feasible implementation has to be capable of are inference over latents, model parameters, and models. As these computations might take place on different time scales, it is plausible to assign them to different levels of neuronal computations. While inference over latent variables might be corresponded to activity dynamics of the neuronal network, inference over model parameters and models can be corresponded to synaptic dynamics, i.e. plasticity of synapses. Practically, evolutionary processes form a third level of dynamics that might be beneficial in tuning the initial priors of parameters.

Candidate circuits in the brain to perform calculations necessary for statistical learning in graphical models involve the prefrontal cortex and inferotemporal cortex (Miller et al., 2002, 2003). Combined psychophysical experiments and unit recordings in primates have revealed neurons in these brain areas that can adapt to response selectively for relevant environmental variables (Rainer et al., 1998). Neurons in the prefrontal cortex were shown to respond selectively to abstract rules (White and Wise, 1999; Wallis et al., 2001), and in a task-selective manner (Asaad et al., 2000).

Although neural implementations of Bayesian computations have only recently been started to develop, a number of promising attempts are available. In firing rate-based models Pouget et al. (1998) have shown how to represent environmental variables in a near optimal way, that later was extended to be able to perform multi-sensory integration (Deneve et al., 2001) and to arbitrary smooth transformations of sensory variables (Latham et al., 2003). In their study, Sahani and Dayan (2003) presented a model for parallel representation of stochasticity and multiplicity in a neural network model. Rao (2004) has recently shown an implementation of Bayesian computation in recurrent neural network. A number of studies have been published on the implementation of Bayesian computations in spiking neurons (Hinton and Brown, 2000; Deneve, 2004; Lengyel and Dayan, 2004; Zemel et al., 2004). Although these models are appropriate for specific purposes, a number of them represent one latent variable and multiple observed variables with independent noise on each observed variables or use hidden Markov models for computations. Performing inference over latents is one component of the computations, but the remaining two types of inference, one over model parameters, and the other on models, are also a must for neural networks in order to account for learning in the belief network framework.

The challenge of the research is finding the neural network architecture that is capable of accomplishing the computations in a belief network structure in order to optimally extract information from the available training experience.

11.3.5 Concluding remarks

If the learning systems are not doomed to only passively observe variables but are also allowed to actively interact with the environment, a whole new set of questions need to be investigated. From a theoretical point of view it could be asked what the optimal interaction is that reduces uncertainty maximally under some loss function? Predictions from such theories then could be tested experimentally to see whether subjects actually perform optimal interventions. For example, these interventions could take the form of object manipulations for concept formations, or utterances for language learning.

A crucial question links the belief network approach discussed in this section, with the evolutionary neurogenetic approach presented in the previous section. Is it possibe to select for neuronal networks that, in final analysis, perform the calculations phenomelogically decsribed by the graphical models?

Finally, adopting a game theoretic approach in which multiple agents interact with each other and the environment based on the above principles could be used to investigate the emergence of conventionalized communication systems (Steels, 1998a, chapter 10).

11.4 Designing selective scenarios of language evolution

The origin of human language is one of the hardest problems of science. Speech and gestures do not fossilize, and even hominid fossils only give an indirect evidence of linguistic capacity. It seems that the only possible approach to get a coherent picture is to try to re-enact the evolution of language with the help of agent-based computer simulations. One crucial ingredient of these simulations is the selective scenario.In this chapter, we focus on the identification of a selective scenario that might lead to human-like forms of communication. For a similar attempt to related to more simple forms of communication, see Chapter 2.

Under selective scenario we mean a set of relevant tasks that affect the fitness of agents. Obviously not every task would enable selection with respect to the use of language-like communication. There are a number of tasks that can either be solved without communication at all, or by means of simple signals (e.g. the overwhelming majority of animal communication is done by so called ,,self-reporting" signals, which means signals that transfer information about the state of the signaller – Maynard Smith and Harper, 1995); or language-like communication (further abbreviated as LLCS*[1]) might be useful, but the task itself is too complex to expect LLCS to evolve in that context. All in all, the success of such a computational approach very much depends on picking the appropriate task; that is, on constructing the appropriate selective scenario.

There are at least two possible approaches to come up with a selective scenario fostering some degree of linguistic competence: (i) construction of a selection scheme as engineers would do it; (ii) trying to infer from biology a SET of constraints that significantly increases the likelihood at arriving the desired result.

Once we set ourselves to the task of constructing a biomimetic scenario the following questions arise:

- What properties should such a scenario fulfil?
- What can we learn from the scenarios that were put forth to explain the origin of human language?
- How can we design such a scenario without hardwiring the end result into it?
- How can such a scenario be implemented?

In the following sections we consider these questions in turn.

11.4.1 Basic criteria for a selective scenario of language evolution

What properties should a selective scenario of language evolution fulfil? For a detailed analysis of the following criteria see Szamado and Szathmary, 2005

[1] I call language-like communication systems (LLCS) any system that has the following properties: use of symbols, compositionality and recursion, and cultural inheritance. These are the most important design features that distinguish human language from animal communication systems (Hockett, 1960a).

(TREE, in prep.). (1) First of all, if we assume an evolutionary computational paradigm, then an LLCS should obviously increase the fitness of those agents that are able to use it. (2) Second, recent game theoretical findings suggest (see Box.1.) that a communication system, using cost-free signals, such as human language, is highly unlikely to be able to evolve, if there is a conflict of interest between the potential communicating agents. (3) Third, the work of Steels and his colleagues (2002a) shows that the first concepts of LLCS must be grounded in reality, otherwise no shared symbolic representations can evolve. (4) Fourth, the most prominent practical feature of human language is its unparalleled expressive power. It enables human beings to make an unlimited set of statements – for this reason human language is an example of so-called unlimited hereditary systems (Szathmary, 2000);– moreover, it enables them to communicate these statements. Consequently, one shall look for a situation in which there was a need to use such an expressive power, or at least there is a positive feedback mechanism between the existing expressive power of LLCS and the optimal expressive power under the given situation. (5) The fact that no other species evolved the equivalent of human language strongly suggests that human language is a special adaptation. Hurford et al. (1998) argues in favour of a similar approach: "...in general, more realistically and more eclectically, for any set of circumstances proposed as individually necessary and collectively sufficient to explain the emergence of Language, one has to show that this combination of circumstances applies (or applied) to humans and to no other species." Thus, we shall probably look for a selective scenario that is not present in any of the extinct and extant species. (6) Human language requires some pre-adaptations (Hurford, 2003; see Box.2.), and "picking" the right set of preadaptations can be crucial for the success of the simulation. Determining the "minimal set" of these pre-adaptations should be a major challenge for future research.

A debate has been underway for a long time about the honesty of animal communication. The debate focuses on the proposition that signals need to be costly in order to be honest (Zahavi, 1975, 1977). While some models seemed to prove this statement (Grafen, 1990; Maynard Smith, 1991), others were able to show that cost-free signals can be evolutionarily stable even under a conflict of interest, provided certain conditions are met (Hurd, 1995; Bergström and Lachmann, 1998; Szamado, 1999). One of the important results is that cost-free signals, in general, are expected to be evolutionarily stable, provided there is no conflict of interest between the communicating parties (Maynard Smith, 1991). In line with this result, recent game theoretical investigations strongly suggest that prominent features of human language could arise under shared interest, rather than under a conflict of interest. On the one hand, a number of models show that some key features of human language can evolve, provided there is no conflict of interest between the participants (Nowak and Krakauer, 1999; Nowak et al., 2000). On the other hand, (Lachmann et al., 2001) were able to show that human language can be used honestly, even if there is a conflict of interest between participants. However, no one has been able to show that the key features of human language can evolve under a conflict of interest. The problem

	#1	#2	#3	#4	#5	#6
1. Language as a mental tool (Burling, 1993)	-	+	+	+	-	+
2. Grooming hypothesis (Dunbar, 1998)	-	+	-	-	-	-
3. Gossip (Power, 1998)	+	-	-	+	-	-
4. Tool making (Greenfield, 1991)	+	+	+	+	-	+
5. Mating contract (Deacon, 1997b)	-	-	-	-	-	-
6. Sexual selection (Miller, 2001)	+	-	-	-	-	-
7. Status for information (Dessalles, 1998)	+	-	-	+	-	-
8. Song hypothesis (Vaneechoutte and Skoyles, 1998)	-	-	-	-	+	-
9. Group bonding/ ritual (Knight, 1998)	-	+	-	-	-	-
10. Hunting theories (Washburn and Lancaster, 1968)	+	+	+	+	-	-
11. Motherese (Falk, 2004)	+	+	-	-	-	-

Table 2. The explanatory power of the various theories (Szamado and Szathmary, 2005, in prep.). (1) Selective advantage; (2) No conflict of interest; (3) Concepts grounded in reality; (4) Expressive power; (5) Uniqueness; (6) Origin of cognitive capacities.

with conflict-of-interest situations is that they are very susceptible to abuse, which can manifest itself in dishonesty or cheating, which in turn might ruin the signalling system. This logic was confirmed by various computer simulations of a simple communication game (Noble, 2000; Harris and Bullock, 2002). These simulations show that communication only evolves, if there is no conflict of interest between the signaller and the receiver.

Hurford (2003) gives a list of biological properties that served as cognitive pre-adaptations for language capacity: pre-phonetic, pre-syntactic, pre-semantic, pre-pragmatic and elementary symbolic capacities.

Donald's (1993) "executive suit" can be taken as another starting point: metacognition; self-monitoring; imitation; multitasking; auto cuing and explicit memory; self-reminding; self-recognition; purposive rehearsal; reciprocal intentionality; interpretational ability (symbolic ability).

11.4.2 Comparative analysis of selective scenarios put forth to explain the origin of human language

What can we learn from the theories that were put forth to explain the origin of human language? Table 2. (Szamado and Szathmary, 2005, TREE, in prep.) summarizes these theories and gives a short evaluation whether the given theory fulfils our criteria or not.

As we can see from table 2 there are only three theories that can answer a significant number of questions. These are as follows: language as a mental tool, tool making, and hunting theories. Thus, the most parsimonious explanation seems to be assuming that the origin of human language has something to do with primary representational systems, tool making, and hunting.

11.4.3 Language as an emergent communication system

Special care should be taken to avoid hardwiring expected solutions into the investigated model system. Thus, there is a need for a flexible (open ended) neuronal control mechanism that has the potential to evolve complex networks out of very simple small-scale networks, and thereby to give unexpected solutions.

11.4.4 Implementation – possible research strategies

Top-down: Start with our ,,educated guess" that is creating a selective scenario in which agents face a problem of cooperative hunting, which they can only solve by communicating their mental models with each other; in addition, the successful implementation of the task should require the use of certain tools. Unfortunately, this would be a fairly complex scenario assuming highly developed mental skills on the part of our agents.

Bottom-up: Start with simple selective scenarios characteristic of animal communication systems and then gradually increase the complexity of the task. The aim is to achieve the complexity of the selective scenario outlined above. This is definitely a more risky and more time-consuming approach than the previous one. Unfortunately –as noted above – the current state of the art makes it impossible to implement that highly complex scenario, thus this simple but more tedious bottom-up approach seems to be the only way to go at the moment.

11.4.5 Major research questions

1. What is the minimal set of pre-adaptations / cognitive skills that are necessary for LLCS to evolve?
2. What is the minimal complexity of a selective scenario that could select for LLCS?
3. What is the minimal sequence of selective scenarios – and what is the actual sequence – that can lead from a simple system typical of animal communication to the minimal complexity selective scenario?

11.5 Outlook: Stages in the emergence of language

In Chapter 10 the emergence of language is conceptualized to have occurred in VI stages, some of them carved up into sub-stages. The aim behind the open questions posed in the present chapter is also to implement the emergence of language in artifical, simulated or embodied agents. One can rightly ask about the connection between the two approaches. A few remarks are in order.

First, nobody knows how language originated and how this transition can be re-enacted in artificial systems, hence the need for parallel investigations.

Nobody knows where the crucial breakthrough may (or we hope: will) happen. Second, we do know that language did emerge in the human lineage. It may be the case that the easiest, and for all practical purposes the only way at arriving at language-like communication system is to follow the set of selective constraints that acted on the human lineage. Alternatively, a more enginnering-type approach may bear equal fruits, possibly even earlier. Third, as explained in Chapter 10, the six stages proposed there may not correspond to the actual stages of language emergence in biological evolution. It is an open, exciting question how emgineering and biomimetic scenarios might differ, even if they reach a similar goal in the end. The Table at the end of Chapter 10 clearly shows the lack of knowledge as to what the possible neural/neuronal implementations of the suggested stages could be. In this chapter we have discussed a few open questions in the neurobiological realm. We believe that it is possible to give answers to these questions. These would enable us to fill the empty slots in table, or contruct another table that may be closer to realistic stages that are in fact neuronally implementable.

Theoretical and Complex Systems Challenges

Luc Steels and Vittorio Loreto

Sony CSL Paris and La Sapienza Rome

12.1 Introduction

Recently the study of the self-organisation and evolution of language and meaning has lead to the idea that language can be viewed as a complex dynamical system (Steels, 2000), and hence the theoretical methods of complex systems science have become very relevant to study language. Recently there has been a growing body of work successfully investigating various statistical aspects of language (such as explanations of Zipf's law, Ferrer and Sole, 2003). In the ECAgents project we will focus however on challenges that are relevant for IT applications. Indeed we are now beginning to see concrete applications in the field of Embodied Communicating Agents, both in robotics and in large-scale peer-to-peer distributed information systems, so that it becomes necessary to have mathematical models of the semiotic dynamics generated by evolving language systems before the technology can be usefully employed. Most importantly, we want to have methods for predicting the macroscopic behavior of artificial systems based on the microscopic behavior of the agents.

A language is a semiotic system in the sense that it relates signs (words, grammatical constructions) to the world through the intermediary of conceptualisations (meanings) (see figure 1). Such a semiotic system can then be used for communication in the sense that a conceptualisation of an object, when it is distinctive compared to that of other objects, can then be used to draw attention to this object (or situation or aspect of a situation).

When language is viewed as an evolving system, each of these relations may change: new words and grammatical constructions may be invented or acquired, new meanings may arise, the relation between language and meaning may shift (e.g. if a word adopts a new meaning), the relation between meanings and the world may shift (e.g. if new perceptually grounded categories are introduced), both causing shifts between language and the world. All these changes happen both in the individual, as he or she is learning a language and its underlying ontology, and in the group as it is settling or maintaining a shared language and a shared ontology. Semiotic dynamics is the subfield of dynamics that studies the properties of such evolving semiotic systems. Because language use of one individual is obviously influenced by that of another, we are dealing with a multi-component system with individual behaviors as well as interactions between the components, similar to n-body systems of interacting particles, molecules, or organisms.

In chapter 10 of the White Paper a number of stages have been identified leading progressively to more complex human-like language. At each stage we

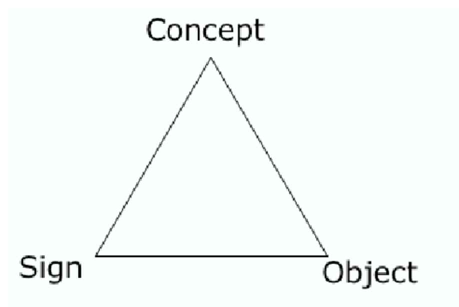

Fig. 1. The semiotic triangle groups the three basic semiotic relations. In the study of semiotic dynamics, the change in time in each of these relations is studied both in individuals and in a population.

define a minimal communication system that is fully adequate to handle the expressive complexity required at that stage. A minimal communication system includes mechanisms for conceptualising what to say, coding this into a sentence, decoding the sentence to reconstruct the intended meaning and interpreting this meaning back into the world. A minimal communication system also requires mechanisms for the expansion and adoption of the ontology and mechanisms for the invention and adoption of linguistic conventions.

At each stage, a number of theoretical challenges re-occur, but each stage also introduces its own additional challenges. The set of stages is summarised again in the following table:

Stage	*Meaning*	*Form*	*Issue*
I	Individuals	(proper) Names	Convergence on convention
II	Single Categories	Single Words	Co-evol. language/meaning
III	Multiple Categories	Multiple Words	Compositionality
IV	Multiple Obj. + Predicates	Grammar	Origins of grammar
IVa	id.	Syntax	Exploitation of syntax
IVb	id.	Grammar	Intermediary layers
IVc	id.	Recursion	Hierarchical re-use
IVd	id.	Meta-grammar	Two-level evolution
V	Second Order predicates	Grammar	Second order
VI	Meta-level	Grammar	Level formation

The 'stages' chapter of the White Paper (chapter 10) also introduced the framework of language games as a useful minimal model of Embodied Communicating Agents. This framework is also assumed for the present paper.

It is useful starting with an example. We shall consider the so-called "Talking Heads" experiment performed at the CSL-SONY laboratory in Paris a few years ago (Steels, 1999). In these experiments a population of robotic agents interacts performing, at each time step, binary language games, i.e. language games between two agents chosen randomly in the population. The outcome of these experiments is the emergence of a shared and efficient communication system in the whole population. These experiments represent then an ideal playground to identify the main ingredients which are supposed to be crucial for the development of efficient communication systems. It is possible to isolate in particular hardware and algorithmic, e.g. software, ingredients. Hardware ingredients are represented by the ensemble of the sensory-motors devices which allow a robotic agent to interact with the actual environment. The algorithmic ingredients are represented by the ensemble of procedures an agent performs to process the inputs received from the external environment and update its internal registry. The ensemble of these procedures can be simulated in an artificial environment and analyzed in great details. In this perspective three main theoretical challenges can be identified.

(a) identifying and defining the simplest (minimal) models (i.e. algorithmic procedures) which could lead to efficient communication systems. It is important to stress the need in this field of shared and general models to create a common framework where different groups could compare their approaches and discuss the results. On the other hand the models should exhibit the extreme level of simplicity compatible with the desired phenomenology. This has several advantages. It could allow for discovering underlying universalities, i.e. realizing that behind the details of each single model there could be a level where the mathematical structure is similar. This implies, on its turn, the possibility to perform mapping with other known models and exploit the background of the already acquired knowledges for those models.

(b) identifying the most suitable theoretical concepts and tools to attempt the solutions of the models. It is important to outline how the possibility to obtain analytical and general solutions for the models proposed could open the way to a positive feedback providing further inputs for understanding and designing new experiments and devices. In general there are two main questions one should be able to answer. On the one hand, we need to find the general laws that govern the semiotic dynamics of a particular system, for example, how the maximum number of words in use is related to the number of agents in the population. On the other hand, we need to find the explanation of these laws as a mathematical property of the dynamics.

(c) coupled to the theoretical activity there should always be an experimental activity with a twofold aim. On the one hand the experiments, as well as the observation of the realities one is interested in, provide inputs for the modeling and the theoretical activity. On the other hand they represent the framework where the theoretical predictions are checked. The outcome of these experiments will be compared with the theoretical and the numerical results and it will be used to better focus the modeling and the theoretical approach. There should then exist a positive feedback mechanism between the theoretical and the experimental activities in order to make the progresses robust, well-understood and concrete.

Just to give an example of a possible path to attack the problem let us focus on a specific example already carried out for the so-called Naming Game, possibly the simplest stage of the language games performed by a population of agents. In this game, a set of agents can communicate pair-wise by naming objects of the environment. Each agent relies on his own inventory listing name/object relations. Starting from empty inventories, the agents create or learn names by performing games with each other until a final state is reached where all agents share a common system of conventions. This final state is approached through a characteristic path depending on the parameters of the game. Besides the number of agents and objects in the system, the approach to the final state is determined by the parameters entering in the rules governing the inventory evolution through word acquisition and loss. Whereas the evolution of the system can be studied in detail through computer simulations, a theoretical description of the complex dynamics would be highly desirable. It is thus important to define models of increasing complexity in order to understand the role and the effect of the different ingredients of the models. At the simplest level, i.e. at the level of the so-called Random Naming Game, where neither biased selection of words in competition nor elimination of them is present. In this case each agent keeps for every object all the associations form-meaning that has ever experimented. As a result the population reaches a successful communication system which is not efficient, due to the high number of associations form-meaning each player should keep in mind. This model has been solved at the mean-field level and the agreement between the theoretical predictions and the simulations is excellent. What should characterize an efficient communication system is a unique association form-meaning shared by all the individuals of the population. In order to reach this goal we had to pass to a higher level of complexity by introducing models where a selection and competition of associations form-meaning was at play. After a careful and systematic exploration of the parameters space we ended up finally with a much simpler version of the naming game, which leads to a successful and efficient communication system. This model is then the ideal candidate for an analytical treatment as well as for further generalizations. A preliminary mean-field approach has been already developed and it gave already some encouraging though preliminary results. It had been possible for instance to clarify several aspects of the general phenomenology: role of the population size, role of synonymy and homonymy, etc. On the other hand this model could be used as a starting point to address more complex and realistic language games framework (cf. chapter 10).

Next section lists the basic ingredients which characterize the semiotic dynamics of a population of agents during the emergence of a successful communication systems.

12.2 Main features of semiotic dynamics

An evolving semiotic system has a number of external parameters which circumscribe the complexity of the semiotic task:

1. The size of the population N_{ag}. This is the number of agents that form a community and have to agree on a shared language inventory and underlying ontology, related to the environment.
2. The flow-rate of the population $\rho_{N_{ag}} = \dot{N_{ag}}$. The rate at which new agents enter and leave the population. In the simplest cases the number of agents is considered constant. At higher levels of complexity one could ask how a change in the population size affects an already well-established communication system, i.e. when the change occurs after the population reached coherence. If the population change occurs during the establishment of the communication system the question is whether or not a successful communication system will emerge (eventually of what type).
3. The complexity of the environment which is quantified in terms of the number of possible meanings M that needs to be expressed. For example, in Stage I (naming game) the set of possible meanings is the set of possible objects in the domain. In Stage II (single category guessing game) the set of possible meanings is the set of possible categories.
4. The flow-rate in the meaning space $\rho_M = \dot{M}$, i.e. the rate with which new meanings become relevant and others become irrelevant.
5. The error-rate in transmission ϵ_s, which determines how reliable utterances are transmitted or received by other agents. For example, if a speech medium is used it will not be possible to get absolute accurate transmission of the utterance.
6. The error-rate in feedback ϵ_f, which determines how far the non-verbal communication is reliable. For example, on real robots there are likely to be serious difficulties in pointing or joint attention.
7. The error-rate in cognition ϵ_c, which may cause deviations in the application of the inventory due to forgetfulness, sloppiness, etc.

An evolving semiotic system also has a number of properties which can be used to determine how far the system is optimal. They depend on the nature of the inventory. Here are some examples:

1. Total number of words in the lexicons of all agents N_w
2. Average number of non-zero strength word-meaning pairs in the population (inventory) I_{ag}
3. Number of syntactic and semantic categories in the grammars of the agents N_c.
4. Number of grammatical constructions in the grammars of the agents N_g.

12.2.1 Communicative success

Obviously the most desirable property of an evolving semiotic system is that the agents reach communicative success. What constitutes success depends on the nature of the game. For example, in the guessing game it means that the speaker agrees that the topic identified by the hearer is equal to the one he initially decided to draw attention to.

The cumulative communicative success $S_a(t)$ of an agent a at time t (i.e. after t games) averaged over the last n games (time-steps) in which a participates, can be defined as:

$$S_a(t) = \frac{1}{n} \sum_{i=t-n+1,t} s_a(i) \tag{1}$$

where $s_a(i)$ is the score of a single guessing game played by two members of $\mathcal{A}$, one of them being a.

The cumulative success $\mathcal{S}(t)$ for a population of N_{ag} agents $\mathcal{A}$ for the last series of n games at game t is defined as

$$\mathcal{S}(t) = \frac{1}{2} \sum_{a=1,N_{ag}} S_a(t). \tag{2}$$

12.2.2 Convergence

It is next useful to compare the ontologies and language inventories of different agents in a quantitative way in order to track whether convergence is reached. Convergence is obviously related to communicative success but there can be high degrees of success without absolute convergence if there is sufficient sharing.

The category variance *cv* between the agent ontologies can be measured by computing the cumulated distance between the ontologies of the agents of a population $\mathcal{A} = \{a_1, \ldots, a_{N_{ag}}\}$, as in

$$cv(\mathcal{A}) = \frac{2}{N_{ag}(N_{ag}-1)} \sum_{i=1}^{N_{ag}} \sum_{j=i+1}^{N_{ag}} D\left(a_i, a_j\right), \tag{3}$$

where we have introduced $D(a_i, a_j)$ as a distance measure between the ontologies of agents a_i and a_j, to be defined in terms of the specific implementation of these ontologies.

Similarly we can define a measure of linguistic variance *lv* between the language inventories of different agents by computing the cumulated distance between the inventories of the agents of a population $\mathcal{A} = \{a_1, \ldots, a_{N_{ag}}\}$, as in

$$lv(\mathcal{A}) = \frac{2}{N_{ag}(N_{ag}-1)} \sum_{i=1}^{N_{ag}} \sum_{j=i+1}^{N_{ag}} L\left(a_i, a_j\right), \tag{4}$$

where $L(a_i, a_j)$ is a distance measure between the language inventories of agents a_i and a_j. This measure is to be instantiated depending on the structure of the inventory or what aspect of the inventory one is interested to track.

Given these measures of variance at a particular point in time, we can define the cumulative variance similar to the way cumulative success is defined for discrimination or guessing games. Other useful measures quantify the size of the language inventory or the ontologies of individual agents as well as the group, the degree of synonymy or homonymy present in the inventories of the agents, the number of syntactic and semantic categories, etc.

12.2.3 Optimality

Communicative success or convergence do not yet determine whether the agents will use an optimal communication system. Definitions of optimality depend on the stage of complexity and the type of inventory. Criteria for optimality generally have to do with the number of linguistic elements (words, linguistic categories, grammatical constructions) that are used by individual agents or by the population.

12.2.4 Network topology

A critical aspect of communication is the communication topology, that is, who communicates with whom. Network theory has developed very useful theoretical and modeling tools for understanding the structure and consequences of different types of networks. It is important to understand the role of the communication topology in collections of agents, both in concrete simulations and more abstractly. In particular there are two main questions: (a) what is the effect of a non-trivial communication topology, e.g. hierarchical structures in which information converges on a single agent or a sub-set of the agents, of a group of embodied agents (e.g., contrasting with the infinite dimension case, in which everyone can communicate with everyone); (b) what communication topologies spontaneously emerge in simulations in which collections of agents have some particular task to perform.

12.3 The lesson of complex systems

It is useful to stress the importance of developing a common language and a fruitful exchange between complex systems research and IT for the specific area of evolving communication systems. IT-researchers need to be able to phrase problems in terms understandable to complex systems researchers and vice-versa there needs to be an infusion of culture from IT into complex systems research.

More specifically the following areas of complex systems research are particularly relevant for the approach to the emergence of successful communication systems.

12.3.1 Evolutionary Game Theory

The project partners believe that it is possible to transport the methods and techniques from Game Theory, as currently applied to biology and economics, to the domain of evolving communication, particularly by the development of a formal theory of evolutionary language games. Systems of interacting agents have been widely studied in economic and biological contexts, however, as pointed out previously in this proposal, the traditional theories often assume hypothesis which are not realistic when compared to real or even artificial embodied agents. Moreover, in most of the literature on conflict-cooperation in both animal and human societies (e.g., behavioral ecology, economy), the important question of the characteristics of communication among interacting agents is not analyzed. A main objective should then be that of analyzing agents collective behaviour under the most general conditions: first of all one needs to understand how coordination and learning may arise when agents are endowed with given properties and interact among themselves and/or with the environment in a given way. Still, one wishes to allow the widest possible variety on the choice of the individual profile or the nature of the environment.

Beyond traditional Game Theory

The experience of traditional Game Theory is rather restrictive, in that severe assumptions are made on the rational behaviour of individuals, the role of the environment where they act, the nature of the information they process to take decisions. Rather, realistic problems often involves agents who are not rational, are extremely heterogeneous, act in a changing and inhomogeneous environment, interact through non trivial topologies and receive different and noisy information. To deal with this sort of systems, one has to relax specific prescriptions on agents or environment properties, and rather adopt random variables to model their variety in time and/or space (this, for example, is what has been partially done in Game Theory to deal with Incomplete Information). The analysis of these models become of course much more complex, especially when the number of agents is large, and new tools are needed to treat the presence of stochastic variables. An important help in this direction comes from the experience gained in Physics in the last thirty years. Indeed, systems of interacting units in presence of disorder are an old problem for which Statistical Mechanics has developed some very powerful techniques. Neural Networks and Random Optimization Algorithms are just two well known examples where these techniques have been successfully applied. More recently, lot of interest within this perspective has been devoted to Competitive Heterogeneous Agents with inductive reasoning. Starting from the first attempts of the Santa Fe market model (Palmer et al., 1994), to simpler and more focused problems, there is now a wide and increasing literature in the econophysics community that can be regarded as a starting point for further research (see, for example Farmer, 1999). Many of the studied examples wish to ultimately refer to the reality of financial markets (Bouchaud et al., 2001; Giardina and Bouchaud, 2003), or systems with global constraints (as the Minority Game problem (Challet and Zhang, 1998; Cavagna, 1999; Challet et al., 2000), however their structure can be easily modified to deal with systems with a non-market interaction (i.e. systems where agents do not simply interact through a global quantity as the price), as most of social and biological systems are (Glaeser and Scheinkman, 2000).

Modified Evolutionary Games

In general terms, the typical structure of these modified evolutionary games consists in a network of interacting agents endowed with heterogeneous evolving strategies of action. The choice mechanism of agents is specified via appropriate utility (reward) functions and encompass adaptive behaviour through learning. Agents may interact among themselves (directly or indirectly) and with the environment, and receive information on specific issues which they then process as input for their decisions. As a result of individual actions, the system exhibits a certain collective behaviour whose specific properties depend on the kind of the adopted choice mechanism and interaction patterns. Such a collection of interacting agents may approach a dynamical stationary state with different degrees of coordination, or rather keep evolving on cycles or more complicated dynamical patterns; it may reach some kind of equilibrium state - in the Game Theory (Nash) sense or in a more robust way; it may exhibit self-organization or not. Within this context, many different and specific questions can be addressed. How the choice mechanism determines the qualitative system's behaviour? One can start with a set of fixed random possible strategies of action per agent and progressively introduce more refined measures of performance, allow for mixed rather than pure strategies, allow for perturbations in the strategy space, introduce selection by allowing the set of strategies to change in time, and so on up to more complex models where each agent has a neural network to retrieve the more convenient strategies. How the nature of the interactions influence global and individual performances? Agents may interact directly or through global constraints, they may respond to competitive or socializing instances, they may feel exogenous fields or not. What is the role of the network topology ? This question is very much related to the previous one, indeed when assuming that agents interact among themselves we need to specify not only the nature of the interaction, but also how it affects the individual agents: each agent may interact with all the others (fully-connected topology), or with just a few of them; the interacting agents can be chosen at random (random graph topology), they can be located in nearby position in space (3-d nearest neighbour topology) or interact on a more specific network structure. Obviously the network topology very much influences the correlations among agents, clustering properties and crowd effects. Also, in dynamical terms it affects crucially the speed of information flow through the system (see also Network Theory section below). At a more sophisticated level one could imagine to extend these techniques for evolutionary language games. In most of the literature on conflict-cooperation in both animal and human societies (e.g., behavioural ecology, economy), the important question of the characteristics of communication among interacting agents is not analysed. How communication may be shaped in relation to its function (to prevent conflict, to identify cheaters, to diffuse reputations) is one of the essential questions for which game theory based approaches may provide an answer. Game theory might also be relevant to develop theoretical models that allow us to prove whether certain behaviours by embodied communicating agents effectively lead to shared communication systems or not and under what conditions this can be maintained and complexify.

12.3.2 Information and Optimisation Theory

At the end of 40s C. E. Shannon, in two fundamental works (Shannon and Weaver, 1962) faced the problem of an efficient transmission of messages laying the foundations of a mathematical theory of communication. Information Theory has since then acquired a leading role in such areas as computer science, cryptography, biology and physics (Zurek, 1990). One of the most important contributions of Information Theory is the discovery that the amount of information contained in a message can be measured in an objective way: the tool for this is the concept of entropy. In Information Theory, in fact, the word "information" acquires a very precise meaning, namely the "entropy" of the string. In a sense, entropy measures the surprise the source emitting the messages can give us. Suppose the surprise one feels upon learning that an event E has occurred depends only on the probability of E. If the event occurred with probability 1 (certainty) our surprise at its occurrence would be zero. On the other hand if the probability of occurrence of the event E was quite small our surprise would be proportionally larger. For a single event occurring with probability p the degree of surprise is proportional to $-\log p$. For a generic source the so-called Shannon entropy gives, at the same time, both the number of bits per symbol that are necessary to codify each one of the (long) sequences emitted by the source and the rate of growth of their number with the length of the sequence. It is interesting to recall the deep relations linking the compressibility and predictability of a sequence of symbols to the entropy (of the source) or to the Algorithmic Complexity of the sequence itself. Let us consider, without loss of generality, the sequences of characters representing English texts. The entropy of the English language can be defined as the minimum number of bits per character necessary to encode an ideally infinite message written in English. In order to estimate this quantity one should be able to subtract the unavoidable redundancy that comes always along with any linguistic message. The redundancy can also be seen as the number of constraints (for instance lexical or grammatical rules) imposed on the English text. For example the fact that a q must always be followed by a u or the impossibility to have two subsequent h are dependencies that make the English language more redundant. Rules of grammar, parts of speech, and the fact that we cannot make up words make English redundant as well. Redundancy is actually beneficial in order to make the message transmission efficient in noisy conditions or when only part of a message comes across. For example if one hears "Turn fo th lef!", one can make a fairly good guess as to what the speaker meant. Redundancy makes then language more predictable. Imagine watching a sequence of symbols emitted by on a ticker tape. The question one could ask is how much information will be added by the next symbol s_i once one knows already the sequence $s_1, ..., s_{i-1}$. How much information will be gained upon seeing i s fixes also the amount of surprise we experience. An extreme surprise will convey large information, while if one can reliably predict the next symbol from context, there will be no surprise and the information gain will be low. The entropy will be highest when you know least about the next symbol, and lowest when you know most.

Information Theory and Language

Except very recent studies (see for instance Plotkin and Nowak, 2000) the connection between Information Theory and language or communication evolution has been mostly disregarded. This project aims to bridge this gap since the partners involved believe Information Theory could give a very important contribution in several directions. Information Theory allows thus for a quantitative measure of the information content through concepts like entropy and Algorithmic Complexity or more sophisticated measures like relative entropy or mutual entropy. In a world in which new information is generated constantly and only progressively and partially represented explicitly, Information Theory can then provide with new measures of information content and information growth. Moreover Information Theory can give important contributions in learning phenomena, either vertical learning (from a source) or horizontal learning (from other agents, i.e. mutual), providing with objective measures of how the learning process evolves. Last but not the least Information Theory can play an important role in problems of communication optimisation. Most of the optimisation approaches adopt a static point of view, neglecting the link between the behavioural mechanisms of the agents and optimality. It is then important to analyse and clarify this link between the individual algorithms processing information and governing the flow of information and the efficiency of the collective dynamics. Finally, information theory can also be relevant to identify the properties of media that enable them to become carriers of information.

12.3.3 Population dynamics

Communicating embodied agents cooperate to achieve a task. Complex systems research can help to analyse the relationship between the characteristics of the communication system, the dynamics of the group, and the characteristics of the task to be performed, and can propose a taxonomy of the dynamics of collective activities. Interacting agents influencing each other's behaviour are by definition coupled non-linear dynamical systems. Over the past decade, research has been conducted in such systems from a complexity point of view and the project partners believe that these results could yield basic methods and techniques for studying the interactive foundation on which communication can be built. A crucial problem in modeling dynamics of embodied agents relies in an appropriate description of the environment properties. Often the environment presents features which change in space and time and that heavily influence the population behaviour. To describe such situations useful insights and techniques can be drawn from the wide literature in Statistical Mechanics on surface growth and diffusion problems. Indeed the connection between population dynamics problems and diffusion/growth problems is direct, if not trivial. The research on stochastic growth and anomalous diffusion has dominated the Statistical Physics of the last years (Bouchaud and Georges, 1990; Halpin-Healey and Zhang, 1995), and can be directly used to deal with problems of biological or social origin where the presence of unknown parameters is relevant. The basic idea is that inhomogeneities in environment properties, or fluctuations in the growth/death rate can heavily affect the evolution of a population: regions of favorable conditions attract individuals and favor their wealth determining localization in space, while for example external conditions variable in time may change reproduction rates. The existence of stable fixed points of the dynamics, i.e. stationary states for the population, and the way in which the population diffuse in space and time, are subtle problems of difficult analysis. Some interesting attempts have been performed in biological context (Dahmen et al., 1999) and diffusion problems (Giardina et al., 2001) and could be fruitfully extended to more specific problems. The project partners believe that these results could yield basic methods and techniques for studying the interactive foundation on which communication can be built.

12.3.4 Network and patterns

Communities of interacting and communicating agents create dynamically evolving network topologies. It is of great importance to use the techniques recently developed in network theory to be able to describe and track these networks, to study the conditions under which they form and collapse, and the conditions and consequences of different interaction topologies. Networks of interactions in ECAgents might be very different from simple ordered (Euclidean) lattices. The Internet is only the most conspicuous example of this kind of complex structures. While it is immediately clear that such networks can not, even approximately, be described in terms of ordered lattices, it has been recognized recently that their topological properties are very different also from those of random graphs, for which a well-established mathematical theory exists (Bollobás, 1985). The understanding of the properties of complex networks is currently the goal of an extremely active field (Albert and Barabasi, 2001; Newman, 2003). The emerging picture is that complex networks are in some sense intermediate between random graphs and Euclidean structures. In particular, many networks exhibit a small average distance between vertices (small-world effect), typical of random graphs, together with local clustering properties, that are typical of ordered lattices (Watts and Strogatz, 1998). Moreover, many social networks are scale-free (Barabási and Albert, 1999), in the sense that they exhibit a power-law distribution of the degree (i. e. the number of connections of each vertex). These new and nontrivial properties have profound implications for physical or social processes occurring on complex networks. Traditional concepts and tools for the study of such processes must be adapted, and in some cases completely reformulated, when dealing with complex networks. A further important issue in this context is the observation that in growing networks the local character of decisions introduce sub-optimal features. As an example, information will flow in sub-optimal ways or the network will be particularly fragile against the removal of some key nodes. Since agents make decisions based on local knowledge, the global performance of a net will be typically lower than optimal. Future models and evolutionary algorithms should incorporate top-down information that might help change local decisions in order to avoid falling into sub-optimal global performance. Transitions from a suboptimal to an optimal organization can be sharp (Ferrer and Sole, 2003) and thus understanding of the intrinsic nonlinearities arising in communicating agents are necessary.

12.4 Challenges

Given the framework depicted and the ensemble of knowledges and techniques one can borrow from the experience of the study of complex systems, it is possible to formulate a certain number of open questions. A first account of them can be summarized as follows:

> **Challenge a)** Predicting whether a given population of size N_{ag} will eventually reach, within a finite time T, maximal communicative success $\mathcal{S}$ to express a given set of meanings M and linguistic and ontological coherence and optimality, given a specific set of behavioral rules adopted by each agent.
>
> **Challenge b)** Identifying and understanding the laws governing the scaling between population size N_{ag}, the size of meaning set M and the time T for reaching communicative success and give an explanation for this law in terms of the microscopic dynamics.
>
> **Challenge c)** Identifying the dependency between the flow-rate of the population $\rho_{N_{ag}}$ and the flow-rate of the meaning set ρ_M and the maintenance of communicative success $\mathcal{S}$, coherence and optimality.
>
> **Challenge d)** Showing the dependency between the flow-rate of the meaning space ρ_M and the maintenance of communicative success $\mathcal{S}$, coherence and optimality.
>
> **Challenge e)** Exploring the role of the network topology on the establishment of a successful, coherent and optimal communication system.
>
> **Challenge f)** Exploring the interplay between the network topology and the semiotic dynamics. In particular one is interested in understanding what type of network structures emerges when a population of agents tries to establish a successful, coherent and optimal communication system.

It is evident how the questions posed are at this stage very general and in some cases very fuzzy. It will thus be important, as our knowledge will progress, to clarify and better focus the questions, in order to make the theoretical contribution as effective as possible.

Part III

Applications

Application Opportunities for ECAgents

Lars Erik Holmquist, Frederic Kaplan, Luc Steels

Viktoria Institute and Sony CSL Paris

13.1 Introduction

Technologies, models and concepts created in the ECAgents project can have great potential to enable novel applications. As a multitude of devices with advanced computing and communication capabilities become available for everyday users, the concept of physically embodied agents that communicate and evolve through interactions between themselves and users should be applicable to many application areas.

However, developing new applications is not unproblematic. The step from research result to useful application is often large, especially in a high-risk research project like ECAgents. Therefore, there is need for a systematic effort to turn theoretical project results into application concepts that have relevance for the real world. Furthermore, in order to evaluate such concepts, it is also necessary to create application prototypes that can be tested in realistic settings. While several innovative design methods have been presented in recent years, in particular in the interaction design research community (e.g. Buchenau and Suri, 2000; Gaver and Martin, 2000), it is not always clear how they can be adapted to other areas. Therefore we need to pay particular attention to how existing methods can be applied to the project, and what novel methods can be developed that suit our particular area.

At present, there are as far as we can see no existing products that would fully fit the definition of an "ECAgent application", i.e. a system incorporating embodied communication between a collection of autnomous agents that evolve their own language. Currently, the concepts explored in the project are still too advanced to be the basis for a commercial system. This will change, just like many earlier concepts in artifical intelligence have moved from research prototypes to everyday applications. But while there are no "true" ECAgents applications, there are several products that exhibit properties that can be useful to study as a source of inspiration and reference. In particular, domestic robots are becoming more and more advanced, both in the form of toys (like the popular *Robosapien* humanoid robot) or other devices (e.g. houscleaning robots such as the *Trilobite* automatic hoover). There is therefore need for a systematic investigation into the properties of existing and future products based on embodied interaction and evolving communication.

It is no probably coincidence that many of the existing "ECAgent-like" systems we will discuss fall into the area of entertainment. We have found no relevant applications yet deployed in areas such as helthcare, public safety, logistics, telematics, etc. A reason could be the following: the same properties that make autonomic and evolving systems *unsuitable* for certain applications - the unpredictable nature of the results, the need for a long training period, etc. - is just what makes them *suitable* for entertainment. To stay interesting, entertainment applications require constant change and stimulation, whereas safety-critical applications must be predictable and reliable, something which can take a long time and be difficult to achieve in an evolving system. For instance, while we would for instance probably never accept unpredictable responses from the steering wheel of our car, unexpected and autnomous behaviour can on the other hand be desirable in an electronic agent that embodied in a doll for children. However, as the concepts and technologies in the ECAgent projects continue to be developed, we may see a situation where the results can be applied to a much broader area of applications.

The following section of the White Paper will first discuss the requirements for development methods that take into account the unique properties of ECAgents, and suggest one such method, *user-driven innovation.* We will then discuss how some existing products can be used as “good examples” for how to design interactive applications based on ECAgents results. Finally, a number of potential applications and application areas will be discussed, including robot-based applications and peer-to-peer networking.

13.2 Development methods

13.2.1 Approaches

To be relevant outside the research lab, new technologies must be compelling and useful for others than the developers themselves. The last twenty years has seen *user-oriented system development* become an important component in computer science. Originally, human-computer interaction research primarily concerned the post-hoc evaluation of computer systems, i.e., the systematic investigation of how efficient a system or interface is when in use. However, it soon became apparent that as a complement to evaluation, users can also be a valuable resource before the construction of a system has even begun. For instance, *participatory design* was introduced as a way to involve users as co-designers at a very early stage when developing a system (Greenbaum and Kyng, 1991). Here, mock-ups and scenarios are often used in a dialogue with users to better understand the real-world implications of a proposed system. Another popular approach has been to use *ethnographically inspired methods,* such as participant observation (e.g. MacKay, 1999). Here, researchers observe and document the current work practices of a group of users, taking this information into consideration when creating a new system. This can help to ensure that the system is anchored in an existing work practice, rather than imposing a completely new way of working.

However, a major issue with user-centred methods is that they often fail to take into account novel properties of new technologies. When the user requirements, rather than innovative ideas and theories, are the main shaper of an application, the results may be useful but often fail to introduce novel concepts or take advantage of the latest avances in technology. Therefore we suggest a new approach, *user-driven innovation* (Holmquist, 2004). This is closest in spirit to participatory design, in that potential users are regarded as a resource in the design process. However, there are some important differences. One is that in this approach projects are often already well on their way in the prototype stage when we start to involve users. We are interested in cutting-edge technology, and in many cases, the initial idea for a project will be based on technical possibilities rather than any particular user need.

Furthermore, unlike participatory design and most ethnographic approaches, user-driven innovation does not necessarily regard the groups that are engaged in the process as the final users of a proposed system. Instead they are seen as a springboard that will help to push the technical ideas further. For this reason, there is particular interest in finding users that have very specialized and perhaps peculiar requirements. The belief is that such specialized groups are more likely to put technology in a new light, thus giving rise to interesting ideas. We can think of them as "extreme users", an analogue to the concept of "extreme characters", which are fictional persona that are created to generate ideas in interaction design (Djajadiningrat et al., 2000). As with extreme characters, the purpose is to inspire novel ideas that can be generalized for a larger audience. Often, insights gained from working with specialized users can push the original technology much further than would otherwise have been the case.

13.2.2 Challenges

The main challenge for ECAgents is to find methods that marry the results in the project with real-world application domains. The various existing approaches, such as ethnographically inspired design and participative design, have been developed for other domains and can not be directly applied to the project. However, elements of such methods can be used to increase the chance of generating high-quality and interesting applications, since purely technology-driven application development is not likely to succeed. User-driven innovation is an attempt to create a sort of hybrid between technology-driven development and user-oriented design. It is however in no way the last word in this department, and it may be necessary to adopt or develop other methods to achieve the goals of the project.

The user-driven innovation method has already been succesfully applied in several projects (e.g. context aware photography (Ljungblad et al., 2004) and smart noticeboards (Helin et al.)). In the ECAgents project, we are currently applying it to two areas: music distribution in mobile peer-to-peer networks, and robot applications for everyday use. However, the method is still fairly new and under development. It is an open issue how well we will succeed in creating application concepts and demonstrators that combine relevance for a real-world user-group with advanced concepts from the project. This will be further explored in the coming stages of the project.

13.3 Learning from existing applications

13.3.1 Overview of existing products

While there are no "true" ECAgent applications that we know of, by laxing the definition slightly we can find many products that have properties that are of interest in the project. While these products may not for instance evolve communication languages in the sense explored in the project, to the user they do exhibit some degree of autonomy and communication capabilities. This autonomy is usually only simulated but in some case it is in fact based on quite advanced concepts in AI and artificial life. Therefore, to gain inspiration for future ECAgents applications, it is interesting to briefly examine a few categories of existing products.

An early example that caught the attention of many consumers was the *Tamagotchi*, launched by the Japanese company Bandai in 1996. It sold over 40 million units world-wide. The Tamagotchi was a form of small-scale artifical life that has to be nurtured by the user. The agent was graphically embodied on a small handheld device with a screen and a few buttons. Only by "feeding" and "playing" with the agent could it be kept alive and happy - if it was ignored it would whither and die. The algorithms in the Tamagotchi were very basic, and the creature and its behaviour were derived from a series of simple rules. However, the rules were cleverly designed in such a way to continually generate interest despite there being no real evolution involved. In 2004, a networked version of the Tamagotchi was introduced: Tamagotchi Connection. This allows Tamagotchi creatures to communicate and even "mate".

A similar toy but physically embodied as a furry doll was the *Furby*, released by Tiger Electronics in 1998. It sold over 10 million units in 1999 alone. In a similar way to the Tamagotchi, the Furby requires the user's attention to thrive. But unlike the virtual pet it is actually embodied in the real world, using sensors (microphone, light sensor, accelerometer, pressure sensors) and actuators (a loudspeaker, and a step motor that gives it a large variety of exrpessions and movements). It reacts to the users actions, and gives a good impression of being aware of the world and responding to external events. Over time, it seemingly "develops" from an infant to a grown up. It can also communicate with other Furbys using infrared light. However, as with the Tamagotchi, no real evolution or communication takes place; all movements and utterances are pre-recorded, and the evolution follows a pre-programmed pattern. Still, for the uninitated, the Furby gives a very good impression of being an autnomous agent with its own goals and agendas - at least over a shorter time span.

Personal computer games have often made use of advances in artificial intelligence and artificial life. An interesting example is *Creatures*, released in 1996 by Cyberlife. It allowed users to breed and train evolving artifical life forms, based on bilogical metaphors. The creatures are embodied as a form of carton characters who live in a richly detailed graphical world, with simulated physics and eco-system. The player does not directly control the creature but rather shapes its environment and gives it indirect commands, observing it as it learns and develops over time. Users can even exchange artficial DNA over the internet to create new species. The game was first considered difficult to market but was in fact a surprise hit and sold more than 2 million copies. The technology behind the game (a third version was recently released) is probably the most advanced artificial life system available in a consumer product. Another product that used similar technologies in a more action-oriented framework is *Evolva* from Computer Artworks, released in 2000.

A similar genre of computer games has been the management simulation, which allows players to control a simulated virtual world, e.g. a city in *Sim City* or an ant farm in *Sim Ant.* The most popular such game is *The Sims,* released in 2000 by Electronic Arts. The first version had shipped a total 6.3 million copies in March 2002, making it the best selling computer game of all time. The Sims can be considered a "people simulator", because the focus is on simulated individual agents, who resemble real humans and live ordinary lives in a 3D virtual environment. The agents are fully autonomous and act out their own goals and desires in response to changes in the environment. The game can be considered a social simulation, since much of it revolves around the social life and activities of the agents. The simulated humans have an active social and professional life, meet and entertain, have careers and hobbies and so on. The latest version, released in 2004, is even more complex and is probably the most advanced agent-oriented technology available for consumer use.

Another recent trend is that of entertainment robots. While Sony's AIBO robot dog has been geared mostly towards hobbyists, and the humanoid QRIO is still to expensive to sell to consumers, recent advances indicate robots are now becoming consumer products. *Robosapien* is a humanoid consumer robot manufactured by Wowee and released in late 2004. It ha an innovative balance system that allows it to walk efficiently, and while not displaying any advanced intelligence it is claimed to have an "interactive reflex system" that allows it to act on external input. In reality, the robot is basically remote-controlled and can be pre-programmed to perform a sequence of moves. The "reflex system" is a feedback loop that gives it rudimentary capabilities to react to changes in the envrionment, for instance to avoid falling over. Although it exhibits even less "ECAgent" qualities than several of the other examples, it is interesting in that it shows that locomotive robots can now be manufactured at an affordable price, and in that it has introduced the concept of a personal robot as a viable product concept.

13.3.2 Lessons learned

The above has been a brief overview of existing applications that in some way exhibit autonomous actions and evolving behaviour (although in several cases it is only an illusion created by clever design and has little to do with "real" artificial intelligence or A-life). What is common for them all is that they are very commercially successful, among the most succesful in their respective product category. However, what is also common is the tendency for "fads" - Tamagotchi and Furby in particular are examples of trends which were immensely popular for a short period and then lost interest in the public eye. What can we learn from these examples?

One important lesson is that there is a real public interest for toys and games that exhibit evolutionary behaviour. The Sims and Creatures are based on complex algorithms that would not be out of place in many research projects. There is certainly a form of "real" evolution going on in both these examples, although perhaps not so profound as for instance the development of a new language. But in order to turn these concepts into products, the makers have managed to hide the complexity in an attractive form, and translate it into metaphors that people can understand: those of social interactions and pets, respectively.

Another lesson is that autonomous behaviour can be faked - for a limited time. Both the Furby and Tamagotchi were very popular when introduced. But the public eventually lost interest, possibly because the "tricks" the designers used to fake autnomous behaviour and evolution became evident after spending more time with the product. It is difficult to know if the products would have stayed popular longer if they had been constructed so that they somehow continued to evolve beyond the pre-programmed patterns. At the times they were introdcued, available technology simply did not allow this kind of evolution in an affordable consumer product.

Finally, the Robosapien, while perhaps the least interesting in terms of behaviour, shows that the concept of embodied agents in the form of robots is now a viable consumer market. But only if they can be coupled with more interesting behaviour, will they stand a chance of being more than a "fad" and keep the public interest for an extended period of time.

13.3.3 Challenges

The examples above are somewhat limited in that they are taken from the entertainment domain. They will need to be studied further and the lessons learned should be generalized to other domains. There is also a need to explore other examples where these kind of properties have found their way into real-world products, outside the entertainment domain. Even though such examples are harder to find, it is possible that there are some relevant products already in existence that we may learn from.

For the development of ECAgents applications, the challenge is to create products that have aspects of "real" evolutionary behaviour, yet present it in a form that makes it attractive for the general consumer. This can be greatly helped by charting what has made existing applications succesful - or unsuccesful. It can also be valuable to bring in the efforts at instituions where robot applications are studied in a context that includes aspects of design and social science, e.g. the *Project on People and Robots* at Carnegie Mellon University.

13.4 Potential application areas

In the following, we will discuss potential applications for ECAgents that will be explored in the project. They include:

- Robotic applications, where agents are embodied in physical form as autnomous robots. Potential areas incvlude service robots and entertainment robots
- Ubiquitous applications, where agents can move between different forms of physical embodiement. Here, agents can exist on devices such as PDA:s and digital cameras
- Peer-to-peer applications, such as music sharing, where the problem is semantic interoperability. Here, it is possible to evolve a common ontology based on user actions

13.4.1 Robotic Applications

Robots that develop communication systems among themselves Several experiments with robots have successfully demonstrated that shared communication systems could be negotiated between autonomous embodied agents (Steels and Kaplan, 1999a,b, 2002; Steels, 1999; Vogt, 2000; Kaplan, 2001). The communication systems of these robots are grounded in reality. This means that communication conventions get associated with aspects of the robots' sensorimotor environment. The robots have no direct access to the "meanings" used by the other robots, but they gradually bootstrap know-how for using communication conventions in order to have other robots performed particular actions. Some experiments showed that it is not even necessary to assume that robots share a prior repertoire of common concepts. Instead, they could build up their conceptual repertoire in a co-evolutionary process simultaneously with the construction of their communication system. These technologies permit new robotic applications where population of robots constructed shared communication systems without the need of a central coordinator.

- **Robots in unknown remote environments**. Technologies for bootstrapping shared communication systems are adapted to situations where a group of robots have to agree on a convention system to communicate in an unknown environment. These techniques are particularly interesting in cases where human intervention is difficult (e.g. a colony of planetary robots). In such cases, robots could develop their own ontology and associated communication conventions in order to manage to coordinate their actions.
- **Cooperating robots**. Collection of simple cooperating robots can potentially solve problems that a single robot cannot solve and/or lead to more robust and efficient solutions than systems based on a single robot. Indeed, collective robotics in gathering an increasing interest in the last few years (see also section by Floreano and Keller). Progresses in methods and techologies that might allow a team of robots to effectively coordinate and cooperate through indirect or direct form of communication might mave huge

application potentials (e.g. in tasks such us retrival of humans buried under collapsed buildings, exploration and monitoring of dangerous area, and many other activities that require coordination and cooperation).
- **Populations of heterogeneous robots**. Several techniques currently under development do not assume that robots have exactly the same sensorimotor apparatus or control architecture. This means that populations of heterogeneous robots (i.e. different models of autonomous robots) can in some cases still manage to agree on an efficient communication system to interact with one another.

Robots that develop a communication system with a user Another line of applications concerns robots that develop autonomously a communication system with a user. Several achievements were obtained with systems capable of learning complex associations between linguistic events (word, sentences) and perceptual ones (images, videos, tactile information, etc.) (e.g. Roy and Pentland, 2002; Siskind, 2001; Cangelosi and Parisi, 2004; Banard et al., 2003; Dominey, 2003; Steels and Baillie, 2003). However, complex issues arise when these techniques are applied to mobile autonomous robots (Steels and Kaplan, 2000b; Kaplan et al., 2002). It is clear that several challenges remain to be addressed in order to successfully build systems capable of developing all the necessary prerequisites enabling complex human-robot communication (see for instance Kaplan and Hafner (2004) for a survey of the challenges related to joint attention), but progress is made in these directions. Ethical and cultural issues must also be taken into account when designing new usages for these technologies (Kaplan, 2004). Two kinds of application targets can be foreseen for these techniques.

- **Service robot companions**. Service robot companion are designed to enhance and extend the individuals own ability to perform crucial tasks. Robotic aids should in general complement the abilities of the user and can, for instance, help people live better lives as they get older or help disabled individuals. It is crucial that such robots are easy to use and adaptive. An increasing number of systems permits the use of human speech for control. However, these systems are only reliable for rather specific domains. Efficient interactions with service robot companions are likely to need more adaptable techniques. Researches on robots capable of autonomously developing communication systems with a user offer new perspectives in such a context. Instead of the user adapting to a predefined set of conventional commands, human and robot could mutually converge on a shared set of communication conventions. Moreover, such communication systems will continuously be adapted in order to be as adjusted as possible to the individuals current needs.
- **Entertainment robots**. In contrast with robot companions, entertainment robots are not necessarily useful devices but are designed to be interesting for their own sake (Fujita et al., 2000; Kusahara, 2000; Kaplan, 2005). In order to stay interesting for a long time, entertainment robots must autonomously

develop and learn. Technologies for developing a shared communication system between the user and the robot are well adapted to maintain such a continuously renewed interest in the interaction (see (Kaplan, 2005; Dautenhahn, 1999) for a longer development of this argument).

13.4.2 Ubiquitous ECAgents

It is always possible to separate a software part of an ECAgent, in which adaptation and learning take place, from a hardware part, which remains the same. In particular, a robot can be seen as a *software agent* controlling a physical body. Therefore, using wireless network connections, a software agent can transfer itself between two physical bodies. The term *teleportation* is used when the bodies are identical (McIntyre et al., 1999). When the software agent is transfered between two non-identical bodies (e.g. a personal robot and PDA), the term *metamorphosis* can be used. This kind of transfer is intrinsically more complex and requires a set of specific technological innovations. Kaplan and Oudeyer have presented an algorithm for reusing object prototypes learned in one body, in another one with similar sensory spaces. This idea was illustrated with the example of a software agent that could engage in verbal interaction when embodied in an robot and reuse the grounded vocabulary it has learned when transfered to a more simple robotic body (Kaplan and Oudeyer, 2000). However, most of the technology for doing this kind of transfer in a more general context remains to be developped. A central question is how can complex skills such as navigation or grasping developped for one body be recruited and adapted to another one? The easiest thing is to rebuild a new categorization of the world each time the agent changes body. But this is inefficient. Solutions must be found for agents to adapt to the specificity of their new body, reusing as much as they can of what they previously developped. One most interesting experiment in this perspective has been conducted by Floreano and Mondada (1998). They study the possibility of running incremental cross-platform evolution. An agent which has evolved navigation techniques in a Khepera body is then transfered in a Koala (a larger robot), and adapts in order to continue to use its techniques. Another interesting issue is the possibility of doing part of the learning for a robot in a virtual world. There have been several isolated attempts to experiment on the transfer between real and virtual worlds. Michel (1998) showed how virtual robots can develop skills both in a realistic virtual world and in the body of Khepera robots. Several researchers have conducted experiments in simulated and real environment, and compared their results (Billard et al., 1999; Martinoli et al., 1999). These experiments show that part of the training of a robot could be done first in a simulated world, before being embodied in a physical robot. These technologies offer new interesting application perspectives.

- **Ubiquitous robotics**: Using teleportation and metamorphosis, software agent controlling ECDevices can manage to change body in order to find the most appropriate form for any given situation. A robot is not an easily transportable object compared to a PDA or a digital camera. Allowing software agents to “dock” into various kind of ECDevices permits long term interaction with human as a companion software agent can follow its owner even when he or she leaves home. From the point of the view of the agent’s development, the number of learning situations increases consequently. Agents can learn through a variety of real world situated interaction, or even embodied in a virtual character inside a video game.

– **Large-scale collective dynamics**: Teleportation technologies permit to consider the possibility of the emergence of collective dynamics resulting in the interaction between a large number of software agents. By interacting not only with humans, but also with one another, shared convention systems can emerge adapted to both human-robot and robot-robot interactions.

13.4.3 Emergent Semantics in peer-to-peer networks

The mechanisms being developed and studied in the ECAgent projects have the potential to solve an important problem in current distributed information technology, namely the exchange of information in peer-to-peer networks, more specifically the problem of semantic interoperability. Instead of imposing a universal pre-defined ontology over universally defined conceptual schemata, the techniques of emergent ontologies and languages potentially enable each agent to develop a repertoire of grounded categories and labels for these categories and negotiate their use and semantics with other agents. The communication system as well as its semantics is hence emergent and adaptive instead of predefined, leading to a Self-organisation Approach to Semantic Interoperability (SASI).

Semantic Interoperability An information system contains a collection of data and possibly a set of meta-data structured according to some conceptual schema. To enable user interaction, an information system typically allows a user to taxonomically structure the data according to her own named categories so that she can retrieve items through these names. Typical examples of information systems are: (1) The ‘favored’ web pages of a user organised in bookmark folders. The data consists of the URLs to the web pages and the taxonomy is the hierarchy of named folders that the user can browse through to retrieve a web page. The taxonomy implicitly defines a categorisation of web pages by the user. (2) A set of music files maintained by a user, organised as a series of named hierarchical playlists. (3) A set of images organised according to specific interests of the user, e.g. a series of medical images organised along pathologies, or a series of paintings organised along periods, genres, and painters. (4) A set of scientific papers organised along specific research themes.

The human user, further called the owner of the information system, controls her information system by adding data and imposing structure on the data in the form of taxonomies and giving names to the nodes in the taxonomy. Note that the categorisations of the user implied by these taxonomies are based on private cognitive processes which are not accessible to other users nor to information systems. For example, a user may decide to put all the songs she likes in one folder and the ones she does not like in another. This categorisation decision is completely subjective and can never be automated nor emulated by a machine.

We call the taxonomy created and maintained by the user the owner taxonomy. The names used in this taxonomy (which could be words or phrases) are owner names. The taxonomy implies a particular way of categorisation (known perhaps even only at an intuitive level by the owner) which is called the user ontology. The user ontology is implicitly implied in the taxonomy but otherwise not known.

A peer-to-peer information system consists of a collection of such information systems. Each information system is owned and maintained by a different user and assumed to operate independently of the others. The defining characteristic of peer-to-peer information systems is that they allow direct information exchange between peers without the need to go through a central server. Examples of peer-to-peer information systems are peer-to-peer music file sharing, such as Napster or Gnutella, that are already used by millions of people today. Similar sharing networks are growing for movies or game software. Also in the domain of scientific data or educational materials, there are growing networks of peer-to-peer shared systems (Nejdl et al., 2003).

In a peer-to-peer information system, the owner of one information system typically queries directly another information system in order to obtain additional data. The information system of the querying peer is called the client and the system providing information is called the server. For example, a web user may want to query the bookmark folders of another user in order to find web pages that may be of interest, a user may want to query the play lists of another user in order to find music that might be of interest to her, a user may want to query the image data base of another user in order to find images that relate to her own interests, or a user who is looking for papers that are relevant to one of the research topic she is investigating, might want to query the set of papers stored by another user.

In the spirit of peer-to-peer information systems, any node can behave as client or as server. Note that users communicate through the taxonomies with which each owner has organised her data.

There are two key problems in peer-to-peer information exchange. The first one is that the data and names used in the taxonomies of one peer (the client) are typically different from those used by another peer (the server) and so the client's owner cannot know how to formulate the query nor can the server's owner or the server itself know how to respond if the query is not formulated according to its taxonomy. This is a real problem for users of currently operational peer-to-peer systems. For example, in a music file sharing network, users must try to guess the titles of data and the meaning of names given to the folders and subfolders.

The second problem is that the conceptual schemata used for storing data and meta-data in each information system may be very different, particularly if the meta-data is itself open-ended. Even a simple incompatibility such as usage of different languages can be a problem. For example, the client may have a meta-datum 'country(Belgium)' whereas the server may have 'pays(Belgique)'.

Without semantic knowledge, information systems cannot know how the two meta-data map onto each other, and so a client cannot simply formulate a query for a server using his own meta-data.

Both problems are instances of the so called semantic interoperability problem.

Approaches One solution to semantic interoperability is to standardise. The different users of a peer-to-peer network could all agree a priori to use the same taxonomies to structure their data and to use the same conceptual schemata for their data and meta-data. The owner names in the taxonomies can then act as a shared communication protocol between peers. For example, all users could agree to use the taxonomies of Yahoo for organising their data, and adopt the names used by Yahoo (possibly with translations into different languages).

Unfortunately such a standardisation approach is unlikely to work for a truely open-ended peer-to-peer network in volatile domains like music file sharing, medical images or scientific papers. As new topics and new kinds of data come up all the time, styles shift, and interests of users diverge. There are legacy systems which should also be enabled to participate in a peer-to-peer network. It is very hard to capture all this once and for all in a static ontology.

Alternatively, it is possible that each peer has its own local taxonomy, and its own conceptual schema but that these are translated into a global ontology and conceptual schema which is used for querying and information exchange and thus exacts as an Interlingua between peers. The translations could be based on defining as much as possible the semantics of the names in the taxonomies. For example, if a user has a sub-folder in his music file system with songs by the Beatles, then the semantics of the implied category is translated in a query over meta-data: 'performed-by(TheBeatles)'. This query can then be used (possibly after translation by a mediator) into a query over the meta-data of a peer.

This is the approach currently being explored by the Semantic Web initiative (Berners-Lee et al., 2001), w.r.t. web information systems, and, more generally, by 'universal' ontologies such as advocated by CYC or Wordnet (Lenat et al., 1995). It has lead to extensive efforts to develop common ontologies, support systems for defining these ontologies, ways for mapping local schemas into global schemas, and mechanisms to use ontologies in information retrieval, i.e. for mapping categories to data (Davies et al., 2002).

However a consensus is growing that this approach has several major drawbacks (Aberer et al., 2004; Steels, 1997a) as well:

1. The semantic web which relies on universal ontologies just pushes the problem of semantic interoperability to another level. It still requires standardisation based on universal ontologies. It is hard to imagine that a world-wide consensus is reachable and enforcable in every domain of human activity for which information systems are currently in use. Even in restricted domains this is hard because of an increasingly interconnected global world.
2. Human activity and the information systems built for them are open systems. They cannot be defined once and for all but must be adapted to new needs.

3. Peer-to-peer information systems are distributed systems. There is no central control point and so it is not possible to control them centrally.
4. Many information systems already exist and ways should be found to enable their participation in peer-to-peer networks.

An alternative approach to semantic interoperability is to extend information systems with components so that peers can develop and negotiate their own communication protocols in interaction with the data world and the world of human users. So the agents autonomously create an Interlingua which they can each locally interpret. Just as in human natural languages, the consensus will be for ever emergent, adaptive and local. This approach is one of the ways to achieve *emergent semantics* and is the one that we see as major application area for ECAgents.

The technical solutions that rely on techniques drawn from recent work on language games for robot-robot and robot-human communication (Steels, 1998b), (Steels, 2003a), as further developed in the ECAgents project, have to be expamaknded and changed to make them applicable to the current task. Earlier work in this area has been reported in (Steels, 1997a) and (Avesani and Agostini, 2003) but large-scale application has so far not yet been attempted.

In this approach, Semantic interoperability is seen as a coordination problem between the world, information systems, and human users. A particular kind of 'semiotic dynamics' is defined so that both the labels used in peer-to-peer communication and the categories the agents use to interpret these labels become aligned as a side-effect of peer-to-peer information exchange. The labels used in information exchange as well as the semantics of the labels is emergent and the conceptual schemata used for the meta-data in each peer are local and extensable. We note that the Interlingua emerging through agent interactions will never be static and may be locally specialised among a group of peers. The categories defining the ontology of each agent are defined purely in terms of local meta-data and so they are not uniform either.

Challenges At the moment the most basic technologies for achieving peer-to-peer information exchange based on emergent semantics are available, at least for certain domains. For example, for the domain of music, there is a high interest in p2p exchange (so there is a potential 'market'), many algorithms exist to ground descriptors into music sources and even for developing new categorisations in order to discriminate based on evolutionary algorithms (Zils and Pachet, 2004), and there is a high volume of potential data to be exchanged. Moreover this domain is constantly 'on the move' so that it is difficult to apply the classical technique of pre-defined ontologies.

So the main challenge to be taken up in the ECAgents project is to construct a viable demonstrator of this technology. This requires that all parts are technically pulled together in a single platform (e.g. networked mobile PC with music players) and that a user community is identified that is willing to try out the p2p exchange in this domain.

13.5 Conclusion and challenges

Developing application concepts based on novel technologies and concepts is difficult, but likelihood of success can be greatly increased by applying methods from other fields, as well as making use of insights gained from the study of existing products. The challenge for the remainder of the project is to develop new application concepts in a way that makes them not just interesting from a research point of view, but also viable product concepts that can have a genuine interest for consumers. For this, it is necessary to apply the methods outlined in section 2, and the lessons learned in section 3, to potential applications such as those outlined in section 4. By doing this, we will achieve a true integration of research advances and product development. This will mean that results in the project have a chance of impacting not just the research world but society as a whole, through the introduction of services and applications that would otherwise not have existed.

White Paper Concluding Chapter

Luc Steels

SONY Computer Science Laboratory - Paris

This chapter summarizes the common theoretical and methodological foundations of the project and highlights the main challenges the ECAgents project is focusing on.

14.1 Sociality and Communication

It is clear to anyone who has seriously studied communication from a scientific point of view, that this is a very broad research subject in which contributions from many different disciplines are relevant. The great value of the ECAgents project is that it has managed to bring together several of the most relevant disciplines, in particular the biological and cognitive science perspective on the one hand, and the computer science and information technology perspective on the other. The first perspective puts forward natural models, important issues, and theoretical tools, the second one operational models, implementation technologies, and application domains.

Although there are many ways to structure the domain of communication, the ECAgents project has adopted a biological perspective (chapter 2), making a distinction between three types of communication, depending on the characteristics of the agents involved:

- Communication among selfish agents. This is the most common situation in animal communication. Each agent only cares about itself in a ruthless attempt to propagate its genes. A consequence is that communication systems are entirely automatic and genetically determined and/or highly costly to avoid cheating. They are (analog) signaling systems rather than (discrete) symbol-based communication systems.
- Communication among eusocial agents. This is also found in animal communication but only in eusocial groups such as social insects where individuals largely share the same genes (kin selection). Communication systems arising under such conditions also appear to be automatic and genetically evolved, and they also qualify as signaling rather than symbol-based systems.
- Communication among social agents. This is the case for human beings, and perhaps the only case in nature of symbol-based communication (chapter 7). Although agents are not necessarily strongly genetically related, they have nevertheless been able to transcend the Darwinian world to become fundamentally cooperative. Only under these conditions do we see the emergence of symbol-based communication systems with the complexity of human natural languages.

There are many biological questions on how eusociality can evolve within Darwinian assumptions and also how human-like social trusting agents can arise under the same assumptions. These questions are fascinating but not central to the ECAgents project. The ECAgents project rather concentrates on the following two questions:

1. Given a particular type of sociality among agents (selfish/eusocial/social), how can a communication system emerge and continue to function across generations. Depending on the type of sociality we will get very different approaches. (1) and (2) will rely mostly on genetic evolution of the communication system and is studied in the project under the heading of animal/animat-like communication (surveyed in part I of the white paper) and (3) on cultural evolution, studied in the project under the heading of human-like communication (surveyed in part II of the white paper).
2. What is the most appropriate communication system for a particular application domain of IT, specifically how can the communication system remain viable against operating and adversary conditions. As Hammerstein (see chapter 6 on Game-theoretic approaches) points out, often IT-systems assume social trusting conditions, whereas they do not always completely hold as Darwinian conditions may resurface in the human world. An example is spam where a benign system based on trust is exploited by spammers for their own self-interest.

The remaining sections now provide more specific research topics which the ECAgents project is tackling for each of these questions.

14.2 The Emergence of Communication Systems

The first key question addressed in the ECAgents project is how it is at all possible to see the emergence of a communication system in a group of embodied agents. If we are ever going to translate this into technology, we need a mechanistic theory, in other words we need a specification of the kinds of mechanisms that agents need in order to see the emergence of a communication system. Proof that the mechanisms work will have to come from computer simulations and/or robotic experiments. The latter is particularly important because the project focuses on embodied communication.

There are three different research streams, each time bringing together biologists/cognitive scientists and computer scientists/AI experts.

Signaling Systems among Selfish Agents

Signalling systems in selfish agents based on genetic techniques are the main focus of attention of the Budapest (chapter 2) and Berlin group (see chapter 5) and the Rome group (see chapter 3). The biologists focus on theoretical studies in animal communication (such as the need to have costly signaling) and the engineers focus on experiments in which communication systems might evolve under selfish agent conditions. As pointed out by Nolfi (chapter 3) and Hammerstein (chapter 5), the biggest challenge here is to counteract the lack of adaptive benefit for the signaler and the conflict between individual and collective interests. Research is focusing on mechanisms and/or experimental conditions under which this challenge might be overcome.

Signaling Systems in Eusocial Agents

Signaling systems in eusocial agents typically focus on helping large groups of genetically strongly related agents to cooperate, for example in path formation or nest building. The ECAgents research stream that takes this line of investigation contains the Brussels partners (chapter 6), the Lausanne partner (chapter 4), and the Rome partner (chapter 3). The work has focused on the one hand on the study of natural systems and the development of mathematical tools for analyzing under which conditions communication arises and can be sustained. The communication systems in eusocial agents are almost all based on self-organization based on a positive feedback loop that enforces random fluctuations, so a key challenge taken up in the ECAgents project is to develop workable mathematical tools to analyze such systems. On the other hand, work in this line has focused on building robotic systems that work along the same principles, the biggest challenge here is to find the precise mechanisms that cause self-organization and to see how a viable communication system could evolve by evolutionary programming.

Symbol-based Communication system in Social Agents

When agents are social, as in the case of human groups, the emergence of symbol-based communication systems of high complexity becomes possible but it is still a deep question how natural language like communication systems then might arise. These questions are investigated in particular by the Budapest partner from the (neuro)biological side (see chapter 11) and the Paris partner (see chapter 10) from the side of computational and robotic models. There are three main challenges which are all attacked in the ECAgents project.

As Tomasello and others have argued, the first prerequisite for symbol-based communication is social cognition, which itself rests on joint attention and theory of mind, i.e. the capacity to model the other, e.g. to see the situation from the perspective of another agent. So the project is focusing on mechanistic models of both processes, trying to come up with operational models how these are possible and why they would be used by social agents (see chapter 8). Once social cognition is in place, various mechanisms come into play for the self-organization of a symbol-based communication systems. The ECAgents project is 'biting the bullet' and developing operational models for the self-organization of communication media such as speech sounds (chapter 9), lexicons, the co-evolution of categories and words, and the emergence of grammar (chapter 10). The third prerequisite for symbol-based communication is that there is a neurobiological substrate that can sustain the mechanisms thought to be crucial for language.

Partners are taking up this challenge as well focusing on the recruitment theory (or amoeba hypothesis) which puts a large emphasis on the brain's plasticity to dynamically configure a language faculty under the pressure of communication (chapter 11).

14.3 Viability conditions for emergent communication systems

When communication systems start to emerge and propagate in simulations it is often not clear whether systems are viable and under what constraining conditions. Even though there may be a large number of successful simulations, it is still possible that results are an artefact of specific parameter settings, or that they do not scale up to real world conditions. This is why the ECAgents project puts a lot of emphasis on the challenge to develop tools for the analysis of emergent communication systems. These tools will have to come from Complex Systems Science, in particular game theory (chapter 5), dynamical systems analysis (chapter 6), mean field approaches (chapter 12) and network analysis (chapter 12). They should address such questions as: are there any powerlaws relating population sizes with convergence.

14.4 Application areas

The ECAgents project is contributing to IT innovation by constantly seeking potential spin-offs from the fundamental scientific understanding that is reached by the project partners. This research into potential applications is particularly carried out by the Goteborg and Paris partners (chapter 13). We deliberately do not restrict ourselves to applications defined a priori, because then the project would no longer be a research project but become a development project in which much more technological focus is needed and serendipity becomes excluded. As the research in communication is broad, there is also a broad set of imaginable possible application domains (briefly surveyed in chapter 13) and early explorations have already lead to three clear spin-offs. The first one is in the domain of social tagging and emergent semantics. Collaboration between two ECAgents partners (La Sapienza and Sony CSL) has lead to a new FET project TAGora that applies technologies and analysis methods which arose in the context of symbol-based communication systems to understand and orchestrate the emergent semantics of social tagging sites. Collaboration between Viktoria Institute and Sony CSL has lead to novel ideas for entertainment robotics. Finally collaboration between Barcelona and Sony CSL has lead to another NEST project where insights from the study of emergent communication and language networks are applied to bio-informatics. In addition to these concrete actions there have also been a variety of joined papers and ongoing discussions and workshops which attest to the high level of scientific interchange between the partners.

14.5 Cooperation between partners

It is obvious that the ECAgents project is not structured towards the common goal of building a single software system, but rather that different partners are complementary and cover different areas of the research landscape, each addressing challenges of great relevance to the scientific communities studying communication from different angles. Cross-fertilization takes place due to an investigation of mixed systems, for example where selfish agents use some form of social learning (see chapter 3) or where self-organization which plays a primary role in eusocial communication (see chapter 6) is also the foundation for the self-organization of symbol-based communication (see chapter 10). Also because there is expertise about the full range of communication systems, it becomes possible to investigate where one type of approach is more adapted than another one for IT or to try and introduce new forms of communication inspired by living systems. In fact, without the exercise of writing the white paper, and the progress already achieved in the project it would not have been possible to write this particular chapter.

Bibliography

K. Aberer, P. Cudré-Mauroux, A. M. Ouksel, T. Catarci, M.-S. Hacid, A. Illarramendi, V. Kashyap, M. Mecella, E. Mena, E. J. Neuhold, O. D. Troyer, T. Risse, M. Scannapieco, F. Saltor, L. D. Santis, S. Spaccapietra, S. Staab, and R. Studer. Emergent semantics principles and issues. In Y. Lee, J. Li, K.-Y. Whang, and D. Lee, editors, *Database Systems for Advances Applications, 9th International Conference, DASFAA 2004, Jeju Island, Korea, March 17-19, 2004, Proceedings*, volume 2973 of *Lecture Notes in Computer Science*, pages 25–38. Springer, 2004.

D. Ackley and M. Littman. Interaction between learning and evolution. In *Proceedings of the Second Conference on Artificial Life*. Addison-Wesley, 1991.

R. Albert and A.-L. Barabasi. Statistical mechanics of complex networks. *CoRR*, cond-mat/0106096, 2001.

A. Alissandrakis, C. L. Nehaniv, and K. Dautenhahn. Learning how to do things with imitation. In *AAAI Fall Symposium on Learning How to Do Things*, pages 1–6. American Association for Artificial Intelligence, November 2000.

J. Ame, C. Rivault, and J. Deneubourg. Cockroach aggregation based on strain odour recognition. *Animal Behaviour*, 68:793–801, 2004.

C. Anderson. Self-organization in relation to several similar concepts: Are the boundaries to self-organization indistinct? *Biological Bulletin*, 202:247–255, 2002.

C. Anderson and D. McShea. Individual versus social complexity, with particular reference to ant colonies. *Biological Reviews*, 76(211-237), 2001.

C. Anderson and F. L. W. Ratnieks. Task partitioning in insect societies. I. Effect of colony size on queueing delay and colony ergonomic efficiency. *American Naturalist*, 154:521–535, 1999.

P. Andry, P. Gaussier, S. Moga, J. Banquet, and J. Nadel. Learning and communication in imitation: an autonomous robot perspective. *IEEE Transaction on Systems, Man and Cybernetics, Part A : Systems and Humans*, 31(5):431–444, September 2001.

S. Aron, J.-L. Deneubourg, S. Goss, and J. M. Pasteels. Functional self-organisation illustrated by inter-nest traffic in the argentine ant iridomyrex humilis. In W. Alt and G. Hoffman, editors, *Biological Motion*, pages 533–547. Springer-Verlag, 1990.

W. Asaad, G. Rainer, and E. Miller. Task-specific neural activity in the primate prefrontal cortex. *Journal of Neurophysiology*, 84:451–459, 2000.

S. Asch. *Social Psychology.* Prentice-Hall, 1952.

P. Avesani and A. Agostini. A peer-to-peer advertising game. In M. Orlowksa, M. Papazoglou, S. Weerawarana, and J. Yang, editors, *First International Conference on Service Oriented Computing (ICSOC-03)*, pages 28–42, Berlin Heidelberg, 2003. Springer-Verlag LNCS 2910. Electronically available at `http://sra.itc.it/al/htm/pub.htm`.

J. Baird and D. Baldwin. Making sense of human behavior: Action parsing and intentional inference. In F. Malle, L. Moses, and D. Baldwin, editors, *Intentions and Intentionality*, chapter 9. MIT Press, 2001.

K. Balakrishnan and V. Honavar. Evolutionary design of neural architectures: Preliminary taxonomy and guide to literature. Tr #95-01, Dept. of Computer Science, Iowa State U., 1995.

T. R. Balch and L. E. Parker. Guest editorial. *Auton. Robots*, 8(3):207–208, 2000.

G. Baldassarre, S. Nolfi, and D. Parisi. Evolving mobile robots able to display collective behaviors. *Artificial Life*, 9(3):255–268, 2003.

D. Baldwin, E. Markram, and R. Melartin. Infants' ability to draw inferences about nonobvious object properties: Evidence from exploratory play. *Child Development*, 64:711–728, 1993.

J. Baldwin. A new factor in evolution. *American Naturalist*, 30:441–451, 1896.

P. Ball. *the Self-Made Tapestry-Pattern formation in nature*. Oxford University Press, 1999.

K. Banard, P. Duygulu, D. Forsyth, N. de Freitas, D. Blei, and M. Jordan. Matching words and pictures. *Journal of Machine Learning Research*, 3:1107–1135, 2003.

A. Barabási and R. Albert. Emergence of Scaling in Random Networks. *Science*, 286(5439):509–512, 1999.

J. Barkow, L. Comsides, and J. Tooby. *The adapted mind*. Oxford University Press, 1992.

S. Baron-Cohen. *Mindblindness: an essay on autism and theory of mind*. MIT Press, Boston, MA, USA, 1997.

J. Batali. Small signaling systems can evolve in the absence of benefit to the information sender. unpublished document, 1995.

R. Beckers, J.-L. Deneubourg, and S. Goss. Trails and u-turns in the selection of the shortest path by the ant lasius niger. *J. Theor. Biol.*, 159:397–415, 1992.

M. Beekman, D. J. T. Sumpter, and F. L. W. Ratnieks. A phase transition between disordered and ordered foraging in pharoah's ants. In *Proceedings of the National Academy of Sciences*, volume 98, pages 9703–9706, 2001.

R. D. Beer. The dynamics of active categorical perception in an evolved model agent. *Adaptive Behavior*, 11(4):209–243, 2003.

C. Bergström and M. Lachmann. Signaling among relatives. iii. talk is cheap. In *Proceedings of the National Academy of Sciences USA 95*, pages 5100–5105, 1998.

T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. *Scientific American*, May 2001.

B. Bertenthal. Origins and early development of perception, action and representation. *Annual review of psychology*, 47:431–59, 1996.

S. N. Beshers and J. Fewell. Models of division of labor in social insects. *Annual Review of Entomology*, 46:413–440, 2001.

D. Bickerton. *Language and Species*. University of Chicago Press, Chicago, IL, USA, 1992.

A. Billard. Learning motor skills by imitation: a biologically inspired robotic model. *Cybernetics and Systems*, 32(1):155–193, 2000.

A. Billard and K. Dautenhahn. Experiments in learning by imitation - grounding and use of communication in robotic agents. *Adaptive Behavior*, 7(3/4):415–438, 1999.

A. Billard, A. J. Ispeert, and A. Martinoli. Adaptive exploration of a dynamic environment by a group of communicating robots. In D. Floreano, J.-D. Nicoud, and F. Mondada, editors, *Advances in Artificial Life (ECAL 99)*, Lecture Notes in Artificial Intelligence 1674, pages 596–605, Berlin, 1999. Springer-Verlag.

B. Bollobás. *Random graphs*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London, 1985. ISBN 0-12-111755-3; 0-12-111756-1.

E. Bonabeau, J. Deneubourg, and G. Theraulaz. Within brood competition and the optimal partitioning of parental investment. *American Naturalist*, 152: 419–427, 1998.

E. Bonabeau, M. Dorigo, and G. Theraulaz. *Swarm intelligence: from natural to artificial systems*. Oxford University Press, Inc., 1999. ISBN 0-19-513159-2.

E. Bonabeau, M. Dorigo, and G. Theraulaz. Inspiration for optimization from social insect behaviour. *NATURE*, 406:39–42, 2000.

J. Bongard and R. Pfeifer. Evolving complete agents using artificial ontogeny. In F. Hara and R. Pfeifer, editors, *Morpho-functional Machines: The New Species (Designing Embodied Intelligence)*. Springer Verlag, 2003.

J. Bouchaud and A. Georges. Anomalous diffusion in disordered media: statistical mechanics, models and applications. *Phys. Rep. 195*, 1990.

J.-P. Bouchaud, I. Giardina, and M. Mzard. On a universal mechanism for long-range volatility correlations. *Quantitative Finance*, (1):112–126, 2001.

A. Bourke and N. Franks. *Social Evolution in ants.* Princeton University Press., 1996.

A. F. G. Bourke. *Sociality and kin selection in insects*, pages 203–227. Blackwell Scientific Press, Oxford, 1991.

A. F. G. Bourke. Colony size, social complexity and reproductive conflict in social insects. *Journal of Evolutionary Biology*, 12:245–257, 1999.

R. Boyd and P. Richerson. *Culture and the evolutionary process.* Chicago University Press, Chicago, 1985.

C. Breazeal. *Designing sociable robots.* Bradford book - MIT Press, 2002.

E. J. Briscoe. *Coevolution of the language faculty and language(s) with decorrelated encodings*, chapter 14. Oxford: Oxford University Press, 2005.

C. Browman and L. Goldstein. Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de la Communication Parle*, 5:25–34, 2000.

C. Browman and L. Goldstein. Towards an articulatory phonology. *Phonology Yearbook*, 3:219–252, 1986.

J. Bryant. Scalable construction-based parsing and semantic analysis. In *Proceedings of the Second International Workshop on Scalable Natural Language Understanding.*, pages 74–78.

M. Buchenau and J. F. Suri. Experience prototyping. In *Symposium on Designing Interactive Systems*, pages 424–433, 2000.

E. Bufill and E. Carbonell. Are symbolic behaviour and neuroplasticity an example of gene-culture coevolution? *Revista de Neurologia*, 39(1):48, 2004.

R. Burling. Primate calls, human language, and nonverbal communication. *Current Anthropology*, 34:25–53, 1993.

G. Butterworth. Origins of mind in perception and action. In C. Moore and P. Dunham, editors, *Joint attention: its origins and role in development*, pages 29–40. Lawrence Erlbaum Associates, 1995.

G. Butterworth and N. Jarrett. What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy. *British Journal of Developmental Psychology*, 9:55–72, 1991.

G. E. Butterworth and E. Cochran. Towards a mechanism of joint visual attention in human infancy. *International Journal of Behavioural Development*, 3: 253–272, 1980.

M. Caceres, J. Lachuer, M. A. Zapala, J. C. Redmond, L. Kudo, D. H. Geschwind, D. J. Lockhart, T. M. Preuss, and C. Barlow. Elevated gene expression levels distinguish human from non-human primate brains. *Proc. Natl. Acad. Sci. USA*, 100, 2003.

R. Calabretta, A. Di Ferdinando, G. Wagner, and D. Parisi. What does it take to evolve behaviorally complex organisms? *BioSystems*, 69:245–262, 2003.

S. Camazine, P. Visscher, J. Finley, and R. Vetter. House-hunting by honey bee swarms: collective decisions and individual behaviours. *Insectes Sociaux*, 46: 348–360, 1999.

S. Camazine, J.-L. Deneubourg, N. Franks, E. Bonabeau, G. Theraulaz, and J. Sneyd. *Self-Organization in Biological Systems*. Princeton, 2001.

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

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

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

Y. Cao, A. Fukuaga, and A. Kahng. Cooperative mobile robotics: Antecedent and directions. *Autonomous Robots*, 4:7–27, 1997.

E. Carlson and J. Triesch. A computational model of the emergence of gaze following. In *Proceedings of the 8th Neural Computation Workshop (NCPW8)*, 2003.

M. Carpenter, N. Akhtar, and M. Tomasello. Fourteen through eighteen month old infants differentially imitate intentional and accidental actions. *Infant behavior and development*, 21:315–330, 1998.

C. Castelfranchi. When doing is saying - the theory of behavioral implicit communication, in preparation.

A. Cavagna. Irrelevance of memory in the minority game. *Phys. Rev. E*, (59), 1999.

L. L. Cavalli-Sforza and M. Feldman. *Cultural Transmission and Evolution: a Quantitative Approach*. Princeton University Press, Princeton, NJ, USA, 1981.

D. Challet and Y.-C. Zhang. On the minority game: Analytical and numerical studies. *Physica A*, 1998.

D. Challet, M. Marsili, and R. Zecchina. Statistical mechanics of systems with heterogenous agents: Minority game. *Phys. Rev. Lett.*, (84):1824, 2000.

N. C. Chang and T. V. Maia. Grounded learning of grammatical constructions. In *2001 AAAI Spring Symposium on Learning Grounded Representations*, 2001.

J.-P. Changeux. *L'Ĥomme Neuronal*. Librairie Artheme Fayard, Paris, 1983.

G. Chechik, I. Meilijson, and E. Ruppin. Synaptic pruning in development: A computational account. *Neural Computation*, 10(7):1759–1777, 1998.

D. Cheney and R. M. Seyfarth. *How monkeys see the world*. University of Chicago Press, 1990.

N. Chomsky. *Aspects of the Theory of Syntax*. The MIT Press, Cambridge, MA, 1965.

N. Chomsky. *The minimalist program*. MIT Press, Boston, MA, USA, 1995.

N. Chomsky and O. Halle. *The Sound Pattern of English*. Harper Row, New York, 1968.

M. H. Christiansen and S. Kirby. Language evolution: Consensus and controversies. *Trends in Cognitive Sciences*, 7(7):300–307, 7 2003.

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

T. H. Clutton-Brock and S. D. Albon. The roaring of red deer and the evolution of honest advertisement, 1979.

S. Coradeschi and A. Saffiotti. An introduction to the anchoring problem. *Robotics and Autonomous Systems*, 43:85–96, 2003.

A. C. Courville, N. D. Daw, G. J. Gordon, and D. S. Touretzky. Model uncertainty in classical conditioning. In S. Thrun, L. K. Saul, and B. Schölkopf, editors, *Advances in Neural Information Processing Systems 16 [Neural Information Processing Systems, NIPS 2003, December 8-13, 2003, Vancouver and Whistler, British Columbia, Canada]*. MIT Press, 2004.

A. C. Courville, N. D. Daw, and D. S. Touretzky. Similarity and discrimination in classical conditioning: A latent variable account. In L. K. Saul, Y. Weiss, and L. Bottou, editors, *Advances in Neural Information Processing Systems 17*. MIT Press, Cambridge, MA, 2005.

I. Couzin and J. Krause. Self-organization and collective behavior in vertebrates. *Advances in the Study of Behaviour*, 32:1–75, 2003.

S. J. Cox, T. J. Sluckin, and J. Steele. Group size, memory, and interaction rate in the evolution of cooperation. *Current Anthropology*, pages 369–377, 1999.

D. Croon, S. Nolfi, and E. Postma. Toward pro-active embodied agents: On the importance of neural mechanisms suitable to process information in time. In D. Braha, A. Minai, and Y. Bar-Yam, editors, *Complex Engineering Systems*. MIT Press, in press.

R. Crozier and P. Pamilo. *Evolution of Social Insect Colonies: Sex Allocation and Kin Selection*. Oxford University Press, 1996.

K. A. Dahmen, D. R. Nelson, and N. M. Shnerb, 1999. Population Dynamics and Non-Hermitian Localization. Condensed Matter, abstract cond-mat/9903276.

K. Dautenhahn. Embodiment and interaction in socially intelligent life-like agents. In C. Nehaniv, editor, *Computation for metaphors, analogy and agent*, volume 1562 of *Springer Lecture Notes in Artificial Intelligence*, pages 102–142. Springer, 1999.

K. Dautenhahn and C. Nehaniv. *Imitation in animals and artifacts*. MIT Press, 2002.

J. Davies, F. van Harmelen, and D. Fensel, editors. *Towards the Semantic Web: Ontology-driven Knowledge Management*. John Wiley & Sons, Inc., 2002. ISBN 0470848677.

M. S. Dawkins and T. Guilford. Design of an intentional signal in the bluehead wrasse (thalassoma bifasciatum). In *Proceedings of the Royal Society of London*, volume B 257, pages 123–128, 1994.

R. Dawkins. *The Blind Watchmaker: The Evidence of Evolution Reveals a Universe Without Design*. W. W. Norton & Company, 1986.

R. Dawkins and K. J. R. Animal signals: information or manipulation? In J. R. Krebs and N. B. Davies, editors, *Behavioral ecology: An Evolutionary Approach*, pages 282–309. Sinauer Associates, Massachusetts, 1978.

B. de Boer. *The Origins of Vowel Systems*. Oxford University Press, 2001.

T. Deacon. *The symbolic species*. W. Norton and Co, New York, 1997a.

T. W. Deacon. *The Symbolic Species: The Co-evolution of Language and the Brain*. W.W. Norton, 1997b.

J. DeFelipe, L. Alonso-Nanclares, and J. Arellano. Microstructure of the neocortex: Comparative aspects. *Neurocytol.*, 31:299–316, 2002.

D. Denaro and D. Parisi. Cultural evolution in a population of neural networks. In M. Marinaro and R. Tagliaferri, editors, *Proceedings of Neural Nets Wirn-9*, pages 100–111. Springer, 1996.

J. Deneubourg, J. Pasteels, and J. Verhaeghe. Probabilistic behaviour in ants: a strategy of errors? *Journal of Theoretical Biology*, 105:259–271, 1983.

J.-L. Deneubourg and S. Goss. Collective patterns and decision-making. In *Ethology, Ecology and Evolution 1*, pages 295–311, 1989.

J. L. Deneubourg, S. Goss, J. M. Pasteels, D. Fresneau, and J. P. Lachaud. Self-organization mechanisms in ant societies, II: Learning in foraging and division of labor. *From Individual to Collective Behavior in Social Insects*, 54:177–196, 1987.

S. Deneve. Spiking neural networks implementing bayesian belief networks. In *Advances in Neural Information Processing Systems 17*. MIT Press, Cambridge, MA, 2004. in press.

S. Deneve, P. Latham, and A. Pouget. Efficient computation and cue integration with noisy population codes. *Nat. Neurosci.*, 4:826–31, 2001.

S. Desrochers, P. Morisette, and M. Ricard. Two perspectives on pointing in infancy. In C. Moore and P. Dunham, editors, *Joint Attention: its origins and role in development*, pages 85–101. Lawrence Erlbaum Associates, 1995.

J.-L. Dessalles. Altruism, status and the origin of relevance. In J. Hurford, C. Knight, and M. Studdert-Kennedy, editors, *Approaches to the Evolution of Language: Social and Cognitive bases*. Edinburgh University Press, Edinburgh, 1998.

C. Detrain and J. Deneubourg. Complexity of environment and parsimony of decision rules in insect societies. *The Biological Bulletin*, 202:268–274, 2002.

C. Detrain and J. Deneubourg. Scavenging by pheidole pallidula: a key for understanding decision-making systems in ants. *Animal Behaviour*, 53:537–547, 1997.

C. Detrain, J. Deneubourg, S. Goss, and Y. Quinet. Dynamics of collective exploration in the ant pheidole pallidula. *Psyche*, 98:21–31, 1991.

C. Detrain, J. Deneubourg, and J. Pasteels. Decision-making in foraging by social insects. In C. Detrain, J. Deneubourg, and J. Pasteels, editors, *Information processing in social insects*, pages 331–354. Birkhauser Verlag, Basel, 1999.

C. Detrain, C. Natan, and J. Deneubourg. The influence of the physical environment on the self-organised foraging patterns of ants. *Naturwissenschaften*, 88:171–174, 2001.

A. Di Ferdinando and D. Parisi. Internal representations of sensory input reflect the motor output with which organisms respond to the input. In A. Carsetti, editor, *Seeing and thinking*. Kluwer, 2004, in press.

E. A. Di Paolo. Behavioral coordination, structural congruence and entrainment in a simulation of acoustically coupled agents. *Adaptive Behavior*, 8(1):25–46, 2000.

E. A. Di Paolo. An investigation into the evolution of communication. *Adaptive Behavior*, 6(2):285–324, 1998.

F. Dignum and R. Conte. Intentional agents and goal formation. *Intelligent Agents IV: Agent Theories, Architectures, and Languages*, 1998.

R. Dixon. A note on Dyirbal ergativity. In *Chicago Linguistic Society*, volume 15, pages 90–91. 1979.

J. P. Djajadiningrat, W. W. Gaver, and J. W. Fres. Interaction relabelling and extreme characters: methods for exploring aesthetic interactions. In *DIS '00: Proceedings of the conference on Designing interactive systems*, pages 66–71. ACM Press, 2000. ISBN 1-58113-219-0. doi: http://doi.acm.org/10.1145/

347642.347664.

P. Dominey. Learning grammatical constructions from narrated video events for human-robot interaction. In *Proceedings of the IEEE humanoid robotics conference*, Karlsruhe, Germany, 2003.

M. Dorigo and T. Stützle. *Ant Colony Optimization*. MIT Press, 2004.

R. Dunbar. *Theory of mind and the evolution of language*. Cambridge University Press, Cambridge, 1998.

A. Dussutour, V. Fourcassi, D. Helbing, and J. Deneubourg. Optimal traffic organization in ants under crowded conditions. *Nature*, 428:70–73, 2004.

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

J. Elman, E. Bates, M. Johnson, A. Karmiloff-Smith, D. Parisi, and K. P. Plunkett. *Rethinking Innateness*. MIT Press, Cambridge, 1996.

G. Elston, R. Benavides-Piccione, and J. DeFelipe. The pyramidal cell in cognition: A comparative study in human and monkey. *The Journal of Neuroscience*, 21, 2001.

W. Enard, M. Przeworski, S. E. Fisher, C. S. L. Lai, V. Wiebe, T. Kitano, A. P. Monaco, and S. Paabo. Molecular evolution of foxp2, a gene involved in speech and language. *Nature*, 418:869–872, 2002.

M. Enquist and O. Leimar. The evolution of cooperation in mobile organisms. *Animal Behaviour*, 45:747–757, 1993.

P. W. Ewald. *Evolution of Infectious Disease*. Oxford University Press, New York, 1994.

J. D. Farmer. Physicists attempt to scale the ivory towers of finance. *Computing in Science and Engg.*, 1(6):26–39, 1999.

R. Ferrer and R. Sole. Least effort and the origins of scaling in human language. *Proceeding of National Academy of Science USA*, 100:788–791, 2003.

J. H. Fewell and R. Page. The emergence of division of labour in forced associations of normally solitary ant queens. *Evol. Ecol. Res.*, 1:537–548, 1999.

J. Fiser and R. Aslin. Statistical learning of new visual feature combinations by infants. In *Proc. Natl. Acad. Sci. USA. 99*, 2002.

W. Fitch and M. Hauser. Computational constraints on syntactic processing in nonhuman primates. *Science*, 303:377–380, 2004.

D. Floreano and F. Mondada. Evolutionary neurocontrollers for autonomous mobile robots. *Neural Networks*, 11:1461–1478, 1998.

D. Floreano and J. Urzelai. Evolution of plastic control networks. *Autonomous Robots*, 11:311–317, 1998.

M. Fontan and M. Mataric. Territorial multi-robot task division. *IEEE Transactions of Robotics and Automation*, 14(5), 1998.

S. Forrest, A. Somayaji, and D. Ackley. Building diverse computer systems. In *The Sixth Workshop on Hot Topics in Operating Systems: May 5–6, 1997, Cape Cod, Massachusetts*, pages 67–72, 1997.

D. Frye. The origins of intention in infancy. In D. Frye and C. Moore, editors, *Children's theories of mind*, pages 15–38. Lawrence Erlbaum Associates, Hillsdale, NJ, 1991.

M. Fujita, H. Kitano, and T. Doi. Robot entertainment. In A. Druin and J. Hendler, editors, *Robots for kids : exploring new technologies for learning*, chapter 2, pages 37–70. Morgan Kaufmann, 2000.

B. Fullmer and R. Miikkulainen. *Using marker-based genetic encoding of neural networks to evolve finite-state behaviour.* MIT Press, Cambridge, MA, USA, 1992.

S. Garrod and A. Anderson. Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, 27:181–218, 1987.

J. Gautrais, G. Theraulaz, J. Deneubourg, and C. Anderson. Emergent polyethism as a consequence of increased colony size in insect societies. *Journal of Theoretical Biology*, 215:363–373, 2002.

W. W. Gaver and H. Martin. Alternatives: exploring information appliances through conceptual design proposals. In *CHI*, pages 209–216, 2000.

L. Gerdes, C. Gerdes, P. Hansen, I. Klausen, and O. Faergman. Are men carrying the apolipoprotein ?4- or ?2 allele less fertile than ?3?3 genotypes? *Hum. Genet.*, 98:239–242, 1996.

B. P. Gerkey and M. J. Matarić. Sold!: Auction methods for multi-robot coordination. *IEEE Transactions on Robotics and Automation*, 18(5):758–768, Oct. 2002.

F. Gers, N. Schraudolph, and J. Schmidhuber. Learning precise timing with lstm recurrent networks. *Journal of Machine Learning Research*, 3:115–143, 2002.

I. Giardina and J.-P. Bouchaud. Bubbles, crashes and intermittency in agent based market models. *EPJB*, 2003.

I. Giardina, J.-P. Bouchaud, and M. Mezard. Population dynamics in a random environment. *Phys. A : Math. Gen. (Lett.)*, 2001.

H. Gintis, E. A. Smith, and S. Bowles. Costly signaling and cooperation. *Journal of Theoretical Biology*, 213:103–119, 2001.

E. L. Glaeser and J. Scheinkman. Non-market interactions. NBER Working Papers 8053, National Bureau of Economic Research, Inc, Dec. 2000. available at http://ideas.repec.org/p/nbr/nberwo/8053.html.

A. E. Goldberg. *A construction grammar approach to argument structure*. University of Chicago Press, Chicago, IL, USA, 1995.

L. Goldstein. Emergence of discrete gestures. In *Proceedings of the International Congress on Phonetics, Barcelona*, 2003.

A. Gopnik and L. E. Schulz. Mechanisms of theory-formation in young children. *Trends in Cognitive Science*, 8(8):371–377, 2004.

A. Gopnik, C. Glymour, D. Sobel, L. Schulz, T. Kushnir, and D. Danks. A theory of causal learning in children: causal maps and bayes nets. *Psychol. Rev.*, 11:3–32, 2004.

M. Gopnik. Feature-blind grammar and dysphasia. *Nature*, 344:715, 1990.

M. Gopnik. Familial language impairment: more english evidence. *Folia Phoniatr. Logop*, 51:5–19, 1999.

A. Grafen. Biological signals as handicaps. *Journal of Theoretical Biology*, 144: 517–246, 1990.

J. M. Greenbaum and M. Kyng, editors. *Design at Work: Cooperative Design of Computer Systems*. Lawrence Erlbaum Associates, Inc., 1991. ISBN 0805806113.

P. M. Greenfield. Language, tools, and brain: The ontogeny and phylogeny of hierarchically organized sequential behavior. *Behavioral and Brain Sciences*, 14(4):531–551, 1991.

P. Grosjean, C. Spirlet, and M. Jangoux. Experimental study of growth in the echinoid paracentrotus lividus (Lamarck, 1816) (Echinodermata). *J. Exp. Mar. Biol. Ecol.*, 201:173–184, 1996.

F. Gruau and D. Whitley. Adding learning to the cellular development of neural networks: Evolution and the baldwin effect. *Evolutionary Computation*, 1(3): 213–233, 1993.

J. L. Gunderman, A. Horel, and B. Krafft. Experimental manipulations of social tendencies in the subsocial spider coelotes terrestris. *Insectes Sociaux*, 40:229–239, 1993.

M. Gurven, W. Allen-Arave, K. Hill, and M. Hurtado. “it’s awonderful life”: signaling generosity among the ache of paraguay. *Evolution and Human Behavior*, 21:263–282, 2000.

V. Hafner and F. Kaplan. Learning to interpret pointing gestures: experiments with four-legged autonomous robots. In G. Palm and S. Wermter, editors, *Proceedings of the KI2004 Workshop on Neurobotics*, 2004.

J. P. Hailman. *Optical signals: animal communication and light*. Bloomington Indiana University Press, 1977.

T. Halpin-Healey and Y.-C. Zhang. Kinetic roughening, stochastic growth, directed polymers and all that. *Phys. Rep. 254*, 1995.

W. D. Hamilton. The genetical evolution of social behaviour. I. *J. Theor. Biol.*, 7:1–16, 1964a.

W. D. Hamilton. The genetical evolution of social behaviour. II. *J. Theor. Biol.*, 7:17–52, 1964b.

P. Hammerstein and R. S. E. Payoffs and strategies in territorial contests: Ess analyses of two ecotypes of the spider agelenopsis aperta. *Evolutionary Ecology*, 2:115–138, 1988.

S. Harnad. The symbol grounding problem. *Physica D*, 40:335–346, 1990.

D. Harris and S. Bullock. Enhancing game theory with coevolutionary simulation models of honest signalling. pages 1594–1599. IEEE Press, 2002.

M. D. Hauser. *The evolution of communication*. The MIT Press, Cambridge, MA, 1996.

M. D. Hauser, N. Chomsky, and W. T. Fitch. The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298:1569–1579, 11 2002.

K. Hawkes. Showing off: tests of an hypothesis about mens foraging goals. *Ethology and Sociobiology*, 12:2954, 1991.

T. Haynes and S. Sen. Evolving behavioral strategies in predators and prey. In S. Sen, editor, *IJCAI-95 Workshop on Adaptation and Learning in Multiagent Systems*, pages 32–37, Montreal, Quebec, Canada, 20-25 1995. Morgan Kaufmann.

T. Haynes and S. Sen. Co–adaptation in a team. *International Journal of Computational Intelligence and Organizations (IJCIO)*, 1997.

B. Heine. *Cognitive foundations of grammar*. Oxford University Press, Oxford, UK, 1997.

F. Helin, T. Hglund, R. Zackaroff, M. Hkansson, S. Ljungblad, and L. E. Holmquist. Supporting collaborative scheduling with interactive pushpins and networking surfaces. In *UbiComp 2004 Adjunct Proceedings*, Nottingham, UK.

C. K. Hemelrijk. Understanding of social behaviour with the help of complexity science. *Ethology*, 108:1–17, 2002.

G. E. Hinton and A. Brown. Spiking boltzmann machines. In *Advances in Neural Information Processing Systems*, volume 12, Cambridge, MA, 2000. MIT Press.

G. E. Hinton and S. J. Nowlan. How learning can guide evolution. *Complex Systems*, 1:495–502, 1987.

P. Hobson. *The craddle of thought*. MacMillan, 2002.

C. F. Hockett. The origin of speech. *Scientific American*, 203, 1960a.

C. F. Hockett. Logical considerations in the study of animal communication. In W. E. Lanyon and W. N. Tavolga, editors, *Animal sounds and communication*, pages 392–430. American Institute of Biological Sciences, Washington, DC, 1960b.

A. Hoffmann. Complexity bounds of emerging structures in self-organizing networks, 1993.

A. G. Hoffmann. Complex connectionist dynamics and the limited complexity of the resulting functions. *Connect. Sci.*, 9(2):201–216, 1997.

O. Holland and C. Melhuish. Stigmergy, self-organisation, and sorting in collective robotics. *Artificial Life*, 5(2):173–202, 1999.

B. Hölldobler and E. O. Wilson. *The Ants*. Harvard University Press, Cambridge, 1990.

L. E. Holmquist. User-driven innovation in the future applications lab. In E. Dykstra-Erickson and M. Tscheligi, editors, *Extended abstracts of the 2004 Conference on Human Factors in Computing Systems, CHI 2004, Vienna, Austria, April 24 - 29, 2004*, pages 1091–1092. ACM, 2004.

P. L. Hurd. Communication in discrete action-response games. *Journal of Theoretical Biology*, 174(2):217–222, 1995.

J. Hurford. *The Language Mosaic and its Evolution*. Oxford University Press, 2003.

J. Hurford. Biological evolution of the saussurean sign as a component of the language acquisition device. *Lingua*, 77:187–222, 1989.

J. Hurford, C. Knight, and M. Studdert-Kennedy, editors. *Approaches to the evolution of human language*. Edinburgh University Press, Edinburgh, 1998.

H. I., I. Husband, and D. Cliff. Seeing the light: artificial evolution, real. In *From Animals to Animats 3, Proceedings of the Third Simulation of Adaptive Behavior*. MIT Press, 1994.

H. Iizuka and T. Ikegami. Simulating autonomous coupling in discrimination of light frequencies, in press.

H. Iizuka and T. Ikegami. Adaptive coupling and intersubjectivity in simulated turn-taking behaviours. In *Proceedings of ECAL 03*. Springer Verlag, 2003a.

H. Iizuka and T. Ikegami. Simulating turn-taking behaviors with coupled dynamical recognizers. In *Proceedings of Artificial Life VIII*. MIT Press, 2003b.

A. J. Ijspeert, A. Martinoli, A. Billard, and L. M. Gambardella. Collaboration through the exploitation of local interactions in autonomous collective robotics: The stick pulling experiment. *Auton. Robots*, 11(2):149–171, 2001. ISSN 0929-5593.

T. Ikegami and H. Iizuka. Joint attention and dynamics repertoire in coupled dynamical recognizers. In *AISB 03: the Second International Symposium on Imitation in Animals and Artifacts*, pages 125–130, 2003.

M. Imai, T. Ono, and H. Ishiguro. Physical relation and expression: Joint attention for human-robot interaction. In *Proceedings of the 10th IEEE International Workshop on Robot and Human Communication*, 2001.

M. Ito and J. Tani. On-line imitative interaction with a humanoid robot using a dynamic neural network model of a mirror system. *Adaptive behavior*, 2004. in press.

R. Jackendoff. Possible stages in the evolution of the language capacity.c. *Trends in Cognitive Science*, 3(7):272–279, June 1999.

H. Jasso, J. Triesch, and C. Teuscher. Gaze following in the virtual living room. In G. Palm and S. Wermter, editors, *Proceedings of the KI2004 Workshop on Neurobotics*, 2004.

R. Jeanson, J. Deneubourg, and G. Theraulaz. Discrete dragline attachment induces aggregation in spiderlings of a solitary species. *Animal Behaviour*, 67: 531–537, 2004.

R. Jeanson, C. Rivault, J. Deneubourg, S. Blanco, R. Fournier, C. Jost, and G. Theraulaz. Self-organized aggregation in cockroaches. *Animal Behaviour*, in press.

K. Jeffery and I. Reid. Modifiable neuronal connections: An overview for psychiatrists. *Am. J. Psychiatry*, 154:156–164, 1997.

P. Johnson-Laird. Introduction: what is communication? In D. Mellor, editor, *Ways of Communicating*. Cambridge University Press, Cambridge, UK, 1990.

M. Johnston. Developmental disorders of activity dependent neuronal plasticity. *Ind. J. Pediat*, 68:423–426, 2001.

E. Kandel, J. Schwartz, and T. Jessel. *Principles of neural science*. McGraw-Hill, fourth edition, 2000.

F. Kaplan. *La naissance d'une langue chez les robots*. Hermes Science, 2001.

F. Kaplan. Who is afraid of the humanoid? investigating cultural differences in the acceptance of robots. *International Journal of Humanoid Robotics*, 1(3): 465–480, 2004.

F. Kaplan. *Les machines apprivoisees: Comprendre les robots de loisir*. Vuibert, 2005.

F. Kaplan and V. Hafner. The challenges of joint attention. In L. Berthouze, H. Kozima, C. Prince, G. Sandini, G. Stojanov, G. Metta, and C. Balkenius, editors, *Proceedings of the 4th International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic System*, pages 67–74. Lund University Cognitive Studies 117, 2004.

F. Kaplan and P.-Y. Oudeyer. From talking aibo to talking head: Morphing sensory spaces between two robotic bodies. In *Proceedings of Sony Research Forum 2000*, 2000.

F. Kaplan, P.-Y. Oudeyer, E. Kubinyi, and A. Miklosi. Robotic clicker training. *Robotics and Autonomous Systems*, 38(3-4):197–206, 2002.

K. Kaye. *The mental and social life of babies.* University of Chicago Press, Chicago, 1982.

L. Keller and M. Chapuisat. Eusociality and cooperation. In *Nature Encyclopedia of Life Sciences*. Nature Publishing Group, London, 2001.

L. Keller and M. Genoud. Extraordinary lifespans in ants: a test of evolutionary theories of ageing. *Nature*, pages 389, 958–960, 1997.

L. Keller and H. Reeve. Dynamics of conflicts within insect societies. *Levels of Selection in Evolution L. Keller (ed.)*, (7048):153–175, 1999.

C. Kemp, T. L. Griffiths, S. Stromsten, and J. B. Tenenbaum. Semi-supervised learning with trees. In S. Thrun, L. Saul, and B. Schölkopf, editors, *Advances in Neural Information Processing Systems 16*. MIT Press, Cambridge, MA, 2004.

J. Kennan and R. Wilson. Bargaining with private information. *Journal of Economic Literature*, 31(1):45–104, 1993.

J. Kephart. A biologically inspired immune system for computers. In R. A. Brooks and P. Maes, editors, *Artificial Life IV, Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems*, pages 130–139, Cambridge, Massachusetts, 1994. MIT Press.

M. Kerszberg, S. Dehaene, and J.-P. Changeux. Stabilization of complex input-output functions in neural clusters formed by synapse selection. *Neural Netw.*, 5(3):403–413, 1992. ISSN 0893-6080. doi: http://dx.doi.org/10.1016/0893-6080(92)90003-2.

D. Kimura. *Neuromotor mechanisms in human communication.* Oxford University Press, 1993.

S. Kirby. Syntax out of learning: the cultural evolution of structured communication in a population of induction algorithms. In D. Floreano, J.-D. Nicoud, and F. Mondada, editors, *Advances in Artificial Life (ECAL 99)*, Lecture Notes in Artificial Intelligence 1674, pages 694–703, Berlin, 1999a. Springer-Verlag.

S. Kirby. Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation*, 5(2):102–110, 2001.

S. Kirby. Natural language and artificial life. *Artificial life*, 8:185–215, 2002.

S. Kirby. *Function, selection and innatensess: The emergence of language universals.* Oxford University Press, Oxford, 1999b.

C. Knight. *Ritual/speech co-evolution: A 'selfish gene' solution to the problem of deception*. Cambridge University Press, Cambridge, 1998.

B. Kosko. Bidirectional associative memories. *IEEE Transactions on Systems, Man, and Cybernetics*, 18(2):49–60, 1988.

H. Kozima and H. Yano. A robot that learns to communicate with human caregivers. In *First International Workshop on Epigenetic Robotics (Lund, Sweden)*, 2001.

J. R. Krebs and R. Dawkins. Animal signals: Mindreading and manipulation. In J. R. Krebs and N. B. Davies, editors, *Behavioural ecology: An evolutionary approach*. Blackwell Scientific, Oxford, 1984.

M. Krieger, J.-B. Billeter, and L. Keller. Ant-like task allocation and recruitment in cooperative robots. *Nature*, 404(7352):992–995, 2000.

C. R. Kube and E. Bonabeau. Cooperative transport by ants and robots. *Robotics and Autonomous Systems*, 30(1-2):85–101, 2000.

C. R. Kube and H. Zhang. Collective robotics: from social insects to robots. *Adapt. Behav.*, 2(2):189–218, 1993. ISSN 1059-7123.

M. Kusahara. The art of creating subjective reality: an analysis of japanese digital pets. In M. C. and E. Boudreau, editors, *Artificial life VII Workshop Proceedings*, pages 141–144, 2000.

T. Labella, M. Dorigo, and J.-L. Deneubourg. Efficiency and task allocation in prey retrieval. In A. Ijspeert, D. Mange, M. Murata, and S. Nishio, editors, *Proceedings of the First International Workshop on Biologically Inspired Approaches to Advanced Information Technology (Bio-ADIT2004)*, Lecture Notes in Computer Science, pages 32–47. Springer Verlag, Heidelberg, Germany, 2004.

M. Lachmann and C. T. Bergstrom. The disadvantage of combinatorial communication. *Proceedings of the Royal Society of London B*, 271:2337–2343, 2004.

M. Lachmann, S. Szamado, and C. T. Bergstrom. Cost and conflict in animal signals and human language. *PNAS*, 98(23):13189–13194, November 2001.

P. Ladefoged and I. Maddieson. *The sounds of the world's languages*. Blackwell, Oxford, UK, 1996.

P. Latham, S. Deneve, and A. Pouget. Optimal comp. 97:683–94, 2003.

O. Leimar and P. Hammerstein. Evolution of cooperation through indirect reciprocity. B 268:745753, 2001.

D. Lenat, G. Miller, and T. Yokoi. Cyc, wordnet, and edr: critiques and responses. *Commun. ACM*, 38(11):45–48, 1995. ISSN 0001-0782. doi: http://doi.acm.org/10.1145/219717.219757.

M. Lengyel and P. Dayan. Rate- and phase-coded autoassociative memory. In *Advances in Neural Information Processing Systems 17*. MIT Press, Cambridge, MA, 2004. in press.

K. Lerman, A. Galstyan, A. Martinoli, and A. Ijspeert. A macroscopic analytical model of collaboration in distributed robotic systems. *Artif. Life*, 7(4):375–393, 2002. ISSN 1064-5462. doi: http://dx.doi.org/10.1162/106454601317297013.

O. Levi, A. Jongen-Relo, J. Feldon, A. Roses, and D. Michaelson. Apoe4 impairs hippocampal plasticity isoform-specifically and blocks the environmental stimulation of synaptogenesis and memory. *Neurobiol. Disease*, 13:273–282, 2003.

B. Lindblom. *Systemic constraints and adaptive change in the formation of sound structure*. Cambridge University Press, Cambridge, 1998.

J. Lindblom and T. Ziemke. Social situatedness of natural and artificial intelligence: Vygotsky and beyond. *Adaptive Behavior*, 11(2):79–96, 2003.

A. Lioni and J. Deneubourg. Collective decision through self-assembling. *Naturwissenschaften*, 91:237–41, 2004.

S. Ljungblad, M. Håkansson, L. Gaye, and L. E. Holmquist. Context photography: modifying the digital camera into a new creative tool. In E. Dykstra-Erickson and M. Tscheligi, editors, *Extended abstracts of the 2004 Conference on Human Factors in Computing Systems, CHI 2004, Vienna, Austria, April 24 - 29, 2004*. ACM, 2004. ISBN 1-58113-703-6.

J. V. Looveren. Robotic experiments on the emergence of a lexicon. In B. e. a. Krse, editor, *Proceedings of the 13th Belgium-Netherlands Conference on Artificial Intelligence (BNAIC'01)*, Amsterdam, The Netherlands, 2001. Best original paper award.

R. M. Losee. Communication defined as complementary informative processes. *Journal of Information*, 5:1–15, 1999.

W. E. MacKay. Is paper safer? the role of paper flight strips in air traffic control. *ACM Trans. Comput.-Hum. Interact.*, 6(4):311–340, 1999. ISSN 1073-0516. doi: http://doi.acm.org/10.1145/331490.331491.

B. MacLennan and G. M. Burghardt. Synthetic ethology and the evolution of cooperative communication. *Adaptive Behavior*, 2(2):161–187, Fall 1993. for 8 figures, see author's publication page.

A. Mailleux, J. Deneubourg, and C. Detrain. How do ants assess food volume? *Animal Behaviour*, 59:1061–1069, 2000.

A. Mailleux, J. Deneubourg, and C. Detrain. Regulation of ants foraging to ressource productivity. *Proceedings of the Royal Society*, 270:1609–1616, 2003.

G. Marcus and S. Fisher. Foxp2 in focus: what can genes tell us about speach and language. *Trends in Cognitive Sciences*, 6:257–262, 2003.

G. F. Marcus. Rethinking eliminative connectionism. *Cognitive Psychology*, 37: 243–282, 1998.

G. F. Marcus, S. Viyayan, P. Bandi Rao, and M. Vishton. Rule learning by seven-month-old infants. *Science*, 283:77–80.

H. Markl. Manipulation, modulation, information, cognition: some of the riddles of communication, 1985.

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

D. Marr. *Vision : A computational investigation into the human representation and processing of visual information*. Freeman, San Francisco, CA/, 1982.

A. Martinoli and F. Mondada. Collective and cooperative behaviours: Biologically inspired experiments in robotics. In *4th International Symposium on Experimental Robotics, ISER'95, Stanford, CA*, pages 2–7, 1995.

A. Martinoli, A. J. Ispeert, and L. Gambardella. A probabilistic model for understanding and comparing collective aggregation mechanisms. In D. Floreano, J.-D. Nicoud, and F. Mondada, editors, *Advances in Artificial Life (ECAL 99)*, Lecture Notes in Artificial Intelligence 1674, pages 575–584, Berlin, 1999. Springer-Verlag.

M. Mataric. Coordination and learning in multirobot systems. *IEEE Intelligent Systems*, pages 6–8, March/April 1998.

J. Maynard Smith. Honest signalling: The philip sidney game. *Animal Behaviour*, 42:1034–1035, 1991.

J. Maynard Smith and D. Harper. Animal signals: Models and terminology. *Journal of Theoretical Biology*, 117:305–311, 1995.

J. Maynard-Smith and E. Szathmary. *The Major Transitions in Evolution*. New York: Oxford University Press, 1997.

J. Maynard-Smith and E. Szathmary. *The major transitions in evolution*. W.H. Freeman, San Francisco, 1988.

J. Maynard Smith and E. Szathmáry. *The major transitions in evolution*. Oxford University Press, Oxford, 1997.

A. McIntyre, L. Steels, and F. Kaplan. Net-mobile embodied agents. In *Proceedings of Sony Research Forum 1999*, Tokyo, 1999.

A. Meltzoff. Understanding the intentions of others : Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, 31:838–850, 1995.

A. Meltzoff and A. Gopnick. The role of imitation in understanding persons and developing a theory of mind. In H. T.-F. S. Baron-Cohen and D.Cohen, editors, *Understanding other minds*, pages 335–366. Oxford University Press, Oxford, England, 1993.

O. Michel. Webots: Symbiosis between virtual and real mobile robots. In J.-C. Heudin, editor, *Proceedings of Virtual Worlds 98*, Lecture Notes in Artificial Intelligence 1434, pages 254–263, Berlin, 1998. Springer-Verlag.

E. Miller, D. Freedman, and J. Wallis. The prefrontal cortex: categories, concepts and cognition. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 357:1123–36, 2002.

E. Miller, A. Nieder, D. Freedman, and J. Wallis. Neural correlates of categories and concepts. *Curr. Opin. Neurobiol.*, 13:198–203, 2003.

G. Miller. *The Mating Mind*. Anchor Books, 2001.

J. Millor, M. Pham-Delegue, J.-L. Deneubourg, and S. Camazine. Self-organized defensive behavior in honeybees. *P.N.A.S.*, 96:12611–12615, 1999.

M. Mirolli and D. Parisi. Cognitive, genetic and cultural pressure towards the evolution of a good communication system, in preparation. Technical Report, Institute of Cognitive Science and Technologies, CNR, Roma, Italy.

C. Moore and V. Corkum. Social understanding at the end of the first year of life. *Developmental Review*, 14:349–372, 1994.

M. Mrayati, R. Carr, and B. Gurin. Distinctive region and modes: A new theory of speech production. *Speech Communication*, 7:257–286, 1988.

S. S. Mufwene. *The Ecology of Language Evolution*. Cambridge University Press, Cambridge, 2001.

R. A. Müller, M. E. Behen, R. D. Rothermel, O. Muzik, P. K. Chakraborty, and H. T. Chugani. Brain organization for language in children, adolescents, and adults with left hemisphere lesion: a pet study. *Neuropsychology*, 37(545-557).

S. Munroe and A. Cangelosi. Cultural transmission, learning cost and the baldwin effect in language evolution. In J. Hurford and T. Titch, editors, *Proceedings of the 4th International Conference on the Evolution of Language*, page 79, Boston, 2002.

J. Nadel. Imitation and imitation recognition : functional use of imitation in preverbal infants and nonverbal children with autism. In A.Meltzoff and W. Prinz, editors, *The imitative mind : development, evolution and brain bases*, pages 42–62. Cambridge University Press, 2002.

Y. Nagai, M. Asada, and K. Hosoda. A developmental approach accelerates learning of joint attention. In *Proceedings of the second international conference of development and learning*, 2002.

Y. Nagai, K. Hosoda, A. Morita, and M. Asada. A constructive model for the development of joint attention. *Connection Science*, 15(4):211–229, 2003.

W. Nejdl, M. Schlosser, W. Siberski, M. Wolpers, B. Simon, S. Decker, and M. Sintek. Rdf-based peer-to-peer-networks for distributed (learning) repositories. *VLDB Journal*, 2003.

H. J. Neville and D. Bavelier. Neural organization and plasticity of language. *Current Opinion in Neurobiology*, 8:254–258, 1998.

M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167–256, 2003.

J. Newson. *The growth of shared understandings between infant and caregiver*, pages 207–222. Cambridge University Press, 1979.

S. Nicolis, C. Detrain, D. Demolin, and J. Deneubourg. Optimality of collective choices: a stochastic approach. *Bulletin of Mathematical Biology*, 65:795–808, 2003.

P. Niyogi and R. Berwick. The logical problem of language change. Technical report, MIT AI Laboratory and Center for Biological and Computational Learning, Department of Brain and Cognitive Sciences, 1995. Technical Report AI Memo 1516/CBCL Paper 115.

J. Noble. Cooperation, competition and the evolution of prelinguistic communication. *The Emergence of Language*, pages 40–61, 2000.

J. Noble and D. Cliff. On simulating the evolution of communication. In *From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior*. MIT Press, 1996.

J. Noble, E. A. D. Paolo, and S. Bullock. *Adaptive Factors in the Evolution of Signaling Systems*, chapter 3, pages 53–78. Springer Verlag, London, 2002.

S. Nolfi. Categories formation in self-organizing embodied agents. In C. Lefebvre, editor, *Handbook of Categorization in Cognitive Science*. Elsevier, in press.

S. Nolfi. Evolution and learning in neural networks. In A. Arbib, editor, *Handbook of brain theory and neural networks*. MIT Press, 2002a.

S. Nolfi. Power and limits of reactive agents. *Neurocomputing*, 49:119–145, 2002b.

S. Nolfi. Evolving robots able to self-localize in the environment: The importance of viewing cognition as the result of processes occurring at different time scales. *Connection Science*, 14(3):231–244, 2002c.

S. Nolfi and D. Floreano. Co-evolving predator and prey robots: Do arm races arise in artificial evolution? *Artificial Life*, 4(1):311–335, 1998.

S. Nolfi and D. Floreano. *Evolutionary Robotics : Biology, intelligence and technology of self-organizing machines*. MIT Press, Cambridge, MA, 2000.

S. Nolfi and D. Floreano. Learning and evolution. *Autonomous Robots*, 7(1): 89–113, 1999.

S. Nolfi and D. Marocco. Active perception: A sensorimotor account of object categorization. In *From Animals to Animats 7: Proceedings of the VII International Conference on Simulation of Adaptive Behavior*. MIT Press, 2002.

S. Nolfi and D. Marocco. Emergence of communication in teams of robots evolved for solving a collective navigation problem, submitted.

S. Nolfi and D. Parisi. Learning to adapt to changing environments in evolving neural networks. *Adaptive Behavior*, 5(1):75–98, 1997.

S. Nolfi and J. Tani. Extracting regularities in space and time through a cascade of prediction networks: The case of a mobile robot navigating in a structured environment. *Connection Science*, 2:129–152, 2002.

S. Nolfi and J. Tani. Extracting regularities in space and time through a cascade of prediction networks. *Connection Science*, 11(2):129–152, 1999.

S. Nolfi, J. Elman, and D. Parisi. Learning and evolution in neural networks. *Adaptive Behavior*, 3(1):5–28, 1994.

M. Nowak and D. Krakauer. The evolution of language. *Proceedings of the National Academia of Science USA*, 96:8028–8033, 1999.

M. Nowak, J. Plotkin, and V. Jansen. The evolution of syntactic communication. *Nature*, 404(13):495–498, March 2000.

M. A. Nowak and K. Sigmund. Evolution of indirect reciprocity by image scoring. *Nature*, 393:573–577, 1998.

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

M. Oliphant. *Formal Approaches to Innate and Learned Communication: Laying the Foundation for Language*. PhD thesis, Department of Cognitive Science, University of California, San Diego, 1997.

P.-Y. Oudeyer. Origins and learnability of syllable systems, a cultural evolutionary. In P. Collet, C. Fonlupt, J. Hao, E. Lutton, and M. Schonenauer, editors, *Artificial Evolution*, volume 2310 of *Lectures Notes in Computer Science*, pages 143–155. Springer Verlag, 2001.

P.-Y. Oudeyer. *L'auto-organisation de la parole*. PhD thesis, University Paris VI, 2003.

P.-Y. Oudeyer. The self-organization of speech sounds. *Journal of Theoretical Biology*, 233(3):435–449, 2005.

R. G. Palmer, W. B. Arthur, J. H. Holland, B. LeBaron, and P. Tayler. Artificial economic life: a simple model of a stockmarket. In *Proceedings of the Oji international seminar on Complex systems : from complex dynamical systems to sciences of artificial reality*, pages 264–274. Elsevier North-Holland, Inc., 1994. doi: http://dx.doi.org/10.1016/0167-2789(94)90287-9.

D. Parisi and A. Cangelosi. *A Unified Simulation Scenario for Language Development, Evolution, and Historical Change*, chapter 12, pages 255–276. Springer Verlag, London, 2002.

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

L. Parker. Alliance: An architecture for fault-tolerant multi-robot cooperation. *IEEE Transactions on Robotics and Automation*, 14:504–516, April 1998.

L. E. Parker. *Heterogeneous multi-robot cooperation*. PhD thesis, 1994.

J. K. Parrish and L. Edelstein-Keshet. Complexity, pattern, and evolutionary trade-offs in animal aggregation. *Science*, 284:99–101, 1999.

B. Partee. *Compositionality in Formal Semantics: Selected Papers of Barbara Partee*. Blackwell Publishers, Oxford, UK, 2003.

L. Passera, E. Roncin, B. Kaufmann, and L. Keller. Increased soldier production in ant colonies exposed to intraspecific competition. *Revue Suisse de Zoologie, 103*, page 789, 1996.

J. Pasteels, J. Deneubourg, and S. Goss. Transmission and amplification of information in a changing environment: The case of insect societies. In I. Prigogine and M. Sanglier, editors, *Law of Nature & Human conduct*, pages 129–156. Gordes, Bruxelles, 1987.

M. Pena, L. Bonatti, M. Nespor, and J. Mehler. Signal-driven computations in speech processing. *Science*, 298:604–607, 2002.

J. Piaget. *The origins of intelligence in children*. Norton, New York, 1952.

S. Pinker. *The Language Instinct*. Penguin Books, London, 1994.

S. Pinker and P. Bloom. Natural language and natural selection. *Behavioral and Brain Sciences*, 13(4):707–784, 1990.

S. Pinker and R. Jackendoff. The faculty of language: What's special about it? *Cognition*, 2005.

J. B. Plotkin and M. A. Nowak. Language evolution and information theory. *Journal of Theoretical Biology*, 205(1):147–159, 2000.

G. Pollock and L. A. Dugatkin. Reciprocity and the emergence of reputation. *Journal of Theoretical Biology*, 159:25–37, 1992.

S. Portha, J.-L. Deneubourg, and C. Detrain. *Behavioural Ecology*, 13:776–781, 2002.

A. Pouget, K. Zhang, S. Deneve, and P. E. Latham. Statistically efficient estimation using population coding. *Neural Computation*, 10(2):373–401, 1998.

D. Poulin-Dubois. Infantsd́istinction between animate and inanimate objects: The origins of naive psychology. In P. Rochat, editor, *Early Social Cognition*.

Erlbaum, 1999.

D. Povinelli. On the possibilities of detecting intentions prior to understanding them. In F. Malle, L. Moses, and D. Baldwin, editors, *Intentions and Intentionality*, chapter 11. MIT Press, 2001.

C. Power. *Old wives' tales: The gossip hypothesis and the reliability of cheap signals*. Cambridge University Press, Cambridge, 1998.

D. Premack. Is language the key to human intelligence? *Science*, pages 318–320, 2004.

D. Premack. The infant's theory of self-propelled objects. *Cognition*, 36:1–16, 1990.

T. Preuss. Taking the measure of diversity: Comparative alternatives to the model-animal paradigm in cortical neuroscience. *Brain Behav. Evol.*, 55:287–299, 2000.

I. Prigogine and I. Stengers. *Order Out of Chaos*. Bantam Books, New York, 1984.

M. Quinn. Evolving communication without dedicated communication channels. In J. Kelemen and P. Sosík, editors, *Advances in Artificial Life, 6th European Conference, ECAL 2001, Prague, Czech Republic, September 10-14, 2001, Proceedings*, volume 2159 of *Lecture Notes in Computer Science*, pages 357–366. Springer, 2001. ISBN 3-540-42567-5. URL citeseer.nj.nec.com/quinn01evolving.html.

M. Quinn, L. Smith, G. Mayley, and P. Husbands. Evolving controllers for a homogeneous system of physical robots: Structured cooperation with minimal sensors. *Philosophical Transactions of the Royal Society of London, Series A: Mathematical, Physical and Engineering Sciences*, 361:2321–2344, 2003.

J. Raber, D. Wong, G. Yu, M. Buttini, R. Mahley, R. Pitas, and L. Mucke. Apolipoprotein e and cognitive performance. *Nature*, 404:353–353, 2000.

G. Rainer, W. Asaad, and E. Miller. Selective representation of relevant information by neurons in primate prefrontal cortex. *Nature*, 393:577–579, 1998.

R. P. N. Rao. Bayesian computation in recurrent neural circuits. *Neural Comput.*, 16(1):1–38, 2004. ISSN 0899-7667. doi: http://dx.doi.org/10.1162/08997660460733976.

S. Rapoport. How did the human brain evolve? a proposal based on new evidence from in vivo brain imaging during attention and ideation. *Brain Res. Bull*, 50:149–165, 1999.

F. Ratnieks. Reproductive harmony via mutual policing by workers in eusocial hymenoptera. *American Naturalist*, 132(2):217–236, 1988.

F. L. W. Ratnieks and H. K. Reeve. Conflict in single-queen hymenopteran societies - the structure of conflict and processes that reduce conflict in advanced eusocial species. *Journal of Theoretical Biology*, 158:33–65, 1992.

H. K. Reeve, , and L. Keller. Reproductive bribing and policing evolutionary mechanisms for the suppression of within-group selfishness. *American Naturalist*, 150:42–58, 1997.

B. Repacholi and A. Gopnik. Early reasoning about desires: Evidence from 14- and 18-month-olds. *Developmental Psychology*, 33:12–21, 1997.

S. E. Riechart. Games spiders play: behavioral variability in territorial disputes. *Behavioral Ecology and Sociobiology*, 3:135–162, 1978.

C. Rivault and A. Cloarec. Cockroach aggregation: discrimination between strain odours in blattella germanica. *Animal Behaviour*, 55:177–184, 1998.

P. Rochat, R. Morgan, and M. Carpenter. Young infants' sensitivity to movement information specifying social causality. *Cognitive Development*, 12(4):441–465, 1997.

D. Roy and A. Pentland. Learning words from sights and sounds: A computational model. *Cognitive science*, 26:113–146, 2002.

A. Rubinstein. Perfect equilibrium in a bargaining model. *Econometrica*, 50(1):97–109, 1982. available at http://ideas.repec.org/a/ecm/emetrp/v50y1982i1p97-109.html.

A. L. Rypstra. Prey size, social competition, and the development of reproductive division of labor in social spider groups. *American Naturalist*, 142: 868–880, 1993.

J. Saffran, E. Newport, and R. Aslin. Statistical learning by 8-month-old infants. *Science*, 274:1926–1928, 1996.

F. Saffre, B. Krafft, and J. Deneubourg. What are the mechanisms involved in the emergence of cooperativity? the "spide" model. In G. Theraulaz and F. Spitz, editors, *Auto-organisation et comportement*, pages 85–90. Hermes, 1997.

F. Saffre, R. Furey, B. Krafft, and J. Deneubourg. Collective decision- making in social spiders: dragline-mediated amplification process acts as a recruitment mechanism. *Journal of Theoretical Biology*, 198:507–517, 1999.

M. Sahani and P. Dayan. Doubly distributional population codes: simultaneous representation of uncertainty and multiplicity. *Neural Comput.*, 15 (10):2255–2279, 2003. ISSN 0899-7667. doi: http://dx.doi.org/10.1162/089976603322362356.

B. Scassellati. Imitation and mechanisms of joint attention: A developmental structure for building social skills on a humanoid robot. In *Computation for metaphors, analogy and agents*, Vol 1562 of Springer Lecture Notes in Artificial Intelligence. Springer Verlag, 1999.

H. Schaffer. Early interactive development in studies of mother-infant interaction. In *Proceedings of Loch Lomonds Symposium*, pages 3–18, New York, 1977. Academic Press.

C. Scheier, R. Pfeifer, and Y. Kunyioshi. mbedded neural networks: exploiting constraints. *Neural Networks*, 11:1551–1596, 1998.

T. Seeley. The honey-bee colony as a superorganism. *American Scientist*, pages 546–553, 1989.

T. D. Seeley. *The Wisdom of the Hive: The Social Physiology of Honey Bee Colonies*. Harvard University Press, Cambridge, MA, 1995.

T. D. Seeley. Honey bee colonies are group-level adaptive units. *American Naturalist*, 150:22–41, 1997.

J. Seger. *Cooperation and conflict in social insects*. Blackwell Scientific Press, Oxford, 1991.

M. Seidenberg. Language acquisition and use: Learning and applying probabilistic constraints. *Science*, 275:1599–1604, 1997.

M. Seidenberg, M. MacDonald, and J. Saffran. Neuroscience. does grammar start where statistics stop? *Science*, 298:553–4, 2002.

A. Senghas, S. Kita, and A. Ozyurek. Children creating core properties of language: Evidence from an emerging sign language in nicaragua. *Science*, 305 (5691):1779–1782, 2004.

C. Shannon and W. Weaver. *The mathematical theory of communication*. University of Illinois Press, 1962.

D. Siegel. *The developing mind : Toward a neurobiology of interpersonal experience*. The Guildford press, New York, NY, 1999.

J. B. Silk, E. Kaldor, and R. Boyd. Cheap talk when interests conflict. animal behaviour. 59:423–432, 2000.

K. Sims. Evolving 3d morphology and behavior by competition. *Artificial Life*, 1:353–372, 1995.

P. Sinha. *Perceiving and recognizing three-dimensional forms*. PhD thesis, Massachusetts Institute of Technology., 1996.

J. M. Siskind. Grounding the lexical semantics of verbs in visual perception using force dynamics and event logic. *Journal of Artificial Intelligence Research*, 15: 31–90, 2001.

E. A. Smith and R. L. Bliege-Bird. Turtle hunting and tombstone opening: Public generosity as costly signaling, 2000.

K. Smith. The evolution of vocabulary. *Journal of Theoretical Biology*, 228(1): 127–142, 2004.

A. Somayaji and S. Forrest. Automated response using system-call delays. In *Proceedings of the 9th USENIX Security Symposium*, Berkeley, CA, 2000. The USENIX Association.

A. Somayaji, S. Hofmeyr, , and S. Forrest. Principles of a computer immune system. pages 75–82, New York, NY, 1997. The Association for Computing Machinery.

A. M. Spence. Job market signaling. *The Quarterly Journal of Economics*, 87: 355–374, 1973.

D. Sperber, D. Premack, and A. Premack. *Causal cognition: a multidisciplinary debate*. Oxford University Press, 1994.

O. Sporns, D. R. Chialvo, M. Kaiser, and C. C. Hilgetag. Organization, development and function of complex brain networks. *Trends in Cognitive Sciences*, 8:418–425, 2004.

K. O. Stanley and R. Miikkulainen. A taxonomy for artificial embryogeny. *Artif. Life*, 9(2):93–130, 2003. ISSN 1064-5462. doi: http://dx.doi.org/10.1162/106454603322221487.

A. Starrett. Adaptive resemblance: a unifying concept for mimicry and crypsis. *Biological Journal of the Linnean Society*, 48:299–317, 1993.

L. Steels. Language as a complex adaptive system. In M. Schoenauer, editor, *Proceedings of PPSN VI*, Lecture Notes in Computer Science, Berlin, Germany, September 2000. Springer-Verlag.

L. Steels. The origins of ontologies and communication conventions in multi-agent systems. *Journal of Agents and Multi-Agent Systems*, 1(1), 1997a.

L. Steels. Constructivist development of grounded construction grammars. In *Proceedings Annual Meeting Association for Computational Linguistics Conference*, Barcelona, 2004.

L. Steels. Perceptually grounded meaning creation. In M. Tokoro, editor, *ICMAS96*. AAAI Press, 1996a.

L. Steels. The origins of ontologies and communication conventions in multi-agent systems. *Autonomous Agents and Multi-Agent Systems*, 1(2):169–194, October 1998a.

L. Steels. Evolving grounded communication for robots. *Trends in Cognitive Science*, 7(7):308–312, July 2003a.

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

L. Steels. Self-organizing vocabularies. In C. Langton and T. Shimohara, editors, *Proceeding of Alife V*, Cambridge, MA, 1996b. The MIT Press.

L. Steels. Perceptually grounded meaning creation. In M. Tokoro, editor, *Proceedings of the International Conference on Multi-Agent Systems*, Cambridge, MA, 1996c. The MIT Press.

L. Steels. Synthesising the origins of language and meaning using co-evolution, self-organisation and level formation. In J. Hurford, C. Knight, and M. Studdert-Kennedy, editors, *Approaches to the Evolution of Language: Social and Cognitive bases*. Edinburgh University Press, Edinburgh, 1997b.

L. Steels. The synthetic modeling of language origins. *Evolution of Communication Journal*, 1(1):1–34, 1997c.

L. Steels. The origins of syntax in visually grounded robotic agents. *Artificial Intelligence*, 103:1–24, 1998b.

L. Steels. *The Talking Heads Experiment. Volume 1. Words and Meanings*. Antwerpen, 1999.

L. Steels and J.-C. Baillie. Evolving grounded communication for robots. *Trends in Cognitive Science*, 7(7):308–312, July 2003.

L. Steels and T. Belpaeme. Computational simulations of colour categorisation and colour naming. *Behavioral and Brain Sciences*, ?(?):?, 2005.

L. Steels and F. Kaplan. Aibo's first words: The social learning of language and meaning. *Evolution of Communication*, 4(1):3–32, 2000a.

L. Steels and F. Kaplan. Aibo's first words: The social learning of language and meaning. *Evolution of Communication*, 4(1):3–32, 2000b.

L. Steels and F. Kaplan. Spontaneous lexicon change. In *Proceedings of COLING-ACL 1998*, pages 1243–1249, Montreal, August 1998. ACL.

L. Steels and F. Kaplan. Situated grounded word semantics. In T. Dean, editor, *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence IJCAI'99*, pages 862–867, San Francisco, CA., 1999a. Morgan Kaufmann Publishers.

L. Steels and F. Kaplan. Bootstrapping grounded word semantics. In T. Briscoe, editor, *Linguistic evolution through language acquisition: formal and computational models*, chapter 3, pages 53–73. Cambridge University Press, Cambridge, 2002.

L. Steels and F. Kaplan. Collective learning and semiotic dynamics. In D. Floreano, J.-D. Nicoud, and F. Mondada, editors, *Advances in Artificial Life (ECAL 99)*, Lecture Notes in Artificial Intelligence 1674, pages 679–688, Berlin, 1999b. Springer-Verlag.

L. Steels and A. McIntyre. Spatially distributed naming games. *Advances in Complex Systems*, 1(4):301–323, January 1999.

L. Steels and P. Vogt. Grounding adaptive language games in robotic agents. In I. Harvey and P. Husbands, editors, *ECAL97*, Cambridge, MA, 1997a. MIT

Press.

L. Steels and P. Vogt. Grounding adaptive language games in robotic agents. In I. Harvey and P. Husbands, editors, *Proceedings of the 4th European Conference on Artificial Life*, Cambridge, MA, 1997b. The MIT Press.

L. Steels, F. Kaplan, A. McIntyre, and J. V. Looveren. *Crucial Factors in the Origins of Word-Meaning*, chapter 12. Oxford University Press, Oxford, 2002a.

L. Steels, F. Kaplan, A. McIntyre, and J. Van Looveren. Crucial factors in the origins of word-meaning. In A. Wray, editor, *The Transition to Language*, chapter 12, pages 252–271. Oxford University Press, Oxford, UK, 2002b.

K. Stevens. *The quantal nature of speech: evidence from articulatory-acoustic data*, pages 51–66. New-York: Mc Graw-Hill, 1972.

M. Studdert-Kennedy and L. Goldstein. *Launching language: The gestural origin of discrete infinity.* Oxford University Press, 2003.

Y. Sugita and J. Tani. A connectionist approach to learn association between sentences and behavioral patterns of a robot. In *SAB04*, pages 467–476, 2004.

L. S. Sugiyama and R. Chacon. Effects of illness and injury among the yora and shiwiar. In N. A. Chagnon, L. Cronk, and W. Irons, editors, *Human behavior and adaptation: An anthropological perspective.* Aldine de Gruyter, New York, 2000.

D. J. T. Sumpter and S. C. Pratt. A framework for modelling social insect foraging. *Behavioral Ecology and Sociobiology*, 53:131–144, 2003.

L. Sundstrm, M. Chapuisat, and L. Keller. Conditional manipulation of sex ratios by ant workers: a test of kin selection theory. *Science*, 274:993–995, 1996.

S. Szamado. The validity of the handicap principle in discrete action-response games. *Journal of Theoretical Biology*, 198:593–602, 1999.

S. Szamado and E. Szathmary. Language evolution: Competing selective scenarios. *Trends in Ecology and Evolution*, 2005. submitted.

E. Szathmary. The evolution of replicators. *Phil. Trans. R. Soc. Lond. B.*, 355: 1669–1676, 2000.

E. Szathmary. Origin of the human language faculty: the language amoeba hypothesis. In J. Trabant and S. Ward, editors, *New Essays on the Origin of Language*, pages P41–51. Mouton/de Gruyter, Berlin/New York, 2001.

S. T., C. S., and S. J. Collective decision-making in honey bees: how colonies choose among nectar sources. *Behavioral Ecology and Sociobiology*, 28:277–290, 1991.

L. Talmy. *Toward a cognitive semantics : concept structuring systems (language, speech and communication).* The MIT Press, Cambridge, MA, 2000.

J. Tani, M. Ito, and Y. Sugita. Self-organization of distributedly represented multiple behavior schema in a mirror system. *Neural Networks*, 17:1273–1289, 2004.

J. Tenenbaum. *Bayesian modeling*. 1999.

B. Teter, P. Xu, J. Gilbert, A. Roses, D. Galasko, and M. Cole. Defective neuronal sprouting by human apolipoprotein e4 is a gain-of-negative function. *J. Neurosci. Res*, 68:331–336, 2002.

d. Thomson. *On growth and form*. canto, 1961.

M. Tomasello. Joint attention as social cognition. In C. Moore and P. Dunham, editors, *Joint attention: its origins and role in development*, pages 103–130. Lawrence Erlbaum Associates, 1995.

M. Tomasello. *The cultural origins of human cognition*. Harvard U.P., 2000.

M. Tomasello, J. Call, and B. Hare. Chimpanzees understand psychological states: the question is which ones and to what extend. *Trends in Cognitive Science*, 7:153–156, 2003.

M. Tomasello, M. Carptenter, J. Call, T. Behne, and H. Moll. Understanding and sharing intentions: the origins of cultural cognition. *Behavioral and Brain Sciences (in press)*, 2004.

G. Tononi, G. Edelman, and O. Sporns. Complexity and coherency: integrating information in the brain. *Trends in cognitive sciences*, 2(12), 1998.

C. F. Touzet. Robot awareness in cooperative mobile robot learning. *Auton. Robots*, 8(1):87–97, 2000.

C. Trevarthen. Instincts for human understanding and for cultural cooperation: Development in infancy. In M. von Cranach, K. Foppa, W. Lepenes, and D. Ploog, editors, *Human ethology: Claims and limits of a new discipline*. Cambridge University Press, 1979.

A. Treves. Computational constraints that may have favoured the lamination of sensory cortex. *Journal of Computational Neuroscience*, 14(3):271–282, 2003.

M. Vaneechoutte and J. Skoyles. The memetic origin of language: modern humans as musical primates. *Journal of Memetics - Evolutionary Models of Information Transmission*, 1998.

M. Vihman. *Phonological Development: The Origins of Language in the Child*. Oxford, UK: Blackwell Publishers, 1996.

P. Visscher and S. Camazine. Collective decisions and cognition in bees. *Nature*, 397, 1999.

P. Vogt. *Lexicon grounding on mobile robots*. PhD thesis, Vrije Universiteit Brussel, 2000.

C. von der Malsburg. *The Handbook of Brain Theory and Neural Networks*, chapter Self-Organization and the Brain. The MIT Press, 2003.

K. von Frisch. *Animal Architecture*. London: Hutchinson, 1974.

L. S. Vygotsky. *Thought and language*. MIT Press, Cambridge, 1986.

C. Waddington. Canalization of development and the inheritance of acquired characters. *Nature*, 150:563–565, 1942.

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

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

J. Wallis, K. Anderson, and E. Miller. Single neurons in prefrontal cortex encode abstract rules. *Nature*, 411:953–6, 2001.

S. Washburn and C. Lancaster. The evolution of hunting. In R. Lee and I. De-Vore, editors, *Man the Hunter*, pages 293–303. Aldine, Chicago, 1968.

D. J. Watts and S. H. Strogatz. Collective dynamics of "small-world" networks. *Nature*, 393:440–442, 1998.

G. Weisbuch, H. Gutowitz, and G. Duchateau-Nguyen. Information contagion and the economics of pollution. Santa Fe series 94-04-018, 1994.

H. Wellman and A. Phillips. Developing intentional understandings. In F. Malle, L. Moses, and D. Baldwin, editors, *Intentions and Intentionality*, chapter 6. MIT Press, 2001.

G. M. Werner and M. G. Dyer. Evolution of communication in artificial organisms. In C. G. Langton, C. Taylor, and J. Farmer, editors, *Artificial Life II, Vol.X of SFI Studies in the Sciences of Complexity*, Redwood City, CA, 1991. Addison-Wesley.

W. M. Wheeler. *The social insects: their origin and evolution*. Kegan Paul, London, 1928.

I. White and S. Wise. Rule-dependent neuronal activity in the prefrontal cortex. *Exp. Brain Res*, 126:315–35, 1999.

E. Wilson. *The Insect Societies*. Belknap/Harvard University Press, Cambridge, 1971.

E. O. Wilson. *Sociobiology: The New Synthesis*. 1975.

A. Woodward, A. Phillips, and E. Spelke. Infants expectations about the motion of animate versus inanimate objects. In *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society*. Erlbaum, 1993.

C. D. Yang. Universal grammar, statistics or both? *Trends in Cognitive Sciences*, 8(10), 2004.

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

A. Zahavi. The cost of honesty. *Journal of Theoretical Biology*, 67:603–605, 1977.

R. Zemel, Q. Huys, R. Natarajan, and P. Dayan. Probabilistic computation in spiking neurones. In *Advances in Neural Information Processing Systems 17*. MIT Press, Cambridge, MA, 2004. in press.

B.-T. Zhang and H. Mühlenbein. Genetic programming of minimal neural nets using Occam's razor. In S. Forrest, editor, *Proceedings of the 5th International Conference on Genetic Algorithms, ICGA-93*, pages 342–349, University of Illinois at Urbana-Champaign, 1993. Morgan Kaufmann.

A. Zils and F. Pachet. Automatic extraction of music descriptors from acoustic signals using eds. In *Proceedings of the 116th AES Convention*, May 2004.

W. Zurek, editor. *Complexity, Entropy and the Physics of Information. Proceedings of the 1988 Workshop on the Complexity, Entropy and the Physics of Information*, 1990. Addison-Wesley, New York.