Institute of Computer Science
Academy of Sciences of the Czech Republic

Mirror Neurons, Embodied Cognitive Agents and Imitation Learning

Jiří Wiedermann

Technical report No. 894

July 2003

Pod Vodárenskou věží 2, 182 07 Prague 8, phone: (+420) 266 053 520, fax: (+420) 286 585 789,
e-mail:jiri.wiedermann@cs.cas.cz

Institute of Computer Science
Academy of Sciences of the Czech Republic

Mirror Neurons, Embodied Cognitive Agents and Imitation Learning[1]

Jiří Wiedermann

Technical report No. 894

July 2003

Abstract:

Mirror neurons are a relatively recent discovery; it has been conjectured that these neurons play an important role in imitation learning and other cognitive phenomena. We will study a possible place and role of mirror neurons in the neural architecture of embodied cognitive agents. We will formulate and investigate the hypothesis that mirror neurons serve as a mechanism which coordinates the multimodal (i.e., motor, perceptional and proprioceptive) information and completes it so that the agent remains always situated even when parts of the multimodal information are missing. We show that such a hypothesis forms a basis on which plausible explanation of the development of a host of mental abilities could be founded. These abilities range from imitation learning, communication via a sign language up to the dawn of thinking. Our results build a bridge between the theory of embodied cognition and mirror neurons; they also justify the hopes related to the discovery of mirror neurons.

Keywords:
embodied cognition, complete agents, mirror neurons, imitation learning, sensorimotor control

[1]This research was partially supported by GA ČR grant No. 201/02/1456.

1 Introduction

By the end of the twentieth century two important events happened in the field of cognitive science: one event was the change of the classical paradigm of cognition, the other was the discovery of mirror neurons. The classical paradigm sees cognition simply as computation. It considers mostly disembodied cognitive agents and concentrates mainly onto the information processing aspects of cognition; hence it almost completely neglects the embodiment and the necessity of controlling the sensory and motor activities related to the information extraction from the environment. On the contrary, the contemporary view of cognition sees cognition as a process deeply rooted in the body's interaction with the world. In the Nature, mind serves as a control system for biological bodies. In the world of artificial life, computation controls the artificial bodies — simulated or physical ones — of cognitive agents. The corresponding approach is called "embodied cognition". The respective ideas put the approach to cognition through the embodiment at the forefront of current research in not only AI but also in the whole cognitive science (cf. [2] or [7]).

Along with the above-mentioned silent revolution in cognitive science another discovery, this time in neurosciences, is apparently going to profoundly change our ideas of cognition. It is the discovery of so-called mirror neurons (cf. [9]) from 1990's. As V. S. Ramachandran, a prominent scientist in this field has put it, *"the discovery of mirror neurons in the frontal lobes of monkeys, and their potential relevance to human brain evolution is the single most important "unreported" (or at least, unpublicized) story of the decade. I predict that mirror neurons will do for psychology what DNA did for biology: they will provide a unifying framework and help explain a host of mental abilities that have hitherto remained mysterious and inaccessible to experiments"* [8]. Roughly speaking, the mirror neurons are neurons that fire if their owner performs a certain action as well if it observes the same species performing the same action. This can be interpreted as mirror neurons being a mechanism for "mind reading" of other subjects. Other researchers speculated on the existence of similar neurons also in primates and developed far-reaching conjectures on the importance of mirror neurons for understanding the intentions of other people, empathy, imitation learning and even for the language readiness (cf. [1], [6],[8]). Nevertheless, it seems that so far not much attention has been paid to mirror neurons in the context of embodied cognition: for instance, this notion is not even mentioned in a recent excellent monograph [7] on embodied cognition. Both the paradigm of the embodied cognition and the discovery of mirror neurons form a framework for the paper at hand. Note that for imitation learning embodiment is a sine qua non condition because the elementary goal in imitation learning is the ability to imitate an observed bodily movement.

The belief that cognition cannot be studied without a proper embodiment has lead to a so–called synthetic modelling approach (cf. [7]), stressing the idea of "understanding by building". In our approach we abstain from studying concrete physical embodiments of agents. Instead, we aim at a definition of embodiness that captures the main features of embodiment but still allows a "body independent" treatment of cognitive tasks. What we have here in mind is especially imitation learning and derived cognitive abilities, such as sign communication, empathy, and thinking which up to now have resisted to a formal and general approach. We conjecture and we also bring a plausible evidence that the key to all these mental tasks is offered by mirror neurons which serve as a mean for a coordination and, under some well defined circumstances, also for completion of the perceptive–motor information. The paper has a preliminary and exploratory nature: rather than pointing to ready–made efficient solutions it merely focuses to (the existence of) algorithmic principles by which the cognitive abilities could be explained.

In Section 2 we define the notion of a finite cognitive agent that will serve as a mean to explain the main ideas of so-called self-control and completing mechanism. In Section 3 we explain the role of this mechanism in basic modes of agents' activities that in a natural way enable imitation learning and elucidate empathy, elementary communication and eventually a rudimentary form of thinking. A neural architecture of the resulting model of a finite cognitive agent and a learning process that leads to development of the cognitive abilities is dealt with in Section 4. In Section 5 we sketch how in our framework an evolutionary development of mind could possibly be explained. The closing Section 6 stresses the main contributions of the paper.

2 Finite Cognitive Agents

In order to carry forth our ideas of cognition we need a more precise definition of a cognitive agent than a verbal definition saying "it is a human, animal, or artificial creature like robots or simulated organisms" (cf. [7]).

Definition 2.1 *A finite cognitive agent (FCA) consists of*

- *a finite set of perceptional–motor units (PMUs)*
- *a finite state transducer that translates a potentially infinite stream of sensory and proprioceptive data as delivered by the perceptional parts of agents' PMUs into a potentially infinite stream of motor data that are sent back to the motor parts of agents' PMUs.*

Thus, a finite state transducer represents the "brain", or the "mind" of a FCA whereas the PMUs represent its "body". As its name suggests each PMU is a combination of a perceptional and motor unit (similar to Turing machine's read/write head). This unit sends perceptive information obtained from the unit's sensory organs to the transducer and receives from the transducer *motor instructions* for the moves of the PMU (and thereby, of the entire agent). The perceptional information is twofold. The first kind is *sensory information* from the PMU's external sensors reacting to the properties of the external world, delivering data (visual, auditory, haptic, electromagnetic, etc.) on the environment. The second kind is *proprioceptive information* obtained from the PMU's internal sensors delivering data related to the PMU's internal state, such as muscle tonus, status information such as a feeling of pain, hunger, voltage level, etc. Thus, proprioception also corresponds to internal feelings that are related to agents' motor action performed at that time. All three kinds of information — motoric, sensory and proprioceptive — will be jointly termed as *multimodal information*. Note that within each PMU, proprioceptive information provides a feedback — so-called *internal feedback* — for the corresponding motor instruction.

An example of a PMU is a camera, or an eye. Oculomotor instructions correspond to motor instructions (the direction of the view, lens adjustment, diaphragm setting, etc.), visual information from the retina corresponds to the sensory information and finally the proprioceptive information comes from the internal eye sensors informing about the physical (muscle) effort that is necessary for a particular setting of the eye's control system. Similarly, a hand can also be seen as a PMU. Motoric signals are the instructions for various muscle groups in the hand, sensory information are haptic signals from the hand, and we also have proprioceptive signals from individual muscle groups. In robotic systems a wheel is also a PMU: a motor signal could be e.g. "full speed forward" while a proprioceptive information could be "the wheel does not rotate", or "the pressure in the tires is too low". There seems to be no sensory information in this case.

An FCA can move in the environment with the help of its PMUs (such as wheels, legs, wings, etc.) specialized for locomotion and can modify the environment with the help of other PMUs (hands, arms, tentacles, etc.). The environment itself is not a subject of our modelling. A PMU can provide a feedback to another PMU. A typical example is a hand–eye system. If the agents' eye "observes" its own hand we say that there is an *external feedback* between the respective PMUs. We also say that both PMUs are coordinated.

The activity of an FCA is defined with the help of a *transition function* that, given the current state, currently issued motor instructions for the motor parts of all PMUs and current information from perceptional parts of all PMUs assigns a new state and a new set of instructions for motor parts of all PMUs. Formally, given an FCA $\mathcal{A}$ with k PMUs and denoting Q a finite set of states of $\mathcal{A}$, M a finite set of instructions, S a finite set of sensory information, and P a finite set of proprioceptive information, then the transition function δ takes the form $\delta : Q \times (M \times S \times P)^k \to Q \times M^k$. We assume that sets M and S contain distinguished symbols, corresponding to the case when "no motor instruction has been issued", or "no sensory signal has been received", respectively. On the other hand, we assume that some ("nonempty") proprioceptive information is always there. The schema of an FCA is in Fig. 1.

Note that an FCA as a whole does not compute in the classical sense of this word: neither its input nor its output are data. The agents' "input" is the external world that is perceived by the

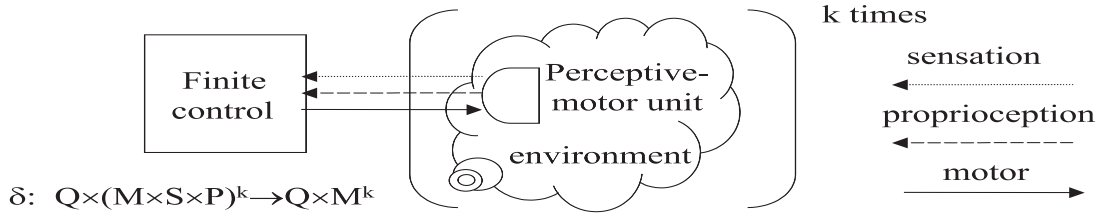

Figure 2.1: The schema of a finite cognitive agent

agent via its sensory units. It is the task of these units to produce data. The “quality” of this data depend on the kind and resolution of the agents’ sensors. The agents’ “output” is the agents’ behavior. Behavior’s appropriateness and effectiveness depend, in addition to the agents’ control, on the quality of the agents’ effectors. Thus, the “mapping” capturing agents’ activity cannot be characterized mathematically as long as we do not have a mathematical description of the environment and a mapping realized by agents’ PMUs. That is why in general the notion of a cognitive task cannot be defined formally. Informally, a cognitive task is defined from the observer’s point of view. In a given environment a *cognitive task* for an FCA calls for performing a sequence of interactions by which, in the eye of the observer, the FCA “passes” the test related to the task at hand. For instance, the Turing test is a test for a cognitive task called “intelligence”.

We say that an FCA $\mathcal{A}$ is *embodied* in environment E w.r.t. a cognitive task τ if and only if $\mathcal{A}$ is equipped by such PMUs that provide $\mathcal{A}$’s control unit with enough perceptive information and sufficient motor abilities to perform task τ in environment E. In this case we say that $\mathcal{A}$’s control is *situated* in E through agents’ PMUs.

3 Basic Modes of an Agents’ Activities

Both internal and external feedbacks are important means for the FCA’s self-control mechanism. For simplicity, consider an FCA $\mathcal{A}$ with two PMUs — a hand and an eye. If in addition equipped by wheels for moving in an environment such agents are called *fungus eaters* in the robotic literature (cf. [7]). Assume that $\mathcal{A}$ is programmed so as its eye always observes its hand. Thus, hand and eye movements are coordinated via the external feedback. To ensure this in the $\mathcal{A}$’s transition function δ the respective hand and eye move instructions must be coupled. Moreover, with each of the former moves there is a proprioceptive information paired with them via the internal feedback. Thus, in δ the corresponding move instructions must also be associated with the respective proprioceptive information. The mode of work of $\mathcal{A}$ in which hand and eye movements are coordinated and the proprioception corresponds to these coordinated movements, is called a *standard mode.* Clearly, any deviation from the standard mode will be indicated by a mismatch in the associated multimodal information. This means that an FCA with a properly formed transition function can detect deviations from the standard mode and can take appropriate measures. Hence, the task of the *self-control mechanism* is to see whether the “multimodal arguments” of δ which are elements of $(M \times S \times P)^k$, correspond to the standard mode.

To simplify the following discussion assume that a multimodal information related to a PMU, of form $(M \times S \times P)$ in the standard mode is uniquely determined by any of its three components[2]. Similarly, if there is an external feedback between a pair of PMUs, then we will assume that the pair is also uniquely determined by either member of this pair. It follows that in situations when some information is missing in the (associated) multimodal information, the “original” multimodal information that would correspond to the standard mode can always be reconstructed. The respective mechanism that completes the incomplete multimodal information will be termed as a *completing mechanism.*

[2]This is an important condition that, if not satisfied, must be detected and restored. Therefore a learning algorithm should include also motor aspects, such as e.g. moving the agent so as to “better see” what is needed in order to satisfy the above mentioned condition.

Consider now the typical situations in which a part of multimodal information is missing. The first such a situation is observing movements of other agents. Imagine two identical fungus-eater agents, A and B, who in the standard mode, under the same conditions, behave in the same way. Now assume that in the standard mode, A performs a particular cognitive task τ calling for moving A's arm in a specific way — e.g grasping a mushroom. Since this happens in the standard mode, A "observes" its own move by its eye. Let (m_1, s_1, p_1) be the respective multimodal information corresponding to τ. Now assume that A is observed by B on this occasion. If it happens that B in state q sees A's arm grasping a mushroom from the same perspective as B would see its own arm grasping a mushroom, then the multimodal information that B gets is of form (m_2, s_1, p_2), with both $m_1 \neq m_2$ and $p_1 \neq p_2$. That is, both current motor and proprioceptive information of B does not correspond to sensory information currently obtained, w.r.t. the standard mode. This "mismatch" indicates that there is a deviation from the standard mode. Then it makes sense to design the transitive function δ of B (or analogously, that of A, since A and B are identical) so as in the next step, δ "corrects" the current motor information by "assuming" that the motor information was in fact m_1. The mode of agents' activity in which the motor instructions that match the observed movement are "filled in" is called *observational mode.* Thus, the respective transition could read as $\delta(q, m_2, s_1, p_2) = (q_{observed}, m_1)$, with $(q_{observed}, m_1)$ really having the semantic *"not–me is observed making move m_1"* (or even, when also s_1 is taken into consideration, *"not–me is observed performing τ"*). Depending on δ B can now e.g. imitate A by repeating move m_1. Doing so B also "reconstructs" the proprioceptive information of A since performing m_1 will "return" proprioception p_1. This is where the mechanisms of imitation and empathy could be rooted. Note that the same mechanism is also responsible for the identification of a known movement (e.g., a gesture). Therefore it can also serve as a platform on which a simple communication between agents can be established. If fully developed this can lead to the use of a sign or body language. Finally note that the distinction between the standard and observational mode leads to differentiation between "self" and "non-self". Formally, in the definition of an FCA this leads to the "split" of set Q into two classes of states: those corresponding to the own activities of the agent and those corresponding to the activities of other agents.

Another occasion when agents deal with incomplete multimodal information is the case when its both perceptive and sensory parts of multimodal information are missing. Imagine that an agent in the standard mode issued a motor instruction m_1 but was prevented (e.g. by some external reason) both from realizing it and also from "seeing" it. As before, a completing mechanism could be designed so as to fill in the missing parts to get (m_1, s_1, p_1), say. Thus, the agent can proceed as if "nothing had happened," and such proceeding can be prolonged over several steps. However, in its states the agent can keep the information that all related activities (except of the move instructions) are faked, corresponding only to some "virtual" reality. An agent (i.e., its transition function) can even possess "false" move instructions whose purpose is not to instruct the PMUs to perform some activity, but to only "invoke" the information completing mechanism. This is where the prologue of thinking could start. Again, as it was the case with the observational mode, an agent can realize its entering a thinking mode; this again leads to the formation of a distinguished class of the respective states within Q.

Thus, based on the completion mode of multimodal information we can distinguish at least three important classes of cognitive states (elements of Q): standard, observational, and thinking states. Note that from the viewpoint of situatedness, in any mode an agent is either really or virtually situated.

4 Place and Role of Mirror Neurons in the Neural Architecture of an FCA

As we have seen in the previous section a properly designed transition function that also implements the self-control and completing mechanism could lead to a cognitive behavior with hallmarks of a rudimentary intelligence. Of course, it is hard to imagine a "manual" design of such a transition function. The following scenario would be very helpful in an automatic design of the computational mechanism of an FCA. First, the agent will learn the correct behavior (i.e., the proper association of multimodal information) in the standard mode. Based on this, in the next phase of learning the agent will acquire the proper handling of situations in which the multimodal information deviates

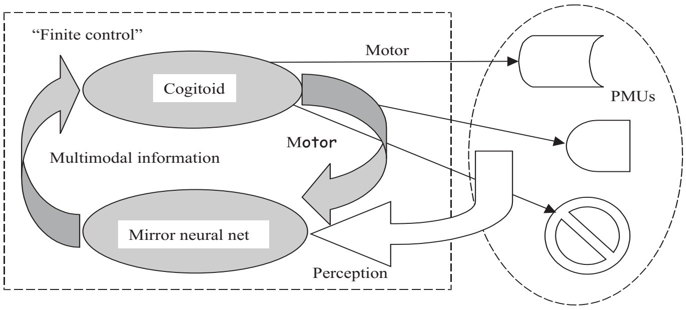

Figure 4.1: The schema of the neural architecture of a finite cognitive agent

from that learned in the standard mode. That is, only in the second phase the self-control and the completing mechanisms utilized in the observational and thinking mode will be formed. Neural nets offer a suitable substrate for such a purpose. Therefore we will consider a neural net implementation of an FCA.

No doubts that any finite state control can be realized with the help of a neural net (cf. [11]). An important part of this net will be the self-control mechanism consisting of the mirror neurons. First we describe how this mechanism is configured into a state in which it works in the standard mode. Then we sketch how it can be extended to realize also the multimodal information completing mechanism.

As we have already stated the self-control mechanism keeps track of the "right" composition of arguments of the transition function. Namely, the tuples of values from $Q \times (M \times S \times P)^k$ should correspond to the tuples occurring in the standard mode. The correct tuples are acquired in the first learning phase by a neural net during the agents' interaction in the standard mode. Let $(q, m_1, s_1, p_1, \ldots, m_k, s_k, p_k)$ are $(3k+1)$ values of such arguments in state q. The neuron checking the simultaneous occurrence of such values of arguments is simply a neuron realizing a Boolean conjunction of $3k+1$ variables corresponding to the given $(3k+1)$ tuple with arguments written in the binary notation. In the neural net there must be a dedicated neuron — let us call it *a mirror neuron* — for each combination of arguments that can occur in the standard mode. Instead of a single neuron we can imagine a group of neurons fulfilling the given task. For instance, the information from sensors passes first through a neural circuit that recognizes the "eligibility" of the observed movement and only then does the information enter into another circuit that verifies the proper "composition" of the entire multimodal information. Thus, a picture of a neural net emerges in that the following data enter: information on the current state, motor instruction issued for each PMU, and sensory and proprioceptive information received from each PMU. This information is funnelled into all mirror neurons in parallel and one of them will react to it as to its "own" information which it has been trained (specialized) to. From the learning point of view this network must be able to learn a disjunction of conjunctions, each conjunction consisting of the same fixed number of variables.

Now let us proceed to the second phase of learning. Consider the case when 'incorrect" (i.e. not corresponding to the standard mode) information enters the network formed during the first phase. For definiteness consider the case when in an observation state there is a mismatch identified between motor and proprioceptive information on one hand and the sensory information on the other hand. In this situation that particular mirror neuron should react that would be active when the agent itself would perform the observed movement in the standard mode. This can be achieved in several ways. Assuming again that each multimodal information in the standard mode is uniquely determined by any of its components and the current state, then we can make use of parallel associative memory search abilities of neural nets. We will not present the resulting net in detail, what we merely want is to point to the fact that such a mechanism does exist and we can imagine its construction. The scheme of the resulting neural net is depicted in Fig.2.

Now we show that the properties of "our" mirror neurons match the properties of mirror neurons described in the literature (see their description in the introductory part of this paper). Assume that the control sends a motor instruction both to the arm and the eye of an FCA in the standard mode. In the respective mirror neuron m these instructions will meet the signals from the arm's and eye's perceptive sensors and because all this happens under the standard mode m will fire. This means that m is active if the agent performs a movement and "checks" (watches) it via the external sensory–motor feedback. Consider now the situation when agent A observes an arm of agent B performing the same movement. The completing mechanism corrects the mismatched multimodal information to a correct one which mirror neuron m will react to.

A final remark: instead of a net consisting of a number of individual mirror neurons one can think of a probably smaller neural net that is able to learn and recognize the necessary self-control multimodal information, at the same time possessing the associative completing ability. Note that such a net can be seen as an implicit model of the word in which the agent exists. The "syntax" of this world is represented by the set of all multimodal information kept in the mirror neurons. That is, the agent has at its disposal a schema describing the world as perceived by its senses and proprioception and as explored by its motor activities. This is where the agents' mechanism for situatedness is rooted. Should a part of the world in this schema be missing then the agent either did not enter this part of the world or simply it is not able to explore it via its perceptive or motor abilities. In [3] a similar conclusion of the existence of an implicit world model in the cognitive agents that is realized with the help of a neural net was also reached. Our ideas go beyond those concerning the importance of the sensory–motor information for the classification of objects, summarized e.g. in [7], because we have proposed a more concrete and experimentally verifiable real mechanism — namely mirror neurons — on which such a model can be based.

5 A Sketch of the Evolutionary Development of the Agent's Mind

Consider now a hypothetical evolutionary development of a cognitive agent which would lead to the formation of its control abilities that would resemble the mind. Doing so it is natural to assume a development passing through a number of evolutionary stages in which mechanisms taking care of the agents' activities aiming at its survival in a corresponding ecological niche will be formed. Clearly, the agents' physical development should go hand in hand with its mental development, both developments also corresponding to possible changes in the agents' living environment. Both developments must be in the so–called *ecological balance*: it does not make sense for an agent to have PMUs delivering more information than its brain can handle, or vice versa, to have a more efficient brain than the one that is sufficient to process the data delivered by the agents' PMUs (cf. [7]).

The so-called *grounding* of an agent in the environment is the basic assumption of the cognition (cf. [5]). This means that an agent must have information available how to handle the environment by means of its actions; this information is obtained via its PMUs. Technically this information is represented as multimodal information stored in the agents' mirror neural net. Grounding thus takes care about agents' elementary understanding the world. Note that at the level of elementary actions an agent cannot do any but "reasonable" actions, since others are not a part of its repertoire, other actions than "useful" ones are not acquired by the agent. For further development an agent needs a faculty of imitation learning. This ability is provided by the self-control and multimodal information completing mechanism. Apparently this developmental stage differs from the previous one since there are creatures not possessing even rudimentary imitation abilities (e.g. such creatures do not have sensory organs enabling them to "see", in order to mirror the movements of other creatures). The ability of mirroring leads to the emergence of the self-concept that is necessary in order to distinguish between the actions observed by one's own sensors on one's own body and the actions of other agents. What develops next is the ability to learn sequences of actions, still with the help of imitation. The mechanism that takes care of these and other abilities, such as learning by analogy, Pavlovian reflexes, operant conditioning, emotion processing, etc., is denoted as a cogitoid in Fig. 2. Note that the cogitoid is not directly connected to any peripheral input device — in fact it is a classical information processing device that has been considered in the computational paradigm of the classical AI. The input into a cogitiod is formed by a stream of preprocessed, corrected or completed multimodal information streaming from a mirror neural net. This data reflects the outer world and causes that a cogitoid is specialized solely to process situated data. What a cogitoid does is "computing" the transition function, i.e. the "next move", given the current state and multimodal information. A cogitoid is designed so as to be able to learn from experience, in various modes (in a supervised or unsupervised mode, by trial and error, positive or negative reinforcement, etc.) and their combinations. In its most developed form a cogitoid learns with the help of thinking — by "mentally" simulating possible scenarios of future situation development based on experience. This simulation consists od steps, each step realizing one transition of δ which is technically realized as one circulation of multimodal information between the mirror net and the cogitoid. Hence, in the thinking mode the multimodal information circulates in the network without any external interference. The cogitiod's mode of work is modulated by emotions. For more details on cogitoid see [10]. Going hand in hand with the evolution of learning mechanisms is the evolution of communication abilities and empathy. Communication is very simple initially, based on visual signaling via gestures, in some cases perhaps based on olfactory signals, and is related to emotion or specific context transfer. The signalling repertoire gradually builds up and new communication channels, especially the acoustic ones, enter the game. By this the limbs used for gestures are freed again to other bodily activities. A multimodal combination of all signals is always built with the key role played by the motor activities. The importance of visual communication decays and the main role is taken over by a vocal communication. The role of motor activities shifts from the gesture and body language towards the organs of speech. The vocabulary of notions keeps developing while keeping them grounded in the perceptive-motor multimodal information. The notions (concepts) and emotions can be activated by hearing the corresponding word(s), not only with the help of sensors. It could be the case that there is also a proprioception caused by word activation and that this proprioception gets associated with further multimodal information related to the words. The spoken language enables a further discrimination of objects, a further structuring of the world since an agent distinguishes more object categories. It is as though the perception organs started to deliver more distinguishable information then before making use of the same channels. What follows is a further evolution of the brain that has to cope with this blow up of information. Agents are now able to communicate among themselves and also one with itself by speaking to oneself. An agent speaking to itself starts to "think loudly". In the further evolution the vocalization accompanying thinking is slowly loosing its importance, the movements of the speech organs decay until only motor signals (especially those for the speech organs) prevail that, however, are no longer realized but are still associated with perceptive and proprioceptive information in which the motoric is grounded. From the viewpoint of its perception an agent finds itself in a virtual environment mediated this time not by its senses but by its completion mechanism (which was formed by the agents' senses in the standard mode). The information on its virtual situatedness has the agent to its disposal and maybe that here are the roots of the agents' consciousness. Namely, it can happen then that an agent becomes a subject of its own observation: in the observational mode the agent is informed about its being in a thinking mode. Could it be the case that the respective proprioceptive information could give rise to the agents' feeling of consciousness?

The evolution of agents' mental abilities sketched above roughly corresponds to similar ideas of other authors, especially those of philosophers of mind (cf. [4]). The difference is that in our case our ideas are supported by concrete mechanisms that are taking care of individual evolutionary stages of the mental development of cognitive agents within a single model of an FCA.

6 Conclusion

The ideas on the mechanisms of thinking in FCAs originate from a single conjecture. It is a hypothesis that the mirror neurons are a part of a mechanism that serves for learning, self-control (verification) and eventually completion of multimodal information. This mechanism was generalized to a mechanism also profiting from proprioceptive information. Proprioceptive information serves here not only as an additional mean for multimodal information grounding and for increasing the robustness of multimodal information against its possible incompleteness, but also as a basis of internal sensation. The mechanism of multimodal information completing, further enabled bringing a seemingly unrelated perception-motor (i.e., standard) and thinking modes onto a common basis. In our model thinking takes a form of suppressed motor actions which are from the agents' viewpoint accompanied by virtual sensations provided by the completion mechanism. A proprioception of this state of the mind could open the door to consciousness.

The proposed theory of thinking is in fact a theory of realization of thinking. The principles presented here have been, to a various extent, described and envisaged by many philosophers, psychologists and people from the field of artificial intelligence and cognitive science. Until now a plausible computational framework enabling a consistent explanation of cognitive mechanisms behind thinking has been missing. Mirror neurons and a neural implementation of finite cognitive agents proposed in this paper present a possible computational framework for the theory of cognition we are after.

Bibliography

[1] Arbib, M. A.: The evolving mirror system: a neural basis for language readiness. In M.H. Christiansen and S. Kirby, editors, Language Evolution: The States of the Art. Oxford University Press, 2003.

[2] R. Brooks: Cambrian Intelligence: The Early History of the New AI. MIT Press (A Bradford Book), 1999

[3] Cruse, H.: The evolution of cognition — a hypothesis. Cognitive Science 27, Elsevier, 2003, pp. 135-155

[4] D. C. Dennett: Consciousness Explained. Penguin Books, 1991, 511 p.

[5] Harnad, S.: The Symbol Grounding Problem. Physica D 42, pp. 335–346, 1990

[6] J.R. Hurford: Language beyond our grasp: what mirror neurons can, and cannot, do for language evolution. In: O. Kimbrough, U. Griebel, K. Plunkett (eds.): The Evolution of Communication systems: A Comparative Approach. The Viennna Series in Theoretical Biology, MIT Press Cambridge, MA, 2002

[7] Pfeifer, R., Scheier, Ch.: Understanding intelligence. The MIT Press, Cambridge, Massachusetts, London, England, 1999, 697 s.

[8] V.S Ramachandran: Mirror neurons and imitation as the driving force behind “the great leap forward” in human evolution. EDGE: The third culture, viz http://www.edge.org/3rd_culture/ramachandran/ramachandran_p1.html

[9] G. Rizzolatti, L. Fadiga, V. Gallese, I. Fogassi: Premotor cortex and the recognition of motor actions. Cognitive Brain Research, 3:131-141,1996

[10] J. Wiedermann, J.: Towards Algorithmic Explanation of Mind Evolution and Functioning (Invited Talk). In: Proc. of the 23-rd International Symposium on Mathematical Foundations of Computer Science, LNCS Vol. 1450, Springer Verlag, Berlin, 1998, pp. 152–166 .

[11] J. Wiedermann: The Computational Limits to the Cognitive Power of the Neuroidal Tabula Rasa. Journal of Experimental and Theoretical Artificial Intelligence (JETAI), in print, 2003