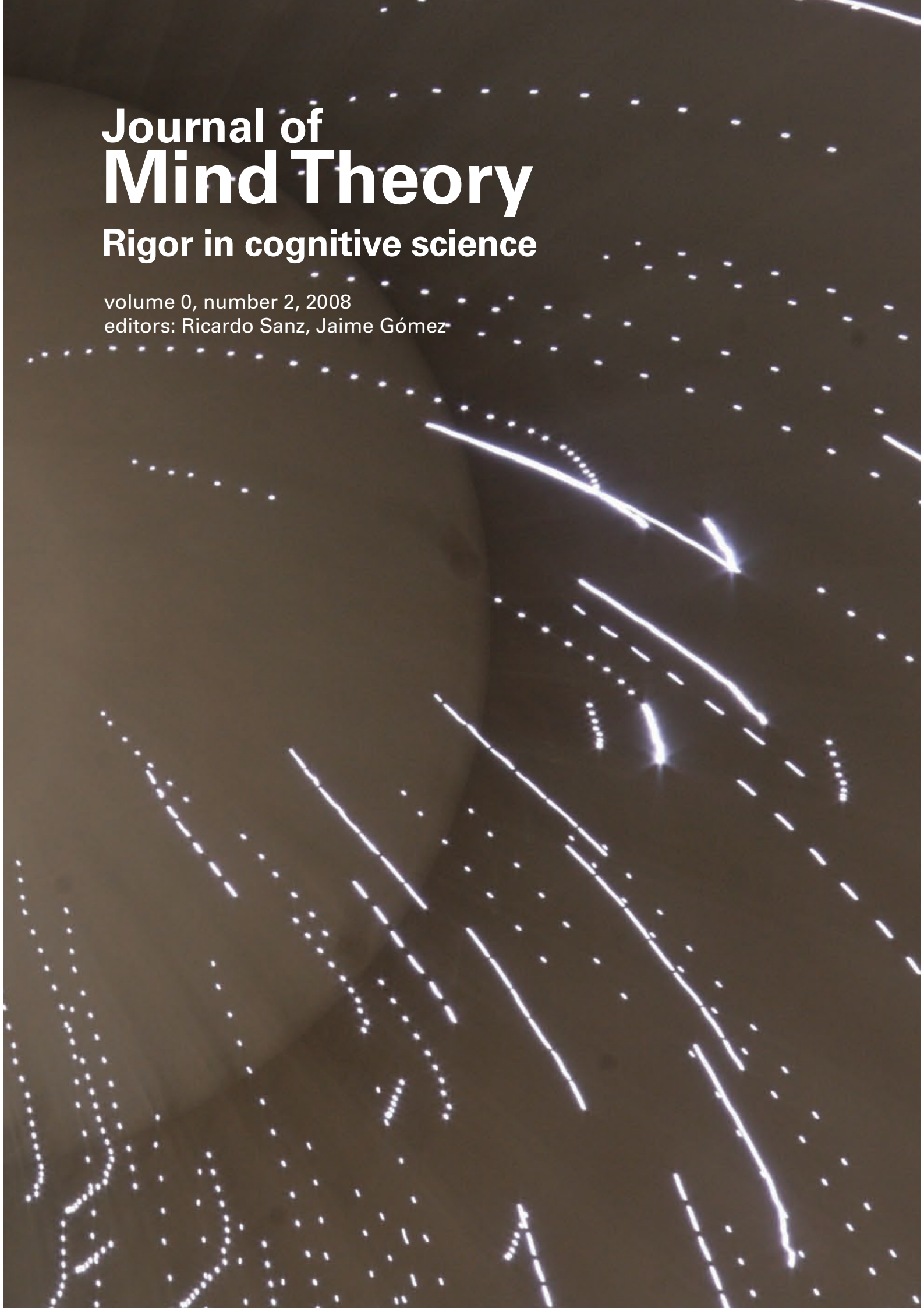

Journal of Mind Theory

Rigor in cognitive science

volume 0, number 2, 2008
editors: Ricardo Sanz, Jaime Gómez

Journal of Mind Theory

Journal scope and Objectives
The Journal of Mind Theory aims to stress a rigourous approach to the investigation of the mind. It is driven by the widespread scientific view that intentions, thoughts and feelings are just natural phenomena and therefore can and must be explored within a strict scientific framework encompassing both theoretical and empirical concerns.

- JMT seeks theoretical rigor in theories of mind.
- JMT seeks contributions that transcend the traditional disciplinary boundaries in cognitive science, encouraging articles from researchers interested in a formal approach to the analysis of cognition.
- JMT emphasizes the synthesis of ideas, constructs, theories, and techniques in the analysis of biological cognition and in the design of cognitive autonomous systems, offering a platform for addressing the problem of formalization of cognition from a systemic and naturalized perspective.
- JMT coverage includes the classic topics of theory of mind but with a formal tint: perception and phenomenology, theory of knowledge, reasoning and causation, the role of mathematics and logic in cognitive systems and philosophical foundations of cognition.
- JMT accepts experimental work insofar it addresses specific theories.
- JMT looks for fresh thinking, vigorous debate, and careful analysis!

Submissions
Submissions for JMT shall be directed to the editors following the guidelines available in the journal website:

http://www.aslab.org/JMT

Cover image: Asís G. Ayerbe (www.lacarreteradelacosta.com)

ISBN Volume 0 (whole volume): 978 84 613 3057 7
ISBN Volume 0 Number 1: 978 84 613 3020 1
ISBN Volume 0 Number 2: 978 84 613 3021 8

Programa de Apoyo a Grupos de Investigación del IV PRICIT, dentro del Contrato Programa Marco entre la Administración de la Comunidad de Madrid y la Universidad Politécnica de Madrid para la regulación del marco de cofinanciación en el Sistema Regional de Investigación Científica de Innovación Tecnológica.

Table of Contents

JMT Vol. 0 No. 1

JMT Vol. 0 No. 2

About the Authors

James S. Albus founded and led the Intelligent Systems Division at the National Institute of Standards and Technology for 20 years. During the 1960's he designed electro-optical systems for more than 15 NASA spacecrafts. During the 1970's, he developed a model of the cerebellum that is still a leading theoretical model used by cerebellar neurophysiologists today. Based on that model, he invented the CMAC neural net, and co-invented the Real-time Control System (RCS). RCS is a reference model architecture for intelligent systems that has been used over the past 25 years for a number of systems, and the latest version of the RCS architecture has been selected by the Army for the Autonomous Navigation Systems to be used on all Future Combat System ground vehicles –manned and unmanned. He has worked with DARPA and other government agencies on a concept for a National Program for Understanding the Mind: "Decade of the Mind". Now he has retired from NIST and is associated to the Krasnow Institute for Advanced Studies

Sarah Rebecca Anne Belden received a B.A. in Art History from New York University and a M.A. in Art History, Connoisseurship and the History of the Art Market from Christie's Education in New York. She also recently completed a curatorial residency program at Konstfack University in Sweden. Since graduating, Miss Belden has worked within the art world in New York at Christie's Auction House at Rockefeller Center, at several commercial art galleries, and within the non-profit sector. In 2006, Miss Belden relocated to Europe where she opened Curators Without Borders, a contemporary art space dedicated to supporting independent curators and promoting emerging artists in Berlin. Miss Belden has curated several important exhibitions in Berlin, Bergen, Athens, and Copenhagen and is now working on several projects related to Neuroaesthetics, Politics and the Arts.

Ron Cottam received his first degree and PhD in Applied Physics from the University of Durham, UK, and in 1971 he transferred to the Department of Metallurgy at the University of Leuven, Belgium. Moving away from academic work, he spent twelve years in commercial organizations and as an independent consultant developing techniques for the enhancement of audio presence in music reproduction. He joined the Department of Electronics and Informatics of the Vrije Universiteit Brussel (VUB) in 1983, where since 1984 he has been a member of the Laboratory of Micro- and Photonelectronics, associated with work on chemical sensors, optical computing, computational theory, and most recently since 1991 on the development of architectures for the implementation of lifelike processes in ULSI beyond 2020. His research specialties are natural birational hierarchy and the establishment of criteria for real intelligence and consciousness in artificial systems. Ron leads the VUB Evolutionary Processing Group (EVOL) and its Living Systems Project, and has authored and co-authored papers on solid-state physics, ultrasonic techniques, computational emergence, natural semiotics, hierarchical evolutionary systems, complexity and anticipatory computation, in many conferences, journals and books. In addition to his research work he teaches at Vesalius College of the VUB and runs an acoustics consultancy and music recording studio.

Andrée Ehresmann is Emeritus Professor at the "Université de Picardie Jules Verne", and Director of the international Journal "Cahiers de Topologie et Géométrie Différentielle Catégoriques". In 50 years of mathematical research she has published about a hundred papers on Functional Analysis and Category theory and edited and commented the 7 volumes of "Charles Ehresmann: Oeuvres complètes et commentées". Since 25 years she has developed with J.-P. Vanbremeersch the theory of Memory Evolutive Systems.

Jaime Gómez is currently an Assistant Professor and Research Scholar at the Universidad Politécnica in Madrid in the Autonomous Systems Laboratory. Prior to this, he received his degree in Computer Science and worked for several years as a consultant in France, and as team leader in Spain for major technology companies. In 2004, he returned to Academia, where he currently acts as an Assistant Professor in Robotics. In 2006 he was visiting researcher at the University of California, Berkeley and during 2008 he continued this research at Humboldt University in Berlin. He has published several papers on such subjects as Cognitive Ontologies, Learning in technical systems, and Naturalized Epistemology for Autonomous Systems. Jaime Gomez is currently completing his PhD, in the construction of a formal theory of cognition that provides a prescribed structure designed to outline the basic cognitive processes.

Pentti O A Haikonen received the M.Sc. (EE), Lic. in Tech. and Dr. Tech. degrees from the Helsinki University of Technology, Finland, in 1972, 1982 and 1999 respectively. Haikonen is presently full adjunct professor at the University of Illinois at Springfield, Department of philosophy. Previously Haikonen was principal scientist, cognitive technology at Nokia Research Center, Finland 1991 – 2009. Haikonen has authored the books "Robot Brains; Circuits and Systems for Conscious Machines" (UK: Wiley & Sons, 2007) and "The Cognitive Approach to Conscious Machines" (UK: Imprint Academic, 2003). Haikonen has 14 patents on signal processing, associative neurons and networks. Haikonen's interests include the theory and philosophy of machine cognition, electronic circuitry for cognition and the design of exotic electronic gadgets.

Lorenzo Magnani, philosopher and cognitive scientist, is a professor at the University of Pavia, Italy, and the director of its Computational Philosophy Laboratory. He has been visiting professor at the Sun Yat-sen University, Canton (Guangzhou), China and has taught at the Georgia Institute of Technology and at The City University of New York. He currently directs international research programs in the EU, USA, and China. His book Abduction, Reason, and Science (New York, 2001) has become a well-respected work in the field of human cognition. In 1998, he started the series of International Conferences on Model-Based Reasoning (MBR). The last book Morality in a Technological World (Cambridge, 2007) develops a philosophical and cognitive theory of the relationships between ethics and technology in a naturalistic perspective.

Willy Ranson received the Telecommunication Engineer degree in 1975 from the University of Leuven, Belgium. He was Assistant Professor in the Department of Microwaves and Lasers at the University of Leuven until 1983, when he joined the Department of Electronics and Informatics (ETRO) of the Vrije Universiteit Brussel (VUB). Willy has participated in projects and contracted research on such diverse topics as planar antenna structures, high frequency wave-guides, chemical sensors, biological applications for breast cancer detection, optical information processing for parallel computation, CO_2 laser applications, microelectronic process technology and revolutionary information and computation theories. He is currently Senior Researcher in charge of the processing technology lab of LAMI and is a founder member of the Evolutionary Processing Group (EVOL). His current research contributions are in the areas of CO_2 laser modulation, millimeter imaging systems, micro machines for ultra-rapid DNA screening, fast enforcing technologies for protein engineering and Evolutionary Living Systems. Willy is (co)author of more than 100 publications in international refereed journals and conferences.

Tariq Samad is a Corporate Fellow in Honeywell Automation and Control Solutions and the 2009 President of the IEEE Control Systems Society. Dr. Samad received a B.S. degree in Engineering and Applied Science from Yale University and M.S. and Ph.D. degrees in Electrical and Computer Engineering from Carnegie Mellon University. He has been with various R&D organizations in Honeywell for 23 years, contributing to and leading automation and control technology developments for applications in unmanned aircraft, electric power systems, the process industries, building management,

automotive engines, and clean energy. His publications also include one authored and three edited books, most recently *Software-Enabled Control: Information Technology for Dynamical Systems* (G. Balas, coeditor; Wiley, 2003). Dr. Samad was editor-in-chief of *IEEE Control Systems Magazine* from 1998 to 2003 and is a Fellow of the IEEE and the recipient of an IEEE Third Millennium Medal, a Distinguished Member Award from the IEEE Control Systems Society, a Neural Networks Leadership Award from the International Neural Networks Society, and the 2008 IEEE CSS Control Systems Technology Award.

Ricardo Sanz is professor in Automatic Control and Systems Engineering at the Universidad Politécnica de Madrid, Spain and coordinator of a research group on autonomous systems (www.aslab.org). His main research topic is advanced control architectures for technical systems. His work sits the frontier between control, computing and intelligence –automatic control, artificial intelligence, embedded systems, real-time distributed systems, software engineering, and cognitive systems. He has been involved in many national and international research projects in the field of real-time distributed systems and complex intelligent controllers. He is co-chairman of the International Federation of Automatic Control Technical Committee on Computers and Control.

Konrad Talmont-Kaminski is at the Marie Curie-Sklodowska University in Lublin, Poland. His research focuses on developing a naturalized account of rationality, with his recent work examining superstitions as a natural, cognitive phenomenon within the context of recent work on the evolution of religion. He argues that superstitions are by-products of cognitive heuristics, rendered more stable by the (usually post hoc) addition of supernatural explanations. Furthermore, numerous superstitions, as well as the mechanisms underlying them, have come to be exapted by religions that add to them a social dimension. The final result is a socially powerful phenomenon whose flexibility and functionality is largely due to its claims having become largely detached and protected from reality

Jean-Paul Vanbremeersch is a physician with a specialty in Geriatric who has both a liberal practice and a coordinator role in a old people's home. He has long been interested in explaining the complex responses of organisms to illness or senescence. Since 1984 in joint work with A.Ehresmann, has developed the Memory Evolutive System model for natural complex systems, such as biological, cognitive, social or cultural systems; it is developed in their recent book, summarizing about 30 research papers.

Roger Vounckx started his career as a teaching assistant in the Physics Department of the VUB's Faculty for Sciences from 1975 to 1980. In 1981 and 1982 he was a visiting scientist and acted as a consultant for AT&T Bell Laboratories, Murray Hill, New Jersey (USA), working on exploratory high speed III-V semiconductor transistors, and was awarded the Dr.Sc. degree in physics in 1984. He was appointed associate professor of microelectronics at the VUB in 1984, full professor in 1993, became director of the Laboratory of Micro- and Photonelectronics (LAMI) in 1987 and was appointed head of the Electronics and Informatics Department (ETRO) in 2008. His current research interests include semiconductor devices and systems for optical and electrical information processing and communication and mm wave imaging systems. He has published over 250 technical papers in international journals and conference proceedings, holds 8 international patents, and serves regularly as an expert for evaluation of industrial research projects for the Belgian Government. He is a co-founder and an executive director of EqcoLogic nv, which designs and produces silicon chips for fast data communication.

Vindication of a Rigorous Cognive Science

Ricardo Sanz and Jaime Gómez

Universidad Politécnica de Madrid

Abstract

The study of mind seems to be in an impasse due to its elusive nature and the inherent difficulties emerging from the sheer complexity of its main realization: the brain. Advance will be possible, however, if we are able to apply the simple method of science: get data, formulate a theoretical hypothesis, and test the hypothesis. In the current state of affairs there is a lack of systematicity in the formulation of the hypotheses and we feel one of the reasons is the lack of an adequate vehicle. In this introductory article we expose the reasons for creating yet another periodic publication in the domain of cognitive science: The Journal of Mind Theory.

1 Motivation

The multidisciplinary nature of the cognitive science endeavour makes it difficult to consolidate theoretical approaches into widely understandable, testable and eventually universaly accepted theories that can serve as cornersonets of a solid science and technology of mind.

In this context we are launching a new forum for theoretical discussion in the form of a journal on mind theory. We all realize that the number of publications in the field of cognitive science is continuously growing. So, what is the rationale for a new one?

The inflactionary academic publication world makes the task of acquiring a coherent state-of-the-art representation of the field an almost impossible task. This is extremely counterproductive when trying to incrementally build a real science. The staircase toward a rigorous, widely accepted, testable, theory of mind is obscure, arduous, tiresome and sometimes exasperating. This mostly happens because there are thousands of pretend-to steps and the real ones are scattered through so many places.

We feel there is a strong need for simplification and focusing of mind-theoretical works. We believe that the pursuit of the ultimate understanding of mind shall be easier if we are able to get rid of the enjoyable but otherwise decorative literature that is used to describe most of the theories. While this kind of text usually embellishes the many insights on the nature of mind and somehow helps grasping their theoretical underpinnings, a narrower focus on

the very core issues is absolutely necessary. Succinctness becomes a major target in this quest for a theory of mind.

Hence, in the old way of the hard sciences, we strive for terse formalizations that will minimize the need for ink and paper and will hopefully convey precise, non-interpretable expressions of theories or hypotheses on mind nature. With this goal in mind we are launching this yet-another-journal, hence contributing to the growing plethora of periodic publications but with the sole and noble aim of capturing, in a single place, a more *rigorous science of mind*.

It is clear that formality and abstraction have been attempted in the past in the study of the mind; but instead of focusing on a concrete formalism and/or a concrete limited target for formalization, we aim to open the domain to the mind at large without committing to one particular language. The commitment is only with the objective: an *unified formal theory of mind*.

If we are successful in this attempt, we hope to see a single journal in the reading pile.

2 Journal focus

The *Journal of Mind Theory* aims to stress a rigorous and even formalist approach to the investigation and theorization about the mind. It is driven by the developing scientific view that all mental issues –intentions, thoughts, feelings– are just natural phenomena and therefore can and must be explored within a strict scientific framework encompassing both theoretical and empirical concerns. This emerging view is coming from the consilience of multiple strands of analysis that are breaking the disciplinary boundaries.

Under this programme the Journal of Mind Theory:

- Seeks theoretical rigor in theories of mind;
- Seeks contributions that transcend the traditional disciplinary boundaries in cognitive science, encouraging articles from researchers interested in a formal approach to the analysis of cognition;
- Emphasizes the synthesis of ideas, constructs, theories, and techniques in the analysis of biological cognition and in the design of cognitive autonomous systems, offering a platform for addressing the problem of formalization of cognition from a systemic and naturalized perspective;
- Addresses the classic topics of theory of mind but with a formal tint: perception and phenomenology, theory of knowledge, reasoning and causation, the role of mathematics and logic in cognitive systems and philosophical foundations of cognition;
- Accepts experimental work insofar it addresses specific theories.

JMT looks for fresh thinking, vigorous debate, and careful analysis!

3 Content of the Journal

JMT is a conventional scientific journal, and hence its main content is a set of research articles. In each number there will be a special "feature" article addressing in detail a concrete, complete theoretical approach.

There will be other several smaller articles on specific topics and, finally, there will be special sections of related content (reviews, interviews, position papers, cultural notes, etc).

4 The question of "formality"

There may be some concerns concerning the meaning of the word "formal" in the context of JMT, but this is a journal for simple people:

- Scientists aiming for a scientific theory of mind, and
- Engineers who are trying to understand enough about minds in order to be able to replicate some of its capabilities- with economically required engineering certainty.

In this sense, we do not constrain the meaning of "formal" in JMT to logics, quantum mechanics or post canonical systems (or whatever formal framework any reader may think about) but to the class of languages used to describe systems that minimize the possibilities of hermeneutical differences (i.e. to be able to write descriptions that do not suffer the vagaries of interpretations).

The point to be retained is that the formalizations are methodological tools and not just ontological simplifications. We want JMT to be a channel of precise mind-theoretical communication and not a demonstration of the powers of specific formalisms.

In this search for a precise theorization about mind, we would say that in JMT there are two intertwined threads:

- What is the mind? (described in a "formal" language)
- What is the language? (suitable for describing "mind")

This last may be FOL, PCS, Java, Dynamical Systems Theory or whatever is suitable for capturing the theory and is more precise than old, good, plain English, German or Latin.

The hope and the core rationale behind JMT is that both threads –the theory and the language for expressing it– will eventually converge into a single "formal language" or "mind theory" conundrum.

Extrapolating beyond what may be reasonable, the language of convergence may indeed be the ultimate LoT; transcending the original idea of LoT that is linguistically biased, obviating other languages of more mathematical nature;

an extremely efficient source of new concepts and tools to understand reality (mental processes included).

5 The question of "reductionism"

It may seem that the endeavor that sublimes JMT is a total reduction of mind to mathematical physics. For some of us it may be the case, but for others it may be not; in any case, it is necessary to be precise in the expression of the way of the reduction or the way of non-reduction, e.g. by emergence. If we are expecting to resolve the issue, both theoretical models shall be commensurate.

Reductionism is a term with considerable bad press within certain cultural milieu that considers the reductionism as the credo (just another -ism) carried out by the reductionists, who are those that approach the understanding of complex phenomena by over simplifying them.

Admittedly, reductionist statements ornamented with some obscure technical terminology made by a few, has served to brutalize social reality and minimize environmental influences for the most self-serving reasons.

However, to tell the whole truth, reductionism and mathematization are dangers only when used to serve private interests and limited knowledge of the mathematical structures introduced in the explanations. In JMT we aim to transcend the pathological fear of reductionism and mathematization within the cognitive sciences, from academics in the humanities, neurosciences and postmodern robotics.

6 About JMT Volume 0

Volume 0 is the first volume of JMT and its sole objective is to start the Endeavour setting a basis for further development and focusing in the long-term objectives of the Journal of Mind Theory. JMT Volume 0 has been edited in two numbers of roughly similar size and variety of content:

- **JMT Volume 0 Number 1**
    - *Feature: Toward a Computational Theory of Mind*
    - *The Mind as an Evolving Anticipative Capability*
    - *The Challenges for Implementable Theories of Mind*
    - *Special section: Questions for a Journal of Mind Theory*
- **JMT Volume 0 Number 2**
    - *Feature: MENS, a mathematical model for cognitive systems*
    - *The Unbearable Heaviness of Being in Phenomenologist AI*
    - *Pragmatics and Its Implications for Multiagent Systems*
    - *Mimetic Minds as Semiotic Minds How Hybrid Humans Make Up Distributed Cognitive Systems*
    - *Special Section: Neuroeconomics and Neuromarketing; Practical Applications and Ethical Concerns*

Our very first article, *Toward a Computational Theory of Mind* by *Albus*, is a tour-de-force, in which, James Albus summarizes his life-long research work dedicated to the analysis and synthesis of mind using an architectural approach. The resulting system, RCS, is an architectural reference model able to both serve as explanatory framework for natural cognition and as blueprint for artificial mind construction.

In *The Mind as an Evolving Anticipative Capability*, Cottam, Ranson and Vounckx make a concrete proposal on the nature of mind and give a rationale for it: Mind is just an *evolving anticipative capability*. This theoretical model is set in a landscape of ecological multiscalar evolution leading to an architecture of mind that exploits internal multiresolutional model structures that serve to guide the behavior of the evolving agent population in multiscalar environments. The article analyzes the implications of their theoretical model for the transposition of genotypic to phenotypic aspects that drive agent operation.

Haikonen contributes *The Challenges for Implementable Theories of Mind*, where he departs from the excessively metaphorical nature of many of the theories of mind that are too loose to serve as blueprints for mind engineering. He clarifies the necessary profile of an implementable theory of mind, identifying some of the core issues that shall be addressed by such a theory: mind-body relation, meaning and understanding, emotion, qualia, etc.

Questions for a Journal of Mind Theory is a special section of JMT: *Interview*. In this case this is a questionaire proposed by one of the editors of JMT (Gómez) and answered by a philosopher (Talmont-Kaminski) and an engineer (Sanz, the other JMT editor). In this questionaire some of the basic questions traditionally addressed by the philosophy of mind are re-considered under the panorama for rigor proposed by JMT.

MENS, a Mathematical Model for Cognitive Systems proposes a mathematical theory to answer the fundamental question of how higher mental processes arise from the functioning of the brain? Ehresmann and Vanbremeersch have spent 20 years working on an entirely new model for studying living organisms. MENS provides a formal unified model for the investigation of the mind, translating ideas of neuroscientists into a mathematical language based on Category Theory.

The Unbearable Heaviness of Being in Phenomenologist AI points out the misuse of Heidegger's philosophical insights within the discipline of artificial intelligence (AI) and robotics. Jaime Gómez and Ricardo Sanz, as engineers, make a passionate and sensible incursion within the philosophical discourse. The article argues that Husserl's phenomenology ("putting the world between brackets") and other post-phenomenologist doctrines from Heidegger to Merleau-Ponty, has led to a positioning in embodied AI that deeply neglects fundamental representational aspects that are necessary for building an unified theory of cognition.

Samad, in *Pragmatics and Its Implications for Multiagent Systems*, illustrates how incorporating pragmatics can play an important part in multiagent system

performance. The author puts the linguistic discipline of pragmatics in a purely engineering context. As a consequence of this, multiagent communication improves key features like security, robustness or efficiency. Additionally, he offers some examples and preliminary remarks towards formalizing this.

Mimetic Minds as Semiotic Minds How Hybrid Humans Make Up Distributed Cognitive Systems by Magnani, claims that the externalization/disembodiment of mind is a significant cognitive perspective able to unveil some basic features of abduction and creative/hypothetical thinking. Magnani coins the term semiotic brains which are able to make up a series of signs and that are engaged in making, manifesting or reacting to a series of signs. Through this semiotic activity the semiotic brains are at the same time engaged in "being minds" and thus in thinking intelligently.

Neuroeconomics and Neuromarketing; Practical Applications and Ethical Concerns by Belden, inaugurates the JMT special section *Science is Culture*. This section is dedicated to giving a voice to those from other disciplines regarding pertinent or controversial scientific and technical issues covered in the journal. Sarah Belden, a Berlin based curator, explores the ethical issues posed by new technologies within the realm of Neuroeconomics and Neuromarketing. This article is an invitation for critical thinking about the goals of science and its financial support, and our increasing power to see and change the basic structure of human consciousness, thinking and identity, which raises a number of important social, political, cultural and ethical issues.

7 Acknowledgements

The edition of this two number volume has been possible thanks of the enthusiastic commitment of our first authors and the economic and infrastructural support of some organizations in our country –Comunidad de Madrid, Ministerio de Ciencia e Innovación y Universidad Politécnica de Madrid.

We sincerely thank all them for this effort,

The editors

MENS, a mathematical model for cognitive systems

Andrée C. Ehresmann & Jean-Paul Vanbremeersch*

**Université de Picardie Jules Verne*

Abstract

How do higher mental processes, learning, intentions, thoughts, feelings, arise from the functioning of the brain? That is the question we attempt to approach in the Memory Evolutive Neural Systems (or MENS). This theory proposes a formal unified model for the investigation of the mind, translating ideas of neuroscientists such as Changeux and Edelman in a mathematical language based on Eilenberg and Mac Lane's theory of categories (which unifies the main mathematical operations). MENS is an application for cognitive systems of our general model MES for autonomous complex hierarchical systems, such as biological or social systems. The 'complexification process', introduced in MES to model the formation of increasingly complex objects, is related to the "binding problem" of neuroscience and it characterizes how higher cognitive processes, the development of a semantic memory, and consciousness, may emerge from physical states of the brain, thus supporting an emergentist monism. In particular, the existence of consciousness is related to the development of a global invariant, the archetypal core that integrates and merges the lasting corporal and mental experiences, giving a basis at the notion of self.

Keywords

Neuron, mental object, memory, cognition, consciousness, emergence, category, colimit, complexification.

1 Introduction

The Memory Evolutive Neural Systems (or MENS) studied in this paper are a model for cognitive systems of animals, up to a theory of mind for man, which incorporates a basic level **Neur** formed by the neural system, and higher levels, deduced from it, representing an 'algebra of mental objects' (in the terms of Changeux, 1983). The main idea is that these higher levels emerge from the basis through iterative binding processes, so that a mental object appears as a family of synchronous assemblies of neurons, then of assemblies of assemblies of neurons, and so on. They develop over time through successive 'complexification processes', up to the formation of higher cognitive processes and consciousness. Their evolution is internally self-regulated and relies on the formation of a memory in which the different data, experiences, procedures can be stored in a flexible manner, to be later recalled or actualized for a better adaptation. The model takes account of the exchanges with the physical environ-

ment, through receptors and effectors which confer to it a kind of "embodiement" (Varela, 1989), and, for higher animals, through education and cultural activities, stressing the role of the society in the development of higher processes. The notion of self relies on the development of a permanent global invariant, the archetypal core, which integrates the main corporal, perceptual, behavioral, procedural and semantic experiences, with their emotional overtones; its self-maintained activation is at the root of consciousness, characterized in particular by temporal extension processes.

MENS is a particular case of the Memory Evolutive Systems which we have developed in a series of papers during the last 25 years (cf. our book, Ehresmann and Vanbremeersch, 2007). Initially, it has been influenced by works on several domains (Bunge, 1979; Laborit, 1983; Merleau-Ponty, 1945; Minsky, 1986; Morin, 1977; Piaget, 1940), and more particularly on neuroscience (Changeux, 1983; Crick, 1994; Edelman, 1989). It could also be applied to artificial systems such as robots equipped with means to sense their environment, but here we focus on the case of higher animals, up to man.

Though the main ideas can be explained in ordinary language (as we try to do as much as possible in this article, referring to the Appendix for rigorous definitions), MENS is a mathematical model based on the theory of categories. This theory is a relational domain of mathematics, introduced in the forties by Eilenberg & Mac Lane (1945) to unify some problems in algebra and topology, and which accounts for the various operations of the "working mathematician" (Mac Lane, 1971). In MENS, it gives tools for modeling the main human capacities: formation, comparison and analysis of the relations between interacting objects, synthesis of complex objects binding more elementary objects (colimit operation), formation of a hierarchy of increasingly complex objects (complexification process) and their later recognition, classification of objects into invariance classes (projective limit operation), allowing for the development of a semantic. In particular, we model a mental object by what we call a category-neuron (abbreviated in cat-neuron), iteratively constructed as the binding of synchronous assemblies of (cat-)neurons. We show that the "degeneracy of the neuronal coding" emphasized by Edelman (1989) implies that a cat-neuron has several such "physical" realizations; it follows that the links between cat-neurons are not only simple links binding clusters of links between their components of the lower level, but also complex links which emerge by composition of simple links binding non-adjacent clusters. The complex links reflect global properties of the lower level which are not observable locally at this lower level. It is the precise mechanism at the root of the emergence of mental objects and processes of increasing complexity.

MENS brings up philosophical problems related to emergence *vs*. reductionism, mind-brain correlation, self and consciousness.

2 Neurons, mental objects, category-neurons

MENS is a model for the cognitive system of an animal. It intends to describe the development of mental objects and cognitive processes of increasing complexity based on the functioning of his neural system. First we recall some

physiological data on this system and its functioning, and we model it by the evolutive system of neurons **Neur**, which is at the basis of MENS.

2.1 The neural system

The neural system consists of neurons and synapses between them, its dynamics results from the propagation of an action potential from a neuron to other neurons through synapses. It slowly evolves during the life of the animal. There are several types of neurons. For some (e.g., intermediate neurons), their activity is entirely dependent from their connections with other neurons, for others, it is modulated by external or internal events. The receptor neurons, in the various perceptual areas are in contact with the environment of the system and are triggered by changes in this environment; this allows the animal to recognize (innately or after learning) some external features (appropriate foods, predators,...) and develop adaptive responses to them. These responses are realized by effector neurons, in the motor areas, which act on the environment through their action on muscles.

The state N(t) at an instant t of a neuron N is determined by its activity around t, which is a function of its instantaneous rate of firing and of its threshold (related to the difference of potential between inside and outside the cell necessary for starting an action potential). We say that an item (external object or neuron) *activates* N at t if it causes an increase in the activity of N at this date; and we think of the resulting activation as a kind of information transmitted by the item to N.

The state at t of a synapse s from N (the presynaptic neuron) to N' (the post-synaptic neuron) is determined by its *strength*, that is related to the capacity of s to transmit an action potential from N to N' around t. There are excitatory synapses and inhibitory ones. The strength of an excitatory synapse is inversely proportional to the number of spikes of N necessary to start a firing of N', supposing that N' does not receive inputs from other neurons. The synapse s has also a *propagation delay* which measures the delay between the firing of N and its possible transmission to N'. Since each impulse has a specific duration and there is a temporal summation of the impulses, the propagation delay is inversely proportional to the strength of s. An inhibitory synapse decreases the activity of the post-synaptic neuron; its strength is negative, with its absolute value inversely proportional to the number of spikes of N necessary to inhibit the activity of N'. While the activity of a neuron N varies quickly, the strength and propagation delay of a synapse vary much more slowly.

There is a spatial summation of impulses: if N receives simultaneously inputs from several neurons N_i through synapses s_i from N_i to N, the activity of N is an upper-bounded function of the sum of the activities of the N_i and of N, weighted by the strength of the synapses s_i.

2.2 **The evolutive system of neurons** Neur

Given two neurons N and N', there can be several "parallel" synapses from N to N'; they can also be linked by a sequence of synapses, say s_1 from N to N_1, then s_2 from N_1 to N_2, and so on up to N'. Such a sequence is called a *synaptic path* (or, more briefly, a *link*) from N to N'; its strength is a function of the product of the strengths of its components, and its propagation delay is the sum of their propagation delays. In particular a synaptic path (s_1, s_2) in which one synapse is inhibitory and the other excitatory has a negative strength, while its strength is positive if both are of the same kind. Synaptic paths are composed by concatenation (meaning one succeeding to the other). A synaptic path from N to N' is *activated* at t if N activates N' at t along it.

We model the configuration of the neural system at t by a (multi)graph: its vertices model the states N(t) of the neurons N existing at t, an arrow from N(t) to N(t'), also called *link* from N to N' at t, models the state at t of a synaptic path s from N to N'; it is determined by the strength $w_s(t)$ and the propagation delay $d_s(t)$ of s at t, which are real numbers. Equipped with the composition defined by concatenation, this graph becomes the *category* Neur_t *of neurons* at t (for the definition of a category, cf. the Appendix).

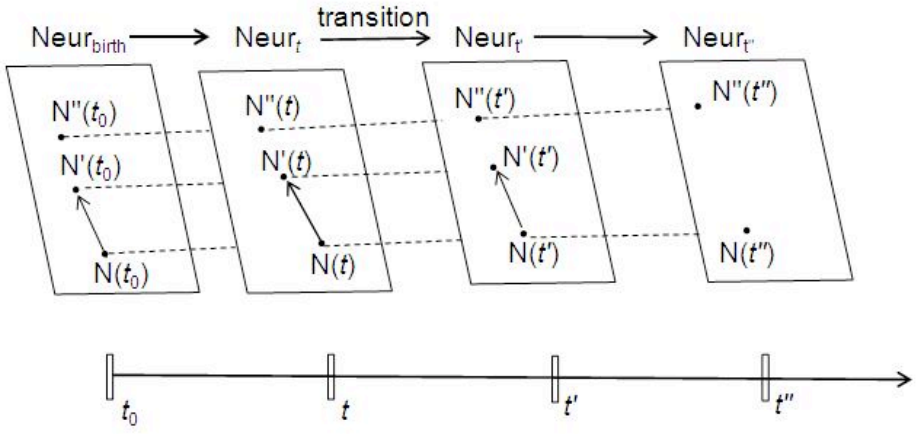

FIGURE 1. For each instant t of the life of the animal, the category of neurons Neur_t consists of the states N(t) *of* the neurons existing at t, and of the synaptic paths between them at this time. The transition from t to t' correlates N(t) and N(t') if N still exists at t'. The categories Neur_t and the transitions between them form the evolutive system of neurons, **Neur**.
A neuron N appears as a component of **Neur**, consisting of its successive states. Here the neuron N' does no more exist at t''.
To an interval (t, t'') we associate the category formed by the neurons and the synaptic paths between them which exist during this period.

Neurons and synapses have a long life. Over time, say from t to a later time t', some neurons are lost, while a few new neurons can grow, and the number of synapses can vary. This change of configuration is modeled by a partial map from Neur_t to $\text{Neur}_{t'}$, called the *transition* from t to t', sending the state of a neuron N at t to its state at t' if N still exists, and similarly for a link. This transition defines a functor from a sub-category of Neur_t to $\text{Neur}_{t'}$. The *Evolutive system of neurons* (Ehresmann and Vanbremeersch, 1987), denoted **Neur**, is

formed by the categories $Neur_t$ and the transitions between them, during the life of the system (cf. Figure 1). Its components model the neurons (via their successive states); they are still called neurons (or, later, cat-neurons of level 0). The operations are not instantaneous but require some period of time; thus what is particularly interesting is the category of neurons and their links existing during such a period, and we generally operate in this category.

2.3 Mental images

Neur models the physical structure of the brain of the animal and its elementary neural dynamics. How can it generate MENS, a model of his cognitive system accounting for the mental operations he can perform, and their evolution over time? The neurons will figure among the components of MENS, but MENS has also other more conceptual objects, which we call *category-neurons* (abbreviated in *cat-neuron*), and which model mental objects (in the terms of Changeux, 1983) of various kinds associated to features of the environment, sensory and motor inputs, internal states, motor skills and various procedures, sensations and emotions, particular events, and so on. A cat-neuron can be thought of as a 'higher order' neuron, The problem is to describe exactly what is a cat-neuron, and how the evolutive system MENS is generated by its sub-system **Neur**, in particular how living and learning leads to the emergence of a hierarchy of cat-neurons modeling more and more complex mental objects and processes.

First we consider a particular kind of mental object, namely a mental image. A mental image corresponds to a long term memory of an item perceived by the sensory organs, say an object in the environment, through which the item can later be recognized or recalled by the animal. It will be represented in MENS by a cat-neuron which gives a record of the item, while keeping some plasticity.

For a simple object, say a small segment of a specific orientation, there is a neuron (the "simple cells" discovered by Hubel and Wiesel, 1962) in a visual area whose firing is specifically triggered by the sight of the object. Such a dedicated neuron may also exist for some complex objects, if they are often met by the animal and/or particularly important for him; for instance an angle triggers the firing of a "complex cell", and there are "place cells" in the hippocampus which have a direct firing with location specific areas (O'Keefe and Dostrovski, 1971). However there is no "grand-mother neuron" (Barlow, 1972).

Brain imagery has shown that more complex items are recognized through the coordinated activation of a whole pattern of neurons more or less distributed in the brain and interconnected by distinguished links (which can be synapses or synaptic paths). This pattern corresponds to an internal memory of the item (Stryker, 1989); its characteristic is that it can act as a *synchronous assembly of neurons* (Hebb, 1949). Here 'synchronous' means that all the neurons of the pattern are activated during the same cycle of the natural oscillation of the neural activity of the brain area to which they belong (e.g. 40Hz in the hippocampus, Fisahn *et al,* 1998); anyway the synchronization lasts only a short time (cf. Miltner *et al.*, 1999; Rodriguez *et al.*, 1999; Usher and Donnelly, 1998).

To act synchronously, the distinguished links of the pattern must have short propagation delays, and therefore great strengths, since delays and strengths are inversely proportional. The formation of such a pattern relies on the following rule, proposed by Hebb in 1949, and experimentally confirmed for synapses in many areas of the brain (*e.g.*, Engert and Bonhoeffer, 1997; Frey and Morris, 1997; Zhang *et al.*, 1998):

> *Hebb rule*: If s is a synapse from N to N' and if the activities of N and N' are simultaneously increasing, the strength w_s of s increases at the same rate. Conversely if the activities of N and N' vary in opposite ways, w_s decreases.

The mental image of an unknown item O will be generated as follows: the perception of O at a given time t activates the neurons of a specific pattern P, thus forming a short-term memory of O. By Hebb rule, their coordinated activation at t increases the strength of the distinguished links between them; and the same repeats at each successive perception of O. Thus, there is a progressive decrease of the propagation delays, which facilitates a coordinated firing of the whole pattern; and over time, the pattern P will take its own identity, being able to act as a synchronous assembly of neurons. In this way the short-term image of O has been consolidated in a long-term memory.

However we cannot identify the pattern as such with the mental image of O. Indeed, depending on the context, the same object O can activate more or less different patterns acting as synchronous assemblies of neurons, and these patterns are not necessarily interconnected (Edelman, 1989, p. 50). The importance of this property has been emphasized by Edelman who speaks of the "*degeneracy* of the neuronal encoding"; we will see later how it is at the root of the emergence of higher mental processes. These other patterns also participate in the mental image of O. This image must be thought of as the invariant that all these patterns P have in common, namely they all have the same functional role, meaning that they can activate the same neurons, and with the same strength; we say that they are *homologous*. This invariant will be modeled by a category-neuron.

2.4 Category-neurons

If O is a simple object or a complex object of importance for the well-being of the animal), the invariant corresponds to a particular neuron. Indeed, in this case there is a neuron which activates the same neurons, and with the same strength, as anyone of the patterns P which are activated by the perception of O, hence which participate in its mental image. This neuron, called the *binding* of P, becomes the mental image (or *record*) ImO of O, and O is later recognized or recalled through the firing of ImO.

For more complex objects, there is no such neuron, and the mental image will be modeled by a cat-neuron, component of MENS. How does one explicitly define a cat-neuron?

In **Neur**, a *pattern* P of neurons is a family of neurons P_i interconnected by some distinguished links (*i.e.*, synaptic paths) through which they may transmit their activation to each other. A collective interaction of P is modeled by a *collective link* from P to a (cat-)neuron N'; it is a family of links s_i from P_i to N', correlated by the distinguished links of P, so that they may collectively activate N'. We model the fact that two patterns are homologous (i.e., have the same functional role) by the fact that there is a 1-1 correspondence between their collective links to any (cat-)neuron N'. If P has no binding *neuron*, it may have a *binding cat-neuron* N in the following sense: the collective links (s_i) from P to any cat-neuron N' are in 1-1 correspondence with the links s from N to N' (in categorical terms, N is a *colimit* of the pattern in MENS; cf. Appendix); in other terms, P as a whole and its binding N have the same functional role (cf. Figure 2). In this case, N is also the binding of any pattern Q homologous to P. The pattern P, as well as each other pattern Q that N binds is called a *decomposition* of N, and the passage from P to Q is called a *complex switch.* (cf. Figure 3).

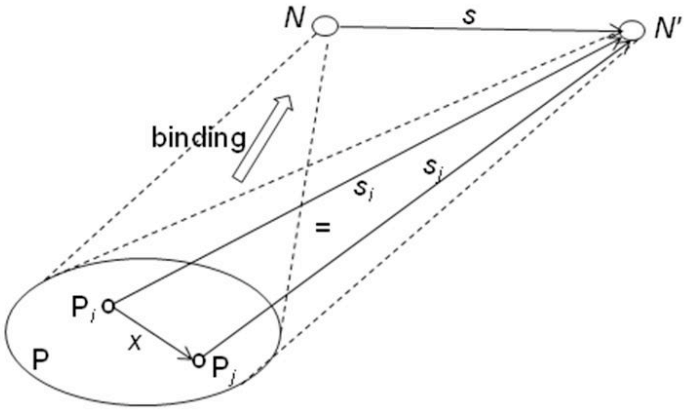

FIGURE 2. P is a pattern of neurons P_i with some distinguished links (synaptic paths) x between them. A collective link from P to a cat-neuron N' is a family of links s_i from each P_i to N' correlated by the distinguished links x. The pattern P admits a cat-neuron N as its binding if each collective link (s_i) from P to an N' binds into a link s from N to N', so that the whole pattern P and the cat-neuron N activate the same cat-neurons, with the same strength.

Let us come back to the item O, and let P be one of the patterns participating in its mental image. This image will be modeled by a new component ImO of MENS (hence a cat-neuron), called the *record* of O. This cat-neuron emerges as the binding of P; by definition of the binding, it will also bind the other homologous patterns which participate in the image. Since two decompositions P and Q of ImO are not necessarily interconnected (as said above), we say that ImO is a *multifold* component of MENS. Its activation at a time t (allowing to recognize or recall O) will consist in the synchronous activation of one of its decompositions, say P, and its activity is then a function of the global activity of the neurons of P.

Over time, ImO takes its own identity, independent of a particular decomposition. It may 'lose' one of its decompositions, for instance if lesions in the brain destroy a number of neurons of P, what remains of P may become too small

for keeping the same functional role, and ImO will no longer be its binding. Conversely ImO may acquire a new decomposition Q; for instance if O progressively changes (e.g., a person who ages), the assembly of neurons synchronously activated by O changes slowly, though remaining a decomposition of ImO. However if the change in O becomes too large or sudden, ImO will not remain the image of (what has become) O.

To sum up, ImO is initially constructed to bind a particular pattern P of neurons and thus become the image of O. Later it takes its identity and can even disassociate from P. Thus it is not a rigid record (as in a computer), but offers a flexible memory which adapts to changing situations. The multiplicity of its decompositions ensures that the animal is able to recognize or recall the object under different forms, even new forms he has not yet met, as long as the change is progressive enough.

ImO is a cat-neuron 'of level 1'. More generally a *cat-neuron N of level* 1 will bind a class of homologous patterns of neurons. It is constructed at a time *t*, through a complexification process (cf. section 4) to bind a given pattern of neurons, so that it can act as a synchronous assembly of neurons. Later on, *N* takes its own identity, possibly acquiring other decompositions which are not necessarily interconnected with P.

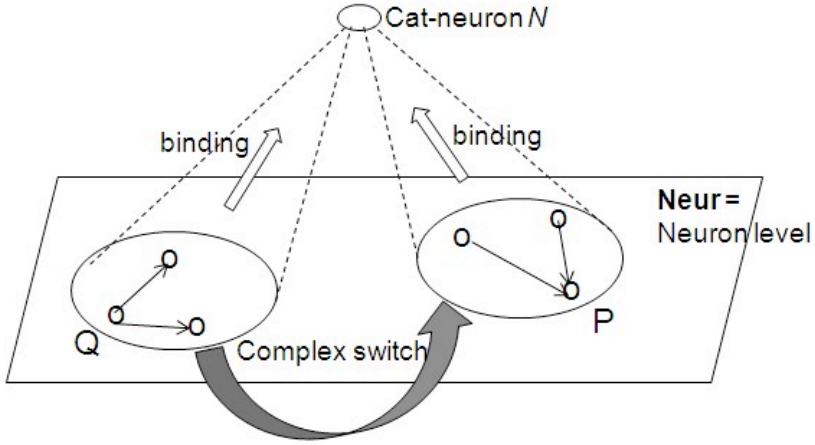

FIGURE 3. The cat-neuron *N* is the binding (or colimit) of the two non-interconnected patterns P and Q. However the fact that P and Q are homologous cannot be observed 'locally' at the level of the neurons in P and Q. The complex switch from P to Q emerges at the higher level of cat-neurons (as indicated by the twisted "Môbius band' style of arrow that crosses the Neur level); however it expresses a global property of Neur, namely that, for any cat-neuron *N'*, there is a 1-1 correspondence between the collective links from P to *N'* and the collective links from Q to *N'*.

It is important to realize that, as a multifold component of MENS, a cat-neuron *N* has 'emergent' properties, that is, properties not observable locally from inside the neuron level **Neur**, though a consequence of its global structure. Indeed, if P is a decomposition of *N*, to recognize that another pattern Q is also a decomposition of *N*, we must verify that P and Q are homologous, meaning that P and Q activate the same neurons and with the same strength. If P and Q are non-interconnected (there is no cluster of links between them),

this verification has to take account of the whole structure of **Neur**, and not only of the links between neurons of P and Q. Thus the existence of a complex switch between P and Q expresses a 'global' property of **Neur**, which emerges as something new at the cat-neuron level 1; in the figures a complex switch will be represented by a twisted ('Möbius band' type) arrow between P and Q, crossing over the neuron level. In spite of its emergent properties, the cat-neuron relies on the physical basis of its different decompositions, and may produce physical effects through the synchronous activation of anyone of its decompositions.

3 The category-neurons and their links

The neurons will be identified to cat-neurons of level 0. The cat-neurons of level 1 that we have just defined bind patterns of neurons. To model more complex mental objects, we have to iterate the construction and form cat-neurons of increasing complexity levels, binding patterns of cat-neurons of lower complexity. This confronts us with two problems:

1. The "binding problem" which in MENS translates into: how do cat-neurons interact?
2. How do the "algebra of mental objects" emerge from the neural system?

3.1 Clusters and Simple links

If we think of a cat-neuron as a kind of 'virtual' higher order neuron modeling a mental object, what will correspond to 'virtual' synapses or synaptic paths between cat-neurons? A cat-neuron N emerges from the neuron level to bind a pattern P of neurons; the links between cat-neurons should also emerge from this level. The first idea is that a link from N to another N' will bind (in some sense) a *cluster* of links between the neurons of P and those of a decomposition P' of N'. So the first step is to find the properties that such a cluster should verify to be able to collectively activate P'. In particular, these clusters model the 'good' interactions between synchronous assemblies of neurons, thus giving a solution to the binding problem as it has been stressed by several neuroscientists, in particular von Malsburg (1995; von Malsburg and Bienenstock; 1986).

If P' is reduced to one neuron N', the pattern P activates N' if all its neurons simultaneous activate N in a coherent way; thus a cluster from P *to* N' is just a collective link from P to N' (as defined above), that is a family of links from the neurons P_i of P to N' correlated by the distinguished links of P which may operate synchronously to activate N'. If P has a binding cat-neuron N, this collective link binds into a link from N to N'.

If P is reduced to a neuron N, a cluster from N *to* P', also called a *perspective* of N for P' (this terminology will be explained later) is a maximal set of links from N to some neurons of P' which are correlated by a zig-zag of distinguished links of P', so that N can synchronously activate a well connected subpattern of P' (but not necessarily the whole of P').

In the general case, we define a *cluster* from P to P' as a collective link of perspectives from the different neurons of P to P' (cf. Appendix).

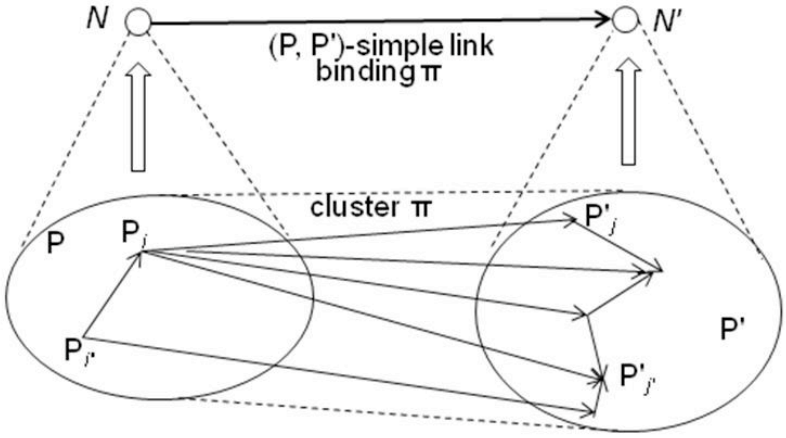

FIGURE 4. P and P' are two patterns; a cluster π from P to P' is a maximal set of links such that: (i) the links in the cluster from each P_j of P are correlated by a zig-zag of distinguished links in P' (they form a perspective of P_j for P'); and (ii) it is closed by composition with distinguished links of P.
If *N* and *N'* are cat-neurons binding P and P', the cluster binds into a (P, P')-simple link from *N* to *N'*; this link activates *N'* with the same strength as P acting collectively through the cluster.

If P and P' are respectively decompositions of the cat-neurons *N* and *N'*, a cluster from P to P' binds into a link from *N* to *N'* in MENS; we call this link a (P, P')-*simple link* (cf. Figure 4). Such a link just sums up the activation that the links in the cluster individually transmit from neurons of P to neurons of P', thus it is entirely reducible to the neuron level. It is completely dependent on the chosen decompositions P and P' and takes no account of the possible other decompositions of the cat-neurons: a (P, P')-simple link may not be (Q, Q')-simple if Q and Q' are other decompositions of *N* and *N'*. In particular the identity of *N* is (P, P)-simple for each decomposition P of *N*, but it is (P, Q)-simple only if P and Q are interconnected (meaning more precisely that there is a cluster between P and Q binding into the identity of *N*).

3.2 Emergence of complex links

The simple links have nothing to do with the emergent properties of *N* and *N'* due to their multiple decompositions. However, the existence of multiple decompositions accounts for the emergence of other links from *N* to *N'*, called *complex* links, which are not simple for any decomposition of *N* and *N'*. Their existence comes from the fact that a multifold cat-neuron *M* may have two non-interconnected decompositions Q and Q'. If we have a (P, Q)-simple link S from *N* to *M* and a (Q', P')-simple link S' from *M* to *N'*, they compose into a complex link SS' from *N* to *N'*; this link transmits the same activation to *N'* as that transmitted by S' when M is activated by S (cf. Figure 5).

More generally, a composite of simple links binding non-adjacent clusters, connected by complex switches is a complex link. A composite of complex links is generally a complex link (though it can sometimes be simple for particular decompositions of the extreme cat-neurons it connects).

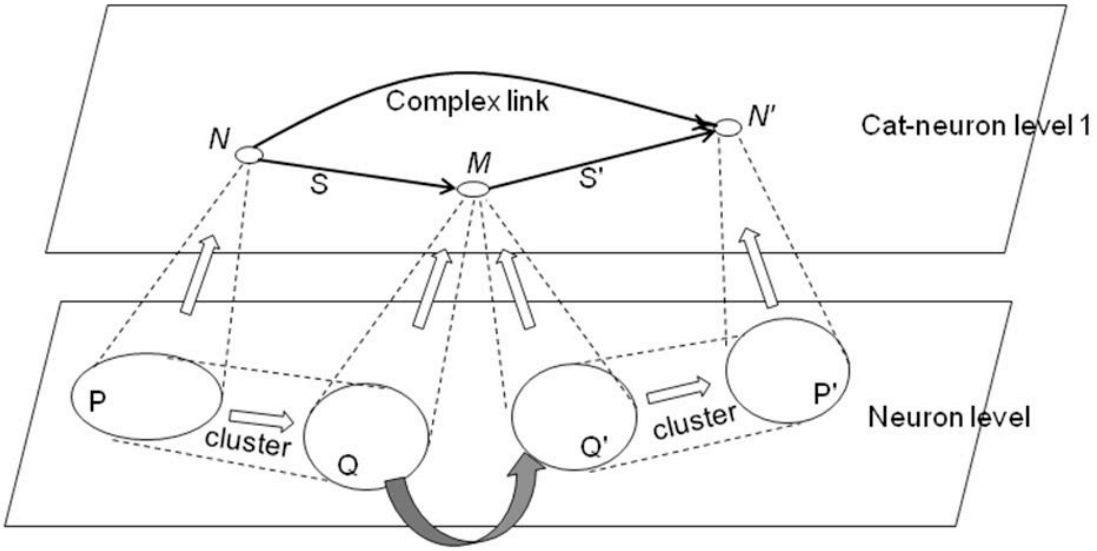

FIGURE 5. Q and Q' are two non-interconnected decompositions of the cat-neuron *M*, the complex link is the composite SS' of the (P, Q)-simple link S and the (Q', P')-simple link S'. It emerges at the cat-neuron level 1 and has properties which are not deducible from local properties of links between P and P', though coming from global properties of the neuron level.

We define the *propagation delay* of a (P, P')-simple link as the maximum of the propagation delays of the links in the cluster, and its strength is an increasing function of the strengths of these links. The propagation delay of a complex link is the sum of the propagation delays of its composing simple links, and its strength is an increasing function of their strengths. We have shown that *Hebb rule* generalizes to cat-neurons connected by such links (cf. Ehresmann and Vanbremeersch, 1999, 2007).

A (P, P')-simple link S is entirely determined by the cluster it binds, hence can be observed directly at the level of the neurons in P and P', and transmits only information already mediated through them. On the other hand, a complex link SS' conveys more 'global' information, since it makes use of a complex switch between the two non-interconnected decompositions Q and Q' of the intermediate cat-neuron *M.* As explained above, the existence of a complex switch is an emergent property of the global structure of **Neur**, even if it can sometimes be experimentally observed (e.g. if *M* models the mental image of an object O, its decompositions correspond to the patterns activated by O). Thus, a complex link from *N* to *N'* relies on properties of the whole level of neurons; it is not a reflection of local properties with respect to particular decompositions of *N* and *N'*. In this sense, it 'emerges' at the level of cat-neurons, but it does not appear 'ex machina', it just actualizes at the higher level a *global* property of the lower level.

The category of neurons is extended in a larger category by adding the cat-neurons of level 1 and the simple and complex links so constructed.

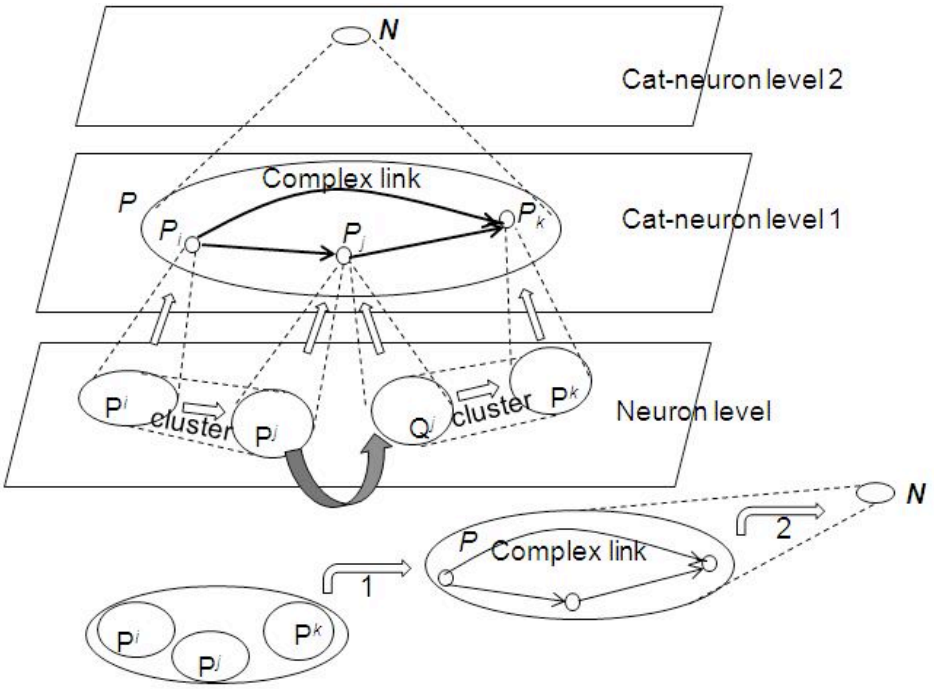

FIGURE 6. ***N*** is a cat-neuron of level 2 binding a pattern *P* of cat-neurons P_j of level 1, one of which P_j has 2 non-interconnected decompositions P^j and Q^j. Thus ***N*** admits 2 ramifications down to the neuron level, one with P^j at the end, the other with Q^j; and it can be activated by anyone of them, for example the first one. This activation necessitates 2 steps: (i) Simultaneous activation of one decomposition P^i of each cat-neuron P_i which forces the activation of this cat-neuron; (ii) The synchronous activation of all the cat-neurons of *P* leads to the activation of *P* (directly at the level 1 because of the complex link), and therefore of ***N***.

3.3 Higher level cat-neurons

The mental image of an item has been constructed as a cat-neuron of level 1, which binds the synchronous assemblies of neurons activated by the item. Higher animals are able to operate on mental objects to form more complex ones. For instance they can form a mental image of a complex object by decomposing it in smaller parts which are recognized, and combining the images of these parts. Or they can learn to perform a new motor skill, by combining more elementary skills already known. In MENS, this corresponds to the construction of a cat-neuron of level 2. Since we have defined what the simple and complex links between cat-neurons of level 1 are, we can speak of patterns of such cat-neurons, of their collective links, and of their binding (literally defined as for neurons), and we easily imitate the construction of cat-neurons of level 1, just replacing the patterns of neurons by patterns of cat-neurons of level < 2. Roughly we can 'compute' with cat-neurons of level 1 as if they were neurons.

Let us develop the construction of a cat-neuron of level 2 modeling a mental image of an object C formed by the juxtaposition of several objects O_i that the animal can already recognize (cf. Figure 6). When the animal perceives it for the first time, C simultaneously activates synchronous assemblies of neurons P^i corresponding to the various O_i. It follows an activation of their records ImO_i. A scan of the object C shows how the O_i are associated in C, and in MENS their inter-relations are modeled by links between the records ImO_i. The records and these links form a pattern *P* of cat-neurons synchronously activated by C. The mental image of C will be modeled by a new cat-neuron ImC added to MENS for binding this pattern (in categorical terms it becomes a colimit of *P*); we say that (*P*, (P^i)) is a *ramification* of ImC. As for cat-neurons of level 1, the record ImC takes its own identity and may acquire various homologous ramifications obtained by replacing each P^i by a homologous pattern of neurons, or *P* by a homologous pattern of cat-neurons. Thus it is not a rigid record, but can adapt to small modifications of C. The later recognition of C consists in the activation of ImC through the unfolding of one of its ramifications, which necessitates 2 steps:

1. first simultaneous activation of the various P^i which leads to the activation of their bindings P_i;
2. then activation of the distinguished links between the P_i to synchronously activate the pattern *P*.

More generally, a cat-neuron ***N*** of level 2 is the binding of a pattern of cat-neurons of level < 2, the distinguished links between them being either simple or complex. It emerges (in a complexification process, cf. Section 4) to bind such a pattern acting as a synchronous assembly of cat-neurons. It also binds all the patterns homologous to *P*; among them some can be non-interconnected. Later it takes its own identity, possibly independent from *P*, and may acquire other decompositions. As above, ***N*** has *ramifications* down to the neuron level; a ramification (*P*, (P^i)) consists in a decomposition *P* of ***N*** in cat-neurons P_i of level < 2, and for each P_i one of its decompositions P^i in neurons. In other terms, ***N*** binds the synchronous assembly of assemblies (or *super-assembly* of neurons) formed by the neurons of the different assemblies P^i with their distinguished links in *P*. Thus ***N***, as a component of MENS, is a conceptual unit, but its later activation (or recall) corresponds to the dynamic unfolding of one of its ramifications which activates the corresponding super-assembly of neurons; the unfolding is done in two steps: simultaneous activation of the various P^i, followed by their synchronization through the links of *P* to activate ***N***. An experimental example of this process has been observed in odor encoding (Wehr and Laurent, 1996). Let us note that at each step we have multiple choices: choice of a decomposition *P* of ***N***, then choice of a decomposition P^i for each cat-neuron P_i of *P*; that gives numerous degrees of freedom to the cat-neuron ***N***, allowing for an adaptation to various contexts.

Simple links between cat-neurons of level ≤ 2 are defined as for cat-neurons of level 1. Since a cat-neuron of level 2 may have non-interconnected decompositions, there will also exist complex links composing simple links binding non-adjacent clusters. The propagation delay and strength of these (simple or complex) links are computed as in the level 1. Thus the construc-

tion can be iterated to construct cat-neurons of level 3, and progressively cat-neurons of increasing levels, modeling more and more complex mental objects.

4 The memory evolutive neural system MENS

Above we have iteratively defined cat-neurons as binding synchronous assemblies of cat-neurons of lower levels. These cat-neurons model more or less complex mental objects, such as mental images of features in the environment, behaviors or internal states. Now we have to explain how the comportment of the animal in his environment, through his successive physical, affective or social experiences, promotes the construction of a particular cat-neuron rather than another, and how it leads to the progressive construction of a hierarchical evolutive system MENS over his life.

4.1 MENS as an evolutive system

In Section 2 we have seen that **Neur** is an evolutive system, its timescale being the life of the animal. It is the same for MENS. At a given time t of his life, the category $MENS_t$ models the (neural and) cognitive system of the animal; its objects are the states at t of the cat-neurons existing at this date, its links are the simple and complex links connecting them around this date. The transition from t to t' keeps trace of the change of state of these cat-neurons and links.

NEUR	MENS
Neurons	Cat-neurons
Synapses	Simple links
Synaptic paths	Complex links

4.2 The complexification process

At birth, the neural system of the animal is practically formed, and he has some innate mental objects associated to some simple actions (breathing, eating, sleeping,...), to specific features of his environment important for his survival (e.g. cues to recognize predators or preys) and to adapted behaviors (sucking, running away from a predator, catching a prey). Thus, the category $MENS_0$ modeling his cognitive system at this date contains, in addition to $Neur_0$, the (states of the) corresponding cat-neurons (probably of level at most 2) and their links. The evolution of the neural and mental system of the animal during his life depends on his successive sensory, proprioceptive, motor, mental, affective and cognitive experiences; they give rise to the formation of new mental objects (e.g., formation of new mental images of objects he perceives) and processes (new motor skills, behaviors), and possibly to modifications or even destruction of others (if they are no more adapted).

In terms of cat-neurons, this evolution consists of realizing some objectives of the following kinds:

1. formation (or preservation, if it already exists) of a cat-neuron binding a given pattern P of cat-neurons; it forces the strengthening of the distinguished links of P, so that P can act as a synchronous assembly of cat-neurons; the formation of a mental image ImO is an example;
2. formation of a new neuron or of new links;
3. elimination of a cat-neuron (e.g., loss of a neuron, destruction or modification of a record if it is no more adapted).

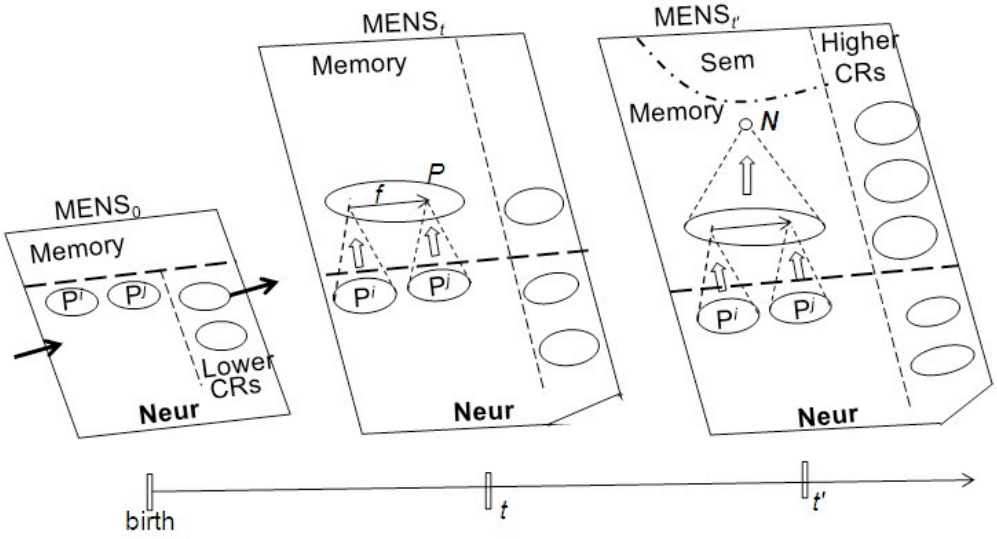

FIGURE 7. Development of MENS during the life. At birth MENS consists essentially of neurons in **Neur** and a few innate cat-neurons in the memory, there are only lower coregulators in Neur. By complexification, MENS extends; at *t*, various patterns P^i have acquired a binding, and simple or complex links such as *f* have emerged between their bindings; a few neurons have been lost. Another complexification process adds a cat-neuron ***N*** of level 2, binding the pattern *P*. More higher coregulators emerge, and the records begin to be classified in the semantic memory Sem.

To model the evolution of MENS, say from a date *t* to a later date *t'*, we suppose that the category $MENS_{t'}$ is constructed as the *complexification* of $MENS_t$ with respect to a procedure having specific objectives of the above kinds (cf. Figure 7); it means that these objectives are realized in the 'best way' in $MENS_{t'}$. We have given an explicit description of this category and of the corresponding transition functor from $MENS_t$ to $MENS_{t'}$ (Ehresmann and Vanbremeersch, 1987, 2007) partially recalled in the appendix. Essentially the complexification fulfills these objectives in the 'most economical' way (in terms of energy and time); the links between the cat-neurons are both the simple links and the complex links as defined in Section 3.

Finally MENS is deduced from **Neur** by successive complexification processes. For higher animals, the procedures may have another kind of objective which will be explained later (Section 6).

4.3 The hierarchy of cat-neurons and their complexity order

The animal has a hierarchy of mental objects, from the mental image of a simple object modeled by a single neuron, to mental objects activated by a synchronous assembly of neurons, up to more complex mental objects combining more elementary ones. Translated in MENS, it means that MENS is a *hierarchical evolutive system* (cf. Appendix). Indeed, by construction:

- the cat-neurons are divided into several levels: the neurons figure at the level 0, the cat-neurons binding an assembly of neurons figure at level 1, the cat-neurons binding a super-assembly of neurons (constructed in 2 steps) at the level 2, and so on;
- a cat-neuron of level n binds at least one pattern of strictly lower levels cat-neurons.

A cat-neuron $\boldsymbol{N}$ of level n emerges in a complexification process to bind a pattern of cat-neurons of strictly lower levels (hence $< n$) which acts as a synchronous assembly, as well as all the homologous patterns. Later it takes its own identity and may acquire new decompositions in patterns of cat-neurons of lower levels. It is multifold, in the sense that two of its decompositions, say P and P', can be non-interconnected (categorically, the identity of $\boldsymbol{N}$ is not a (P, P')-simple link). As explained for cat-neurons of level 2, a cat-neuron of level n admits *ramifications* down to the neuron level. A ramification consists of a decomposition P of $\boldsymbol{N}$, then a decomposition P^i of each cat-neuron P_i of P, then a decomposition of each cat-neuron of the various P^i, and so on down to decompositions in patterns of neurons. The activation of $\boldsymbol{N}$, at a given time, consists in the unfolding of one of these ramifications, through a stepwise process, with multiple choices at each intermediate level, ultimately activating a synchronous assembly of assemblies... of assemblies of neurons, abbreviated in *synchronous hyper-assembly of neurons*. For example, to activate the record (and recognize) an ambiguous image, such as the duck-rabbit, we can use either a ramification that activates the duck record, or one that activates the rabbit record.

Thus, the plasticity of a cat-neuron, or of the mental object which it models, increases with the length of its ramifications, each step adding new degrees of freedom. For a cat-neuron $\boldsymbol{N}$ of level n, this length is generally n, since we have constructed the cat-neurons of level n in n steps: first cat-neurons of level 1 binding patterns of neurons, then cat-neurons of level 2, up to those of level n binding patterns of strictly lower levels. However $\boldsymbol{N}$ may have (or later acquire) ramifications of length strictly less than n. For example a cube C may have initially be decomposed in 6 squares, each modeled by a cat-neuron of level 1, and ImC which binds the pattern they form will be of level 2; but the cube can also be decomposed in its 12 edges, each having for record a unique neuron (a 'simple cell'), and ImC can be obtained in 1 step, as the binding of the pattern of these neurons. Thus, a cat-neuron of level n may sometimes be activated (through some of its ramifications) in less than n steps, and the 'real' complexity of a cat-neuron is related the minimum length of a ramification down to the neurons rather than the level.

To define this complexity, we define the *(complexity) order* of a cat-neuron ***N*** as the smallest *k* such that ***N*** binds (in one step) a pattern of cat-neurons of levels strictly less than *k*. If this order is strictly less than the level, the cat-neuron is *k-reducible*. The above example of a cube shows that its record is 1-reducible, and its order is 1. An example of a cat-neuron of order 2 is given by the record of a Möbius band, obtained by binding a pattern of triangles (this example is taken from Ryan, 2007). The total record of an ambiguous image, such as the duck-rabbit, is of strictly higher order than the records of the duck and of the rabbit taken separately.

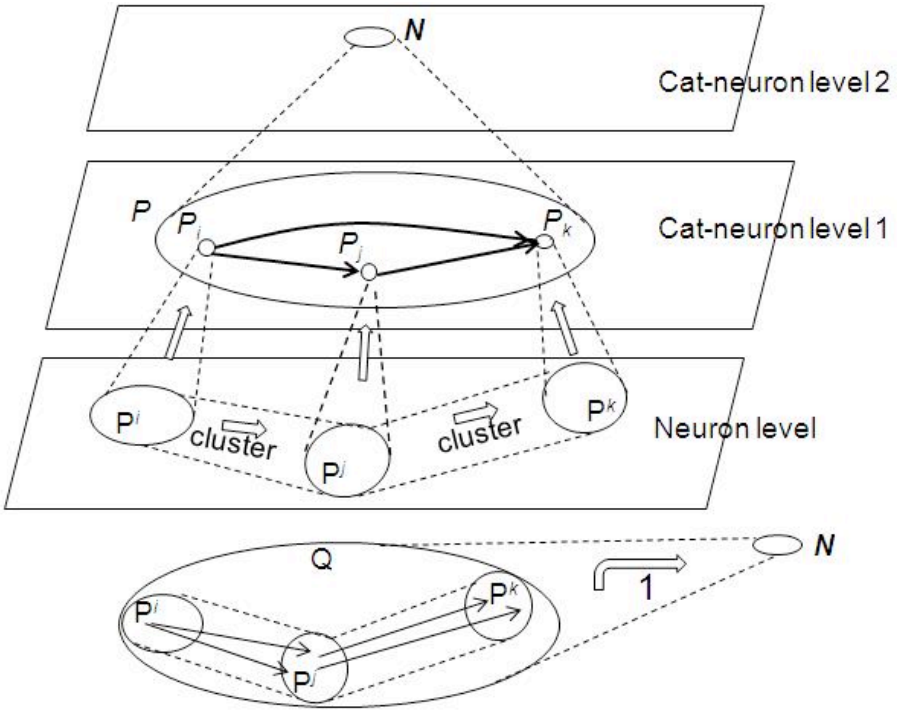

FIGURE 8. ***N*** is a cat-neuron of level 2 binding a pattern *P* of cat-neurons of level 1 in which all the distinguished links are simple. It is also the binding of the pattern Q of neurons containing all the decompositions P^i of the cat-neurons P_i of *P* and the links of all the clusters that the links of *P* bind. Thus ***N*** is of complexity order 1, and it can be activated in one step through the activation of the large pattern Q.

More generally we have proved the following result (Reduction Theorem, Ehresmann and Vanbremeersch, 2007, page 104): If ***N*** is a cat-neuron of level *n* binding a pattern *P* of cat-neurons in which all the distinguished links are simple links, its complexity order is strictly less than *n*; The following figure (Figure 8) illustrates this case. On the other hand, the cat-neuron ***N*** of level 2 binding *P* in the figure 6 of Section 3 is of order 2; it is not 1-reducible because one of the distinguished links of *P* is complex.

4.4 The multiplicity principle as the source of emergence

The construction of cat-neurons binding patterns of cat-neurons of lower levels, and of their simple and complex links, can be interpreted as a computation on mental objects, comparing and combining them to form more complex

ones, thus it explains how to develop an algebra of mental objects (following the proposal of Changeux, 1983, p. 181).

It seems probable that most animals will only develop cat-neurons of complexity order less then, or equal to, 2. A characteristic of man is that he has the capacity of forming mental objects and cognitive processes of increasing complexity order. As we have seen, it relies on the possibility for a cat-neuron to admit several decompositions in non-interconnected patterns. This property generalizes to cat-neurons the degeneracy of the neuronal coding emphasized by Edelman (1989). Instead of 'degeneracy', we prefer to speak of 'multiplicity', saying that MENS satisfies the *Multiplicity Principle* (Ehresmann and Vanbremeersch, 1996). We have explained how the multiplicity (or degeneracy) at the neuron level extends to the cat-neurons of level 1. By iteration it extends to all the levels. (Categorically, the complexification process respects the multiplicity principle; cf. Appendix.)

To sum up, the root of the emergence of higher cognitive processes, up to consciousness (cf. Section 7) is the degeneracy of the neuronal coding. We had already shown this in 1996; later, Edelman and Gally (2001) have also insisted on the relation between degeneracy and emergence.

The neo-connectionist models of neural systems (following Hopfield, 1982), which operate at the sub-symbolic level, can only account for cat-neurons of level 1 (under the form of attractors of the dynamics). They cannot describe the interactions between attractors necessary to iterate the process and solve the binding problem at their level, leading to mental objects of increasing complexity. By contrast, the complexification process gives an explicit construction of the links, both simple and complex, between cat-neurons of any level, allowing for the binding of patterns of cat-neurons to construct more complex ones. Since cat-neurons model mental objects, it gives a solution, not only to the binding problem at the first level, but to a binding problem extended to each level. At the same time, the construction explains why usual methods in terms of assemblies of neurons fail for higher levels. Indeed, the correlation between a cat-neuron of level more than 1 and a synchronous hyper-assembly of neurons is intricate and non-univocal:

- It is intricate because the hyper-assembly of neurons is obtained via the dynamic stepwise unfolding of one of the ramifications of the cat-neuron $\boldsymbol{N}$ down to the neuron level; let us recall that a ramification consists of a pattern P of cat-neurons P_i of level $\leq$ n having $\boldsymbol{N}$ for its binding, each P_i binding a pattern P^i of cat-neurons of lower levels, and so on down to patterns of neurons. For instance for a ramification $(\mathrm{P}, (\mathrm{P}^i))$ of length 2, the neurons of the hyper-assembly are all the neurons of the various patterns P^i; but to recover the cat-neuron $\boldsymbol{N}$ we must also take into consideration the distinguished links of the pattern P; if some of these links are complex, they reflect global properties of the neuron level, not observable at this neuron level but actualized at the higher level (cf. Section 3).

- It is non-univocal (or 'degenerate') because a cat-neuron may have several ramifications, not necessarily interconnected.

In terms of mental states and brain states, this correlation gives a new approach to the *brain-mind problem* (cf. Section 8).

5 Self-regulation of the dynamics

MENS is more than an evolutive system, In Ehresmann and Vanbremeersch (2007) it is presented as an example of a memory evolutive system. We have introduced the memory evolutive systems in 1991 to model autonomous complex natural systems, such as biological or social systems. They are evolutionary systems with an auto-organization directed by a net of internal regulatory organs, called coregulators with only partial information on the system; they have the capacity to learn from their experiences by recording them in a memory from which they can be recalled later in analog situations. The coregulators direct the dynamics of the system in a more or less cooperative/competitive manner, with recourse to the central memory.

5.1 The memory and the coregulators

In MENS, the memory models the long-term memory of the animal. It is a hierarchical evolutive sub-system whose cat-neurons are called *records*, It is divided into:

- the empirical memory, itself divided into: (i) the perceptual memory whose records model the mental images of items perceived by the sense organs, or internal states (hunger, pain, joy, ...) and (ii) the episodic memory with records of particular events or personal experiences,
- the procedural memory with records of motor skills, behaviors or procedures. A record Pr in it operates through the activation (via 'commands') of a pattern E_{Pr} of cat-neurons modeling its effectors; these cat-neurons operate internally by activating other cat-neurons, or act externally (e.g., by activating muscles). The result of the activation by Pr of its effectors is recorded by a cat-neuron N which binds the pattern E_{Pr} (cf. Figure 9).
- Higher animals develop other kinds of memory: a semantic memory, and the archetypal core, a personal memory at the basis of the self (cf. sections 6 and 7).

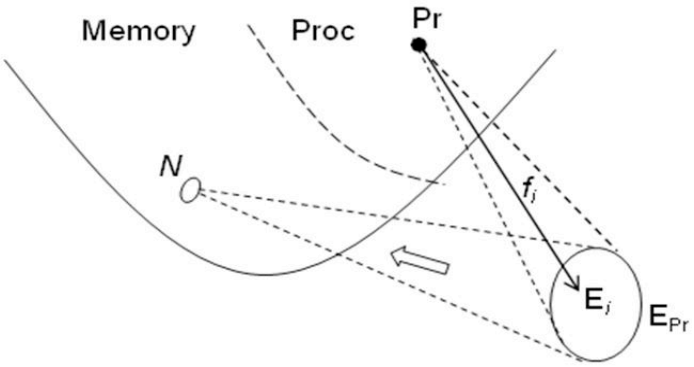

FIGURE 9. A record Pr in the procedural memory Proc can activate a pattern E_{Pr} of cat-neurons, its effectors, through links f_i (commands) from Pr to effectors E_i. The result of this activation of its effectors is memorized by the record *N* which binds the pattern. E_{Pr}.

As said above, at birth MENS is somewhat reduced, and the animal must progressively learn to recognize more objects, to perform more complex comportments, possibly to evaluate their results and remember them for best adapting to various situations. As a result, MENS is progressively extended by the formation of more cat-neurons, obtained through successive complexifications. This extension is not pre-programmed, but will be internally directed by the animal.

The whole comportment of the animal, his actions, his internal states, the formation of his mental objects are all dependent on his nervous system, hence reflected in the evolutive system MENS which it generates. The regulation system responsible for his comportment must be modeled in MENS. However there is no internal 'homunculus' able to have a global vision and to impose its choices. The control is distributed among a net of internal regulatory organs, the *coregulators*, able to collect partial information on the internal and external situation, select appropriate procedures, command their realization, evaluate their results and later participate in their recording in the memory.

A *coregulator* is an evolutive sub-system of MENS based on a particular more or less extended part of the neural system; here based means that its cat-neurons (called its *agents*) have ramifications whose lowest cat-neurons are in this part, so that they are activated by (hyper-)assemblies of neurons situated in this part. The existence of a kind of modular organization in the brain (as emphasized by Fodor, 1983) is now generally accepted; among the possible bases for a coregulator we distinguish various modules: systems of receptors (for vision, audition, smell,...); systems of internal or external effectors (connected to muscles); more or less specialized dedicated areas (in the visual areas, motor cortex, hippocampus, temporal cortex, brain stem, limbic system,...), but also smaller ones (e.g., the treatment units considered by Crick (1994) in vision, such as a color module processing colors). The coregulators are more or less complex depending on the complexity order of their agents; we speak of lower coregulators (their agents model neurons or assemblies of neurons), and higher coregulators.

Each coregulator operates by steps, delimited by its own discrete timescale; the duration of a step is related to the propagation delays of the links which activate its agents and to their refractory periods. It has a differential access to the (central) memory and has a characteristic 'function', determined by its *admissible procedures*, which model the actions it may command:

- For a lower coregulator, they may just consist in an automatic transmission of the information received (through the activation of some agents) from other more complex coregulators or from effectors. For example, a color module will transmit the various characteristics of the colors it perceives to higher visual areas.
- For other coregulators, there are particular admissible procedures modeled by records in the procedural memory with the following properties: they can activate some of the agents of the coregulator; and conversely (some of) the effectors of their objectives can be commanded through agents of the coregulator. The coregulator also participates in the later storage in the memory of the new information it has received, the responses it has triggered and their result.

The dynamics of MENS depends on the 'local' procedures of its coregulators, but the 'global' operative procedure which will really be implemented at a given time is the outcome of an equilibration process which makes the commands sent by the various coregulators as coherent as possible, with a possible fracture for those whose procedure will not be realized.

5.2 Local dynamics of a coregulator.

We have said that a coregulator has its own discrete timescale; a *step* of the coregulator extends between two consecutive instants, say t to t', of this timescale. Let us describe the unfolding of this step for a particular regulator. During this step, the coregulator can be looked at as the pattern CR of its agents and their distinguished links. The step is divided in 3 more or less overlapping phases, the two first ones add up to the *actual present* of CR during which its agents are activated, the last one corresponding to their refractory periods.

5.2.1 First phase: Collect of information (or decoding)

During this phase, CR collects information on the internal state of the animal and/or the external situation, thus forming what we call its *landscape* at t. This information consists in the activation of some agents by various cat-neurons. For instance, if CR is based on a color module and the animal perceives a blue object at t, the 'blue' agents are activated by the record of the object, but a 'red' agent is not. The aspects of the system which can be seen by the coregulator are modeled by the links b from a cat-neuron N to an agent of CR; this aspect is *t-activated* if N activates this agent along b during the actual present. The distinguished links between agents are supposed to be strong enough to transmit this information, so that the whole perspective b^* of N generated by a t-activated aspect b is formed of t-activated aspects; we say that the perspec-

tive is *t-activated*. The same N may have several t-activated perspectives for CR. Two t-activated perspectives of N and of N' are correlated if there is a link from N to N' correlating their aspects. If there is a cat-neuron binding the coregulator, a t-activated perspective b^* of N binds into a link cb^* from N to cr in MENS; this link activates cr during the actual present.

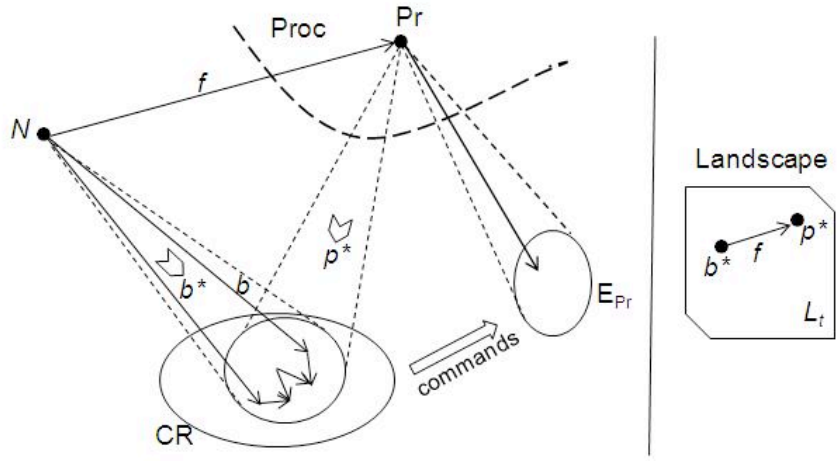

FIGURE 10. A perspective b^* of N for a coregulator CR is a maximal set of links (or aspects) b from N to agents of CR which are correlated by zig-zag of links in CR. It is t-activated if N activates the agents along all its aspects around t. A link f correlates two t-activated perspectives if it correlates their aspects. Here f is an activator link from N to the record Pr of an admissible procedure for CR. When b^* is t-activated, N activates Pr along f, and the perspective p^* of Pr is t-activated. CR can select Pr, and send the corresponding commands to effectors of Pr. The t-activated perspectives and their links form the landscape L_t of CR at t, modéled (on the right) by a category.

The t-activated perspectives and the links which correlate them form the *landscape* of the coregulator at t (cf. Figure 10); it can be compared to the 'perspective space' of Russell, 1971, from which we have taken the word perspective. The landscape accounts only for the part of MENS which is perceived by the coregulator during its actual present; CR can only collect information from the pattern of the cat-neurons N which have a t-activated perspective for CR. Categorically, we model the landscape by a category L_t and measure the loss of information for the CR by the *difference functor* from L_t to MENS which maps a t-activated perspective of N on N.

5.2.2 Second phase: selection of a procedure

Depending on the information received by the coregulator in its landscape, an admissible procedure is selected to respond to the situation. For a lower coregulator, the selection is automatic; for instance a color module will recognize a blue object via a perspective, and transmit this information to higher visual areas. Higher coregulators have admissible procedures recorded in the procedural memory. Some of them may have a t-activated perspective; in this case, one of them (generally the one whose perspective has the greatest strength) is selected.

In particular, the selection takes account of earlier experiences which have been memorized. If a similar situation has already occurred and a successful procedure has been used, this result is recorded via the formation of an *activa-*

tor link *f* from a record *N* of the situation to the record Pr of the procedure; this link correlates their perspectives for CR. In this case the activation of *N* leads to that of Pr, so that the perspective of Pr is *t*-activated, prompting CR to select the procedure. For instance, the recognition of a predator (activation of its record) activates an escape procedure.

If the situation is unknown and there is no admissible procedure with a *t*-activated perspective, a procedure already used in a not too different situation can be adapted, or a new procedure formed. There is a *fracture* if the step must be interrupted because no procedure is found.

5.2.3 Third phase: command and evaluation

The agents transmit the objectives of the selected procedure, in particular activating the effectors which they can command. This activation is carried through the unfolding of a particular ramification; for instance a specifically adapted motor gesture is chosen to uplift an object. The result is evaluated at the end of the step, by comparing the new landscape $L_{t'}$ with the *anticipated landscape,* in which the objectives of the procedures would be realized. If all the objectives are not achieved, they are more or less different, and this difference will have to be compensated later on to avoid a fracture. (In categorical terms, the anticipated landscape is the complexification AL_t of L_t with respect to the procedure, and the difference is measured by a *comparison functor* (if it can be formed) from AL_t to $L_{t'}$).

At the next step, the coregulator will participate to the formation of a record of the situation (if it had not yet been learned), of the procedure used and of its result. This result can be recorded by the formation (or, if it already exists, the strengthening) of an activator link from the record of the situation to that of the procedure.

One important cause of fractures is the non-respect of the *structural temporal constraints* which relate the length of the step, the mean propagations delays of the links in the *t*-activated perspectives, and the stability spans of the activated cat-neurons (cf. Ehresmann and Vanbremeersch, 1996, 2007). Indeed, there must be time enough for circulating the information among agents, selecting a procedure and sending commands to effectors, during which the necessary cat-neurons must be able to be activated. For instance, if the animal does not see a predator soon enough, a procedure of running will not be started in time for the animal to escape.

If these constraints cannot be met during successive steps, there is a *dyschrony*, which might necessitate a change in the period (mean duration of the successive steps) of the coregulator; we speak of a *resynchronization*. For instance, when he ages, the animal will move more slowly.

5.3 Global dynamics

At an instant *t* of his life the animal is confronted with various events in his environment and must produce adapted responses. The situation is analyzed

through his different coregulators, which select a procedure and transmit its objectives to effectors. If all these objectives are compatible, they are all carried out. However, they are not always compatible, because some objectives of the various procedures can be conflicting or not realizable; *e.g.* if a coregulator sends a command to activate an effector which is simultaneously inhibited by another coregulator. For instance, a motor gesture to uplift an object can be interrupted by a higher coregulator which measures that it is not well directed, or because the object is too heavy. The problem is that the coregulators must operate coherently while there are many reasons for their selected procedures not to be compatible: the coregulators receive only partial information on the global situation through their landscapes; they function at their own rhythm and with specific structural temporal constraints; and they compete for the common resources of the system.

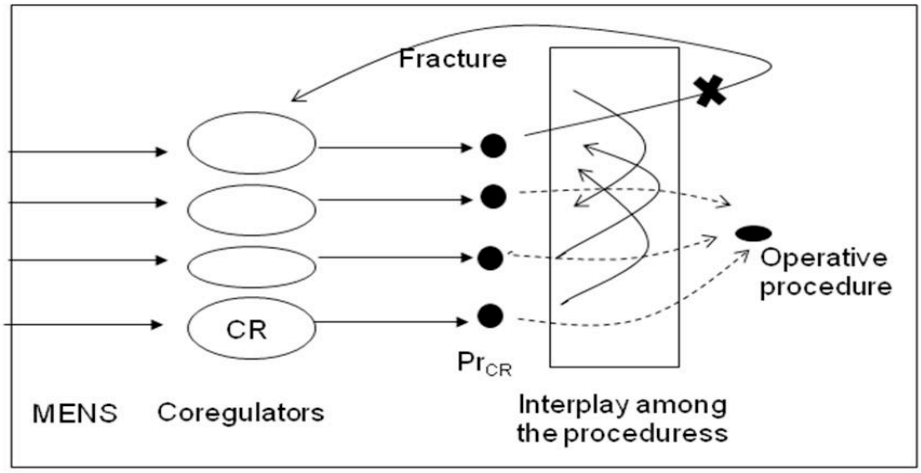

FIGURE 11. At a given time, there is an interplay among the procedures selected by the various coregulators to harmonize their objectives. The operative procedure which will be carried out is the result of this interplay. It will cause a fracture to a coregulator when (some of) the objectives of its procedure are not retained.

Thus there is a need for an equilibration process, or *interplay*, among the procedures, to determine a global *operative procedure* keeping as much as possible of the objectives of the various coregulators (cf. Figure 11). It is this operative procedure which will be finally carried out, possibly causing fractures to coregulators whose objectives are not realized. Categorically, the transition is modeled by the complexification of $MENS_t$ with respect to this operative procedure.

There is no general rule to determine the operative procedure. The interplay among the procedures takes into account:

- the strengths of the commands sent by the various coregulators; the objectives of coregulators with higher order agents generally prevail (the role of 'intentional' coregulators will be discussed later on, cf. Section 7);
- the plasticity of the cat-neurons allowing the unfolding of one ramification rather than another for activating a particular cat-neuron. For

instance, to catch a flying ball, there must be an adaptation of the various motor commands to the position of the ball, which amounts to complex switches between ramifications of the motor effectors;

- the temporal constraints which impose a kind of dialectic between two coregulators with very different rhythms. The coregulator CR, with a longer period, cannot be informed in real time of the small changes due to a lower coregulator with a much shorter period, because of the propagation delays or because they do not individually affect the stability of higher cat-neurons. However, the long term accumulation of small changes makes the unchanging landscape of CR more and more unreliable, ultimately causing a fracture to CR; for instance a fall can be due to a sequence of ill-adapted small equilibration gestures. To repair its fracture, CR will have to initiate a new procedure, which may retroact sooner or later on the lower coregulator (cf. Ehresmann and Vanbremeersch, 1996).

5.4 Development of the memory

During his life, the memory of the animal develops; he learns to recognize more items, to perform new skills; he memorizes successive experiences, his responses to them and their results. Let us see how the different coregulators participate in the formation of a new record. It will be a good illustration of the interplay among the procedures discussed above.

How can the animal memorize a new item O? In section 2 we have said that O synchronously activates a pattern of neurons (or of cat-neurons for more complex items); this pattern is consolidated at each later occurrence of O, its links being strengthened according to Hebb rule. The record of O will be a binding of this pattern. The formation of this record is done by the various coregulators which get different information on O; for instance for a physical object, different coregulators treat its color, shape, direction, possibly its odor, the noises it emits; for a motor skill, motor modules associated to different parts of the body must cooperate. Each coregulator CR will record the information received in its landscape and form a new cat-neuron called the CR-*record* of O; the operative procedure resulting from the interplay among the procedures will bind these partial records associated to the different coregulators into the global record of O.

Let CR be one of the coregulators. At the instant t, O synchronously activates a pattern P of cat-neurons; only a sub-pattern R of P has t-activated perspectives for CR; for instance if O is a blue object and CR a coregulator based on the color module, R activates the agents treating the characteristics of the color blue. The t-activated perspectives and the links of R correlating them form a pattern R* in the landscape L_t of CR at t (cf. Figure 12). The procedure selected by CR will be to bind this pattern (in categorical terms, form a colimit of R* in a complexification of L_t). This objective is transmitted (via the difference functor) and participates in the interplay among the procedures. Once retained in the operative procedure, it is realized by the formation of a cat-neuron binding R, called the CR-*record* of O. This CR-record may have no aspect for any agent of the coregulator. However, if CR (as a pattern) has a binding cr, each t-

activated perspective binds into a link to cr, and the pattern R* of these links defines a collective link from R to cr; the collective links bind into a link *c* from the CR-record to cr, which can be thought of as a 'generalized' aspect of the CR-record, along which O can 'globally' activate the coregulator.

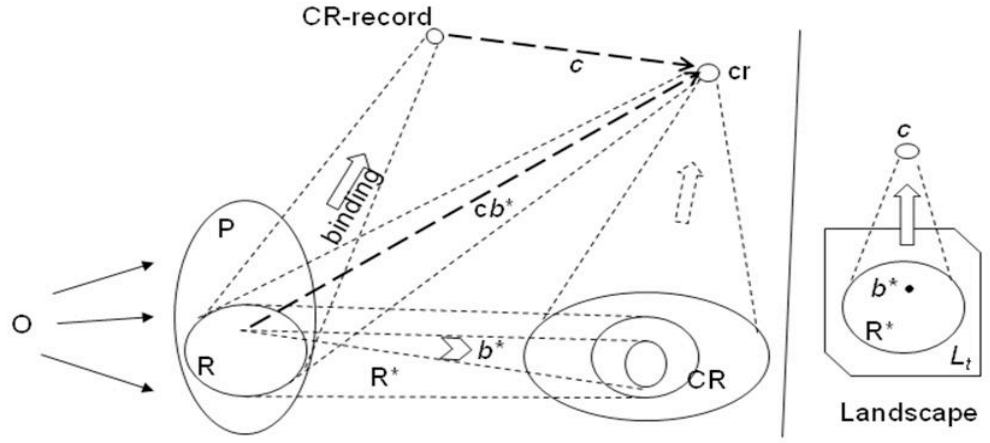

FIGURE 12. The item O activates a pattern P of cat-neurons; R is the sub-pattern whose cat-neurons have a *t*-activated perspective b^* for the coregulator CR. The selected procedure of CR will have for objective to bind the pattern R^* of these perspectives in its landscape L_t. It is transmitted (via the difference functor) into the objective of binding R, retained in the operative procedure; the binding of R is the CR-record of O. If there is a cat-neuron cr binding CR, each perspective b^* binds into a link cb^* to cr; these links form a collective link from R to cr which binds into a link *c* from the CR-record to cr. This link is a 'generalized' aspect of the CR-record.

Each coregulator which receives information from the item O selects a procedure to form its own record of O at its own rhythm. Once all the coregulators have transmitted their objectives, the interplay among procedures collects them in the operative procedure. It leads not only to the formation of the various CR-records, but also to the formation of a cat-neuron which simultaneously:

- binds the pattern P of all the cat-neurons activated by O; and
- binds the pattern formed by the partial records.

This cat-neuron is the *record* of O. Its later recall consists in the simultaneous activation of its partial records, leading to the activation of the record through the unfolding of one of its ramifications.

6 Semantic memory

What we have said up to now can be applied to many animals, in particular mammals and birds; even if there are differences in the complexity of the mental objects they can form. Higher animals alone (particular mammals and birds, and specially man), are able to form cat-neurons of complexity order greater than 2; and these animals are also able to develop another capacity to detect invariants in their environment. It is easy to compare various objects when they are simultaneously observed and to classify them according to some common feature or attribute, for instance to classify some geometric figures according to their shape; experiments with the most varied animals

have proved they are capable, given a little number of triangles and circles, to distinguish between them. What is more complicated is to classify these objects through their records, and still more to form a mental object representing the invariant they have in common; this mental object, or rather the cat-neuron which models it, will be called a *concept* (following Changeux, 1983; Edelman, 1989); for instance the concept 'blue' characterizes the class of all the objects which have this color. The concepts are at the basis of the development of a semantic memory. For us, the formation of concepts does not necessitate the use of language, hence is possible for higher animals; naturally for man, language will allow the formation of more elaborate concepts.

6.1 How to classify?

The notion of invariant depends on what features or attributes of the objects are considered; a blue triangle and a blue circle are similar with respect to the color, but not with respect to the shape. How to represent an attribute in MENS? Here, an attribute will correspond to the kind of information which can activate particular coregulators; such as the color (for a coregulator based on a color module), the odor, the shape, the orientation, and so on. In other terms, the attributes are associated to the function of a coregulator, and to classify some items according to an attribute will mean to classify these items (or rather their records) in function of the information they transmit to the corresponding coregulator.

It should be noted that the division of the brain in modules relies on the characteristic operations that these modules can perform, and these have been determined with the help of experiments, for instance, the neurons of a color module discriminate between the colors; but these experiments presuppose a classification into colors, and prove that such a neuron is activated by such a color. For other animals (e.g. dolphins) which have different environments and capacities, the 'attributes' can be different, and other kinds of coregulators should be taken into account for the formation of concepts. Here we see the difficulty to study our own mind (modeled by MENS); it is the 'self-reflection' problem which underlies all discussion or theory about the mind.

Finally, we interpret a classification according to an attribute as meaning a classification with respect to a particular coregulator CR. This classification is done in 2 steps: first a 'pragmatic' classification indicating if two items should be considered as 'CR-similar', then a more formal classification (operated through a higher coregulator) associating a CR-concept to a class of CR-similar records. After that, we describe how these CR-concepts with respect to various coregulators can be combined to get more elaborate concepts.

6.2 Classifier cat-neurons

The CR-concepts and the general concepts are mental objects which will be modeled by cat-neurons of a different type than those constructed in the preceding sections. While the 'binding' cat-neurons model classes of synchronous assemblies which *activate* the same cat-neurons, the *classifier cat-neurons* will model classes of synchronous assemblies which *are activated* by the same cat-

neurons. Roughly, they are characterized not by the information that they can transmit, but by the information that they can receive.

Only higher animals have the capacity to form such classifier cat-neurons. Up to now, the only requirement for the neural system was the capacity to transmit information through the activation of neurons, to bind synchronous assemblies of neurons, and repeat these operations at least once. Now a supplementary capacity is required: to compare the information transmitted by two items (not necessarily at the same time), and to have coregulators of a sufficient complexity level to detect what is common to both and classify it; thus the animal must be endowed of a kind of partial 'self-reflection' over its internal operations.

Given a pattern Q of cat-neurons, the information globally received by Q from a cat-neuron *N* is modeled by a *distributed link* from *N* to Q, which is defined as a family of links from *N* to each Q_j well correlated by the distinguished links of Q; we say that it globally activates Q. It should not be confounded with a perspective of *N* for Q which may only activate some of the cat-neurons of Q, while a distributed link really 'distributes' the activation between all the Q_j.

A *classifier* of Q is a cat-neuron *C* with the following property: there is a 1-1 correspondence between the distributed links from a cat-neuron *N* to Q and the links from *N* to *C*. (In categorical terms, the classifier of Q is a *projective limit* of Q.) Roughly *C* receives the same information as the pattern Q as a whole; both have the same role as receptors. While a binding cat-neuron may bind several patterns, similarly a classifier cat-neuron may classify several patterns. As for binding cat-neurons, there are simple and complex links between classifier cat-neurons (cf. Appendix).

Classifiers will emerge from a *mixed complexification* process (cf. appendix) with respect to a *mixed procedure*. A mixed procedure is a procedure which has a supplementary objective: to form the classifier of some given patters. (Cf. Ehresmann and Vanbremeersch, 2007; and also the Appendix.)

For animals able to perform some kind of classification, the procedures selected by the coregulators, as well as the operative procedure carried out on MENS, can be mixed; their realization necessitates a mixed complexification process. In this case, MENS is deduced from **Neur** by successive mixed complexification processes, so that not only bindings but also classifiers can emerge.

6.3 CR-*concept*

Let CR be a coregulator, P the pattern of cat-neurons activated at a given time *t* by an item O, so that P is a decomposition of the record of O if it exists. What trace will it imprint on the coregulator? In some case, there is no trace, therefore no CR-concept: a musical sound does not activate a visual coregulator, and it will have no color-concept. A trace will be formed if some of the cat-neurons of P have *t*-activated perspectives for CR and we denote by R the sub-

pattern of P they form. R activates a sub-pattern of CR which we call the CR-*trace* of P and denote by trP; it consists of the agents activated by a cat-neuron of R, indexed by the aspects along which they are activated, its distinguished links are the links of CR correlating these aspects.

Another pattern P' is said to be CR-*similar* to P if its trace trP' is globally activated by the same cat-neurons, and with the same strength, as trP (so that the same cat-neurons send distributed links to P and P'). If O has a record, all its decompositions are CR-similar. However this CR-similarity is not limited to the decompositions of the same record. Another item O' could have a record whose decompositions are CR-similar to those of O; in this case we say that O and O' (or their records) are CR-*similar*. For instance two blue objects will be similar for a color module.

The relation of CR-similarity leads to a classification of items (or records) into classes of CR-similar items. This classification is only 'acted' by the coregulator, meaning that all the records in one class leave CR-similar traces. It is not 'internalized' at its level: CR cannot itself distinguish the CR-similarity of two records, because it would imply a kind of self-reflection of CR on its operations. The CR-similarity can only be apprehended if there is a higher coregulator which receives perspectives from the CR-traces and can recognize their CR-similarity. For a lower enough coregulator CR, the existence of such a higher coregulator does not impose too stringent conditions on the neural system. In fact, CR-records are essentially constructed for lower coregulators of complexity order less than 2.

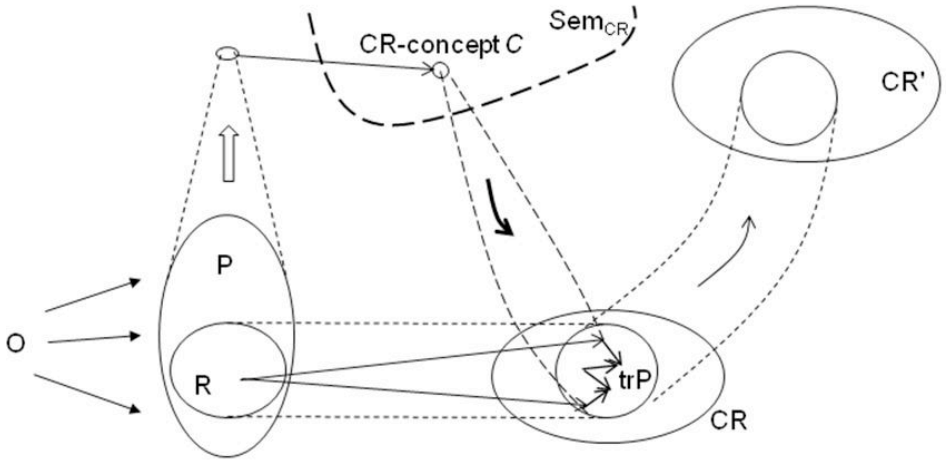

FIGURE 13. The item O activates a pattern P of cat-neurons; R is the sub-pattern whose cat-neurons have *t*-activated perspectives for the coregulator CR. The CR-trace trP of P is the pattern of agents activated by R and the links correlating them in CR. The CR-concept *C* of the record *M* of O is the classifier of trP added through a mixed complexification process directed by a higher coregulator CR'. The defining link *d* from *M* to *C* defines *C* as the 'best approximation' of *M* in the CR-semantic memory Sem$_{CR}$.

The classification into CR-similarity classes is consolidated by the formation of a cat-neuron, called a CR-*concept*, associated to a class of CR-similar items, for instance the color-concept 'blue' to the class of objects whose color is blue. Formally, the CR-*concept* of O, or of the record *M* of O, is defined as the classifier of the CR-trace of a decomposition P of *M* (cf. EV 1992, 2007); it does not

depend on the choice of the decomposition since all its decompositions are CR-similar, hence have the same classifier if it exists. The formation of the CR-concept will be initiated by a higher coregulator, which will take the formation of the classifier as one of the objectives of the mixed procedure it selects (cf. Figure 13).

The constructions of the CR-record *M* of the item O, and of the CR-trace of its decomposition P should be well distinguished.:

- To construct the CR-record, we are interested in the *t*-activated perspectives of O, taken globally (corresponding to links toward the binding of CR if it exists); two aspects in the same perspective are not treated separately; what is important is the global information that CR itself can collect from O and can later use for selecting its procedures and sending commands.
- On the other hand, in the CR-trace we are interested not in perspectives taken as a whole, but in the agents activated along all the different aspects in these perspectives, and the links correlating them in CR. The CR-trace compares with the shape of P with respect to CR (in the sense of Borsuk shape theory; Borsuk, 1975); and the CR-concept converts this shape in an invariant.

Briefly, the CR-record retains how the object can activate other cat-neurons, while the CR-concept extracts what remains invariant in the traces it impresses on CR. These 'dual' viewpoints explain that O and O' may have the same CR-concept, while O and O' have not the same CR-record. Two blue objects have the same color-concept 'blue' whatever their other attributes (different shapes, sizes,...).

The different items (or their records) which have the same CR-concept are called *instances* of the concept. The 2-step construction sketched above shows that:

- We first recognize the CR-similarity of two records (through their traces) without having yet the corresponding concept; roughly, a class of CR-similar items is formed of items having "a family resemblance" (in the terminology of Wittgenstein, 1953).
- It is only in the second step that this class is consolidated through the formation of the CR-concept, which becomes an instance of itself; as such, it can play the role of a prototype for the class (in the terminology of Rosch, 1973).

As any cat-neuron, the concept will take its own identity, and it can acquire new instances later. As said above, at the time a concept emerges, only a few of its instances are already known, and it is initially formed to extract their invariant (as the classifier of their traces). Over time, the classification will be made more precise, e.g. by adding new instances or suppressing other ones. For instance, the child first will form a concept of moving objects encompassing cars and trains, then refine it by distinguishing between them.

A CR-concept emerges in the course of a mixed complexification process with respect to a procedure having its formation as an objective. This process also constructs the links between them and other cat-neurons (cf. Ehresmann and Vanbremeersch, 2007, Chapter 4). The CR-concepts and the links through which they communicate (in the role of classifiers) constitute the CR-*semantic memory*, which is modeled by an evolutive sub-system Sem_{CR} of the memory of MENS. For each record M of an item O which admits a CR-concept C, there is a link d from M to C, called the *defining link* which characterizes C as the 'best approximation' of M in Sem_{CR} in the sense that C activates each CR-concept which is activated by O (categorically this link defines C as a reflection of M in Sem_{CR}; cf. Appendix).

6.4 The semantic memory

The CR-concepts are 'concrete' in the sense that they reflect some specific property of the items they classify, for instance their color. Other concepts can be deduced from them, either more specific ones (a concept of 'blue circle'), or more general ones (a concept of 'dog') or more abstract (a concept of 'justice'). They will be successively formed by combining CR-concepts associated to various coregulators in different ways; leading to the development of the *semantic memory* which is an evolutive sub-system of MENS generated by the various Sem_{CR}.

The formation of a CR-concept starts from a small set of instances; then a mental object, the CR-concept, is formed to extract their similarities; later other instances are added to the concept by comparing their similarity with the concept. And the CR-concept of a record can be characterized in two ways:

- as the CR-concept which 'best approximates' the record (in the strict sense of a reflection in Sem_{CR});
- by the class of its instances, but this class varies over time; new instances can be found while some instances can be later eliminated; this elimination can be forced by a fracture in some higher coregulator if the instances have been wrongly classified because of a lack of knowledge or a poor observation.

The first characterization pre-supposes the existence of the concept, the second one is partly contextual.

The formation of more elaborate concepts will be different; it consists in combining already constructed concepts, and their instances will be recognized afterwards. For these concepts, the classes of instances are not separated: a record can be an instance of several concepts; for instance a spaniel is an instance of the concept 'dog', but also of the concept 'mammal'. More precisely, at a given time t, we consider the sub-evolutive system Sem_t of the memory of MENS containing the various Sem_{CR} associated to various coregulators CR with their links. Two methods are used to form new concepts (cf. Figure 14):

1. We define concepts with respect to several coregulators (or 'attributes' as explained at the beginning of the section). For instance, the concept 'blue triangle' will classify physical objects depending on both their

color and shape; it is obtained as the classifier of a pattern consisting of the color-concept and the shape-concept. More generally, given a pattern P of concepts P_i in Sem_t, a new concept C can be constructed as the classifier of this pattern, which will be added as a result of a following mixed complexification process. Its instances are the records M which are instances of each P_i and whose defining links d_i are correlated by the links in the pattern (thus forming a distributed link from M to the pattern); and the defining link d from M to C classifies them.

2. We define concepts binding several already constructed concepts; for instance the concept of mammals includes the concepts of dogs, cats, men,.... They are obtained as the binding C' of a pattern P' in Sem_t; an instance M' of C' is an instance of one of the concepts P'_j of the pattern; the defining link d' from M' to C' is the composite of the defining link of M' to P'_j with the attachment link of P'_j to C'.

In both cases, the new concepts emerge as the result of a mixed complexification process directed by the cooperation between higher coregulators with the objectives: in the first case, to classify the given pattern of concepts, in the second case to bind it. Thus, Sem_t is extended in a larger evolutive sub-system of the memory. And this process of constructing classifiers and/or binding of patterns of concepts already formed can be iterated. It leads to the formation of the *semantic memory* Sem.

Like any cat-neuron, a concept C takes its own identity over time, possibly acquiring more instances. Each instance M of C can activate the concept along its defining link. Conversely C can recall another instance M' by a 'priming effect' if M' is independently activated via a diffuse activation of the memory, the simultaneous activation of M' and of C strengthening the defining link (Hebb rule). Thus the activation of M can be transmitted to M' via C; we speak of a *shift* between the two instances.

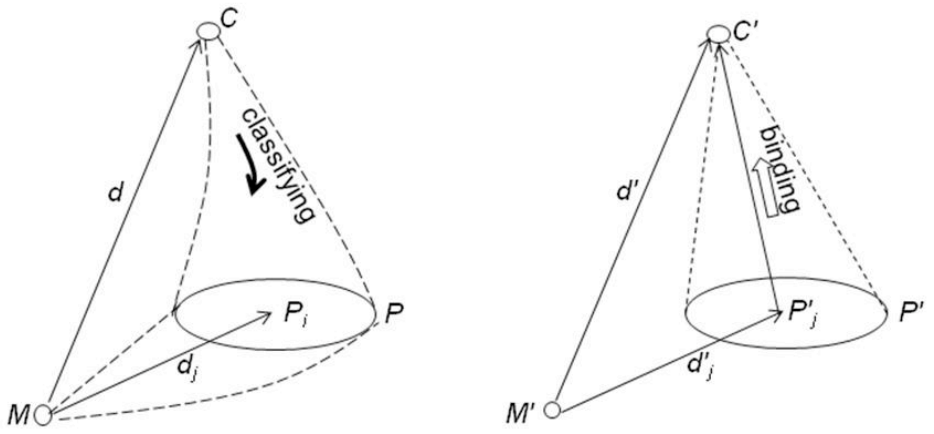

FIGURE 14. C is the concept which classifies a pattern P of concepts P_i. An instance M of C is a record which is an instance of each P_i and such that its defining links d_i to P_i form a distributed link to P'; this distributive link is classified into the defining link d to C. C' is the concept binding a pattern P' of concepts; its instances M' are all the instances of the various P'_j

The shifts between instances of a concept increase the plasticity, in particular in the selection of procedures and the interplay among them. The procedures of a higher coregulator can be recorded as concepts, and then commanded through the effectors of the most adapted instance: a movement such as walking can activate different patterns of muscles depending on the ground. And the interplay among the procedures acquires two kinds of degrees of freedom: possibility of shifts between instances of the different procedures, then complex switches between ramifications of their effectors.

We accept that higher animals other than man can develop a (more or less extended) semantic memory, since its construction is independent from language. For man, the language allows a still more abstract operation: to 'name' the concepts, that permits an economy of means (replace a concept by a name), and thus to handle still more abstract concepts.

7 Archetypal core. Conscious processes

The development of a semantic memory allows the development of a more complex and personal memory, the archetypal core, at the basis of the self. Though higher animals will develop such a memory, it will be particularly important for man, and we first consider this case.

7.1 Development of the archetypal core

At birth, the baby has an innate memory accessible by some lower coregulators which can command simple archaic motor reflexes and sucking reflexes. In the first days, the activation of one of these coregulators, each experience, each emotion, will be memorized, as well as the cause of this activation (internal or external sensorial stimulation) and its possible results if they are perceived. For instance *evaluating coregulators* (based on the emotive brain) are able to evaluate homeostatic drives and states related to pleasure or pain and to measure the consequences on the homeostasis and well-being of the baby; they form partial records which develop the value-dominated memory (Edelman, 1989, p. 99).

All these records extend the innate memory, so that more and more items can be recognized, in particular the emotive memory extends, and the power of action of the baby increases. Many simple and complex links are created in the memory, such as activator links towards records in the procedural memory; and the lower coregulators (in particular the evaluating coregulators), become connected to associative cortical areas, with creation of higher coregulators based on these areas.

During the first months, most of the experiences will be physical and/or affective, but they progressively are completed by a semantic approach, with a pragmatic classification in concepts with respect to coregulators based on the sensory and limbic systems. For the baby:

"I am hungry, I cry, I suck"; "I am hungry, I cry, I suck, it is good";
"I am hungry, I cry, I suck a breast which has a good odor"; and so on

Thus, a primal hard core is formed in the memory, with records of often repeated sensations or behaviors, and of notable experiences. We call it the *archetypal core*; it is an evolutive sub-system of the memory, with numerous powerful internal links whose strengths increase through their constant reactivation. The simple reflexes are replaced by more elaborate skills which are recorded in its part of the procedural memory. Each instant of the baby life activates part of this archetypal core which is the resultant of all his/her experiences and which reflects a memorization of the body through these experiences. The various sensations are remembered, and begin to be classified with formation of sensory concepts.

After a few months, the small child acquires other capacities, such as using some words, recognizing some music. However, throughout this period, the experiences are mainly corporal. The archetypal core is no more restricted to some adaptive reflexes, but has become a rich, more or less stable archive of what the body can do and feel. The later changes will enrich it by adjunction of details; for instance a biting cold may be pleasant if it is associated to games, the feeling of a caress will take a new color at the time of the first love around 10 years. These changes are very progressive, except when there are serious fractures which modify the corporal image such as pain, illness, or violent emotions.

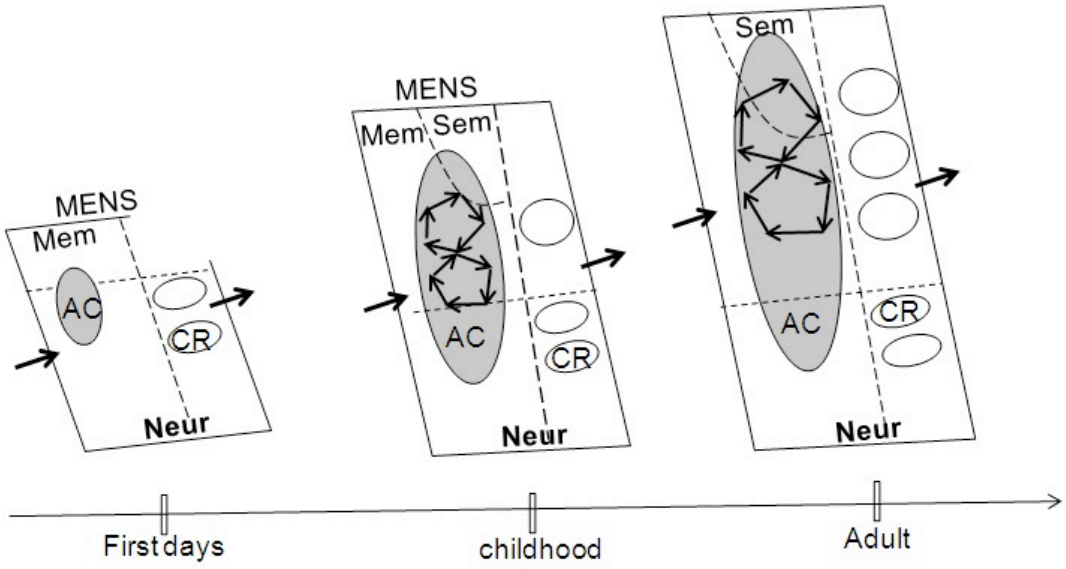

FIGURE 15. Development of the archetypal core AC, a sub-evolutive system of the memory which plays a central role. At birth AC contains a few innate records. During childhood, AC integrates the records of the main persistent physical, mental and emotional experiences, with strong and fast links between them, the links in fans. These links form 'archetypal loops' through which the activation of AC is self-maintained for a long time. AC extends;mo(re slowly during adulthood; its stability is at the root of the notion of Self.

Over time the archetypal core will extend by recruiting more cat-neurons (cf. Figure 15). A record which has a preferential link to an archetypal record will become archetypal if this link strengthens; for instance the memory of a very emotional event, or of an object evoking significant childhood experiences.

However, the extension is much slower than during the childhood, and the archetypal core remains stable enough.
Higher animals other than man also develop an archetypal core, but it has less extension and plasticity, since the animal has to obey to a larger number of innate procedures (means of defense, flight, moving, eating...), he becomes an adult more quickly and probably cannot form cat-neurons of complexity order more than 2, nor abstract concepts. On the other hand, the human baby has not many innate procedures and comportments (hunger, sleep, pain,..). The prime childhood, then the childhood being very long compared with those of the animal, (s)he has more time to construct the archetypal core on experiences related to the environment, education, games.

7.2 Role and structure of the archetypal core

The archetypal core is a permanent memory, developed from the first days on, with often reactivated records intermingling strong memories of the body, its sensations, feelings, emotions, and of the basic procedures associated to them. The links between them are strong and fast, and they are continuously strengthened up to a threshold. It keeps its identity over time, with only slight modifications, thus contributing to the notion of self. It remains in the background, where it acts as a filter, a referent. Each experience activates a semiotic search in it through several higher coregulators, which allows accounting for its sensory and emotional overtones (their importance has been stressed by Damasio, 1999). Being both a filter and a mirror, it modifies the experience in the light of past experiences, as Proust's "madeleine" well illustrates. It circulates the information in loops between various areas, acting as a kind of intranet in the middle of the diffuse neural noise which ensures its permanence.

We had introduced the archetypal core in 1999, as a hypothesis which seemed a natural consequence of our model, compatible with Edelman's view on the importance of the thalamo-cortical loop which supports reentrant activity among various areas. Recently this hypothesis has been confirmed by neuroscientists (P. Hagmann *et al.*, 2008) who have discovered an area in the median posterior cortex to which they attribute exactly the properties needed for its development. This area, which they call "neural connection core", seems the most densely connected zone of the brain, it has the largest energy consumption, even at rest and it is fed by a double artery; they suppose that it is related to consciousness since its activity decreases under anesthesia. And for these authors, it plays an essential role in the integration of information, exactly what we suppose.

In agreement with this, in MENS, the archetypal core is modeled by an evolutive sub-system AC of the memory, based on the neural connection core, which integrates and intertwines recurring sensorial, proprioceptive, motor, emotional, procedural memories and their concepts, as well as notable experiences. Initially it would consist of neurons in this neural core, then higher cat-neurons emerge (through successive mixed complexifications) as the bindings of patterns of these neurons, and later they are classified into concepts also in AC. Each archetypal record (i.e., cat-neuron in AC) has multiple, possibly

non-interconnected, ramifications down to this neural core, each archetypal concept has instances which are archetypal records.

An archetypal record is linked to other archetypal records by very strong links whose activation is self-maintained through specific loops. More precisely, we suppose that, for each cat-neuron A in AC, there is a bundle F(A), called a *fan*, of strong complex links activating A in AC with the following property: there are loops formed of successive links belonging to fans which propagate very quickly the activation of A back to itself; such a loop will be called an *archetypal loop*. [Categorically, we suppose that the fans are covering families for a Grothendieck topology they generate (cf. Appendix), so that AC becomes a site; cf. Ehresmann and Vanbremeersch, 2007.] When an archetypal record is activated, links in fans propagate this activation to other archetypal records. This activation resonates to lower levels via the unfolding of a ramification, and, through complex switches, to other, possibly non-interconnected, ramifications. The activation of an archetypal concept resonates to an instance and, through shifts, to other instances and their ramifications. Thus, an extended part of the archetypal core resonates (this stochastic resonance has been experimentally observed in the brain; cf. Collins *et al.*, 1996; Levin and Miller, 1996; Wiesenfeld and Moss, 1995).

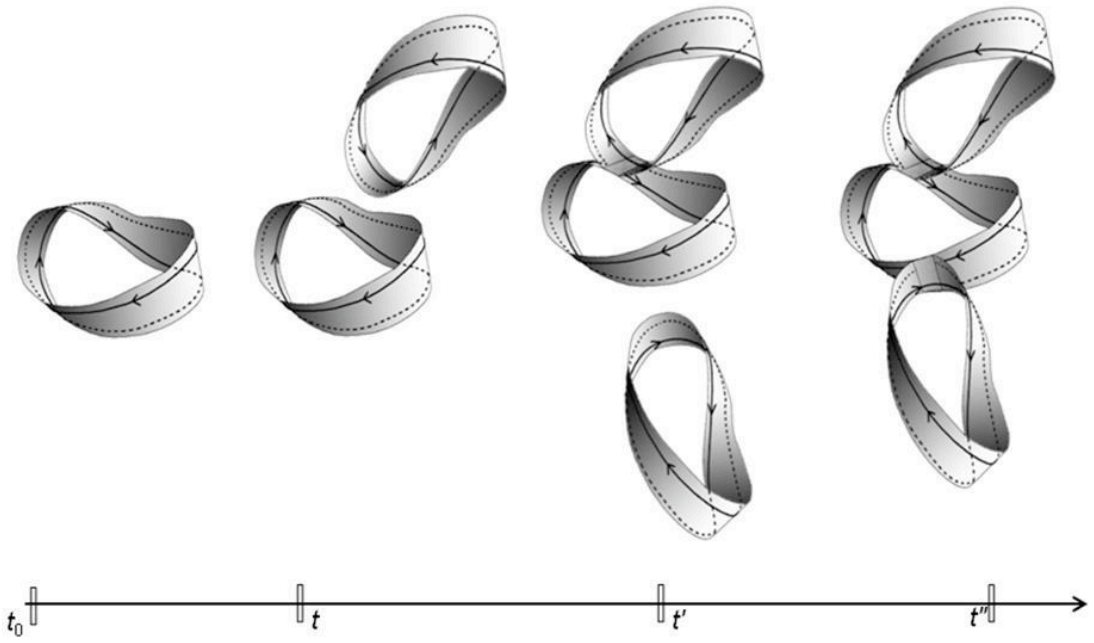

FIGURE 16. A metaphoric representation of the development of the archetypal core. First il is reduced to one Möbius band. An archetypal loop (sequence of successive links in fans) appears as a circuit drawn on it; such a loop propagates and maintains the activation. Later another Möbius band appears; it is glued to the first one so that an archetypal loop can propagate the activation through the two bands. A third band is later glued to the two first ones, and an archetypal loop propagates the activation to the three bands. And so on.

A geometric metaphoric image of the situation could be given by the surface gluing together several Möbius bands. Initially the archetypal core would be reduced to one Möbius band, on which an archetypal loop appears as a circuit. Later it extends by adjunction of records, concepts and their links; an-

other Möbius band is added, and it is glued to the first one; now an archetypal loop can go from one band to the other by crossing their common part. And the process goes on, gluing together more and more Möbius bands (cf. Figure 16); the activation propagates along the circuits formed by the archetypal loops, bouncing back and forth between the various bands.

7.3 The halo and the intentional coregulators

The activation of part of the archetypal core can extend to some cat-neurons *N* outside, but 'near enough' of AC for being integrated in a loop crossing AC. More precisely there is a loop from *N* to *N* consisting of a link from *N* to an archetypal record *A*, then link(s) of fan(s), and finally a link from an archetypal record *A* to *N* by which the activation of *A* can be transmitted to *N* and then self-maintained through the loop (cf. Figure 17). These cat-neurons *N* form the *halo* of AC and may later be integrated in AC, the loop becoming an archetypal loop.

At a given time *t*, an activation of some archetypal records triggers, through archetypal loops, a self-maintained activation of a large domain of the archetypal core, which propagates to part of its halo; if the activation of a cat-neuron is strong enough, it bounces down through the unfolding of a ramification, with possible complex switch to another; the activation of a concept goes back to an instance, with possible shifts between instances. All these activated cat-neurons and the links transmitting the activation form a sub-system D_t of MENS, the *t-activated domain*; its activation is sustained by the long-term activation of the archetypal core, so that it persists for a long enough period which we call the *specious present* (in reference to James, 1890).

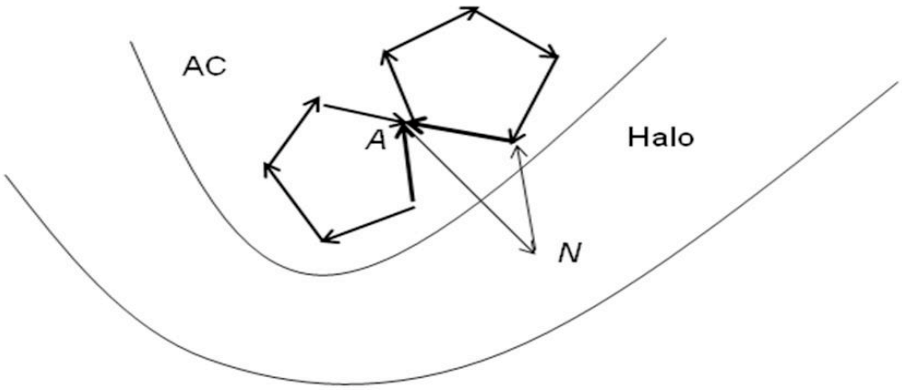

FIGURE 17. Two archetypal loops (formed of links in fans) maintain the self-activation of the archetypal record *A* for a long time. *N* is a cat-neuron outside AC activated by *A* and with the property: there is a loop of links from *N* to *N* containing at least one link in the fan of *A*; this loop self-maintains the activation of *N*. The cat-neurons *N* with this property form the halo of AC; if *N* is a record, it can later become included in AC.

We have explained how the functioning, dynamics and self-regulation of MENS depend on its net of coregulators and the interplay among their procedures. We have not explained how the procedures are selected; we have only said that lower coregulators have a few automatic procedures, and that other

coregulators have a number of admissible procedures (recorded in the procedural memory). Higher coregulators can form and/or learn new procedures, and memorize them in the procedural memory and possibly the semantic memory.

The development of the archetypal core and its halo allows the formation of higher coregulators based on the associative cortical areas and with agents in the halo. Thanks to a large access to this core and their participation in the activated domain, they collect more information and retain it during a longer period, have more opportunity to select and possibly create complex procedures, and evaluate their results, in particular through the feedback received from lower evaluating coregulators. Thus they have some capacity to internally control their own functioning, and we call such a coregulator an *intentional coregulator*, in reference to the "intentional systems" of Dennett (1990). An example is given by the "conscious units" of Crick (1994, p. 336). The *intentional net I*, consisting of these intentional coregulators and links connecting their agents, is essential in the emergence and development of conscious processes.

7.4 The global landscape

The archetypal core, with its persistent activity, gives a dynamic archive of the whole life, reflecting in the present the recurrent salient corporal, sensorial, proprioceptive, procedural or emotional experiences, and strongly interconnecting them. Thus, it reflects (as a mirror) the various components of the *self*, which we propose to define as the (virtual) binding of AC. Its role will be essential in the development of conscious processes. We consider the case of man, though we suppose that higher animals will also develop some kind of consciousness.

The model for consciousness to which MENS leads (cf. Ehresmann and Vanbremeersch, 1992, 2002, 2007) enters the frame of the "global neural network space paradigm" (in the terminology of Wallace, 2004, p. 2), and it has some relation with the models of Edelman (1989) and of Dehaene *et al*. (1998, 2003). In particular it relies on a long-term elaborate memory (afforded by the archetypal core) and on a modular control system intermittently acting on it, modeled by the intentional net.

A conscious process is initiated at a time t by an arousing event, of internal or external origin, which has no automatic response. For instance, it could be a fracture in one of the intentional coregulators, the start of a voluntary action, a sudden sensation of pain. The first response is an increase of attention (Edelman, 1989, p. 205) which activates cat-neurons based on several zones (in particular the reticular formation) connected to the archetypal core. It triggers a self-maintained activation of a large domain of the archetypal core, which extends the t-activated domain. For instance, an unexpected noise arouses all our senses and recalls similar noises and the associated events, and we try to identify it (e.g., going to the window to look for its causes). The t-activated domain constitutes a large 'working memory' (to be compared to the "theater"

of Baars, 1997). The intentional coregulators will use it to collectively extend their landscapes, both:

- 'spatially', collecting more diverse information, in particular in lower levels,
- 'temporally' going back to the recent past to find the causes of the arousing event ('retrospection'), and selecting more long-term procedures for the future ('prospection').

To model this, we define the *global landscape* GL_t *at t*. The intentional net *I* plays the role of a large higher coregulator in D_t; its agents are all the agents of the various intentional coregulators, and its actual present at *t* is the specious present. GL_t is its landscape. As for another landscape, it consists of the perspectives of the cat-neurons in D_t for the various intentional agents (in *I*) which remain activated during the specious present; the links are the links in D_t correlating them.

It is important to realize that a perspective of a cat-neuron *N* in the global landscape can be different from the perspective of *N* in the landscape *L* of one of the intentional coregulators, say CR. In *L* a *t*–activated perspective consists of aspects for the agents of CR; in the global landscape, the perspective may also include aspects for intentional agents which are not in CR, as soon as they communicate with agents of CR along a zig-zag of links in the activated domain. This extension of the perspectives allows for more cooperation between the intentional coregulators, which may exchange their information.

For instance, an intentional coregulator can observe a record *M* of the actual situation in its own landscape, but have no admissible procedure to respond. At the same time, *M* may have an activator link *f* toward (the record) Pr of an admissible procedure for another intentional coregulator CR, while *M* itself is not observable in the landscape of CR. Because of the surge of attention, the activator link enters in the *t*-activated domain D_t, hence figures in the global landscape. Thus Pr can be selected by the joint operation of the two coregulators.

7.5 Conscious processes: Retrospection and Prospection

For Merleau-Ponty, "consciousness unfolds or constitutes time", and time "is not an object of our knowledge, but a dimension of our being" (translated from Merleau-Ponty, 1945, p. 474-475). We agree with him and our hypothesis is that consciousness is characterized by two temporal processes: the retrospection (toward the past) and prospection (toward the future). Both will be operated in the global landscape.

The global landscape is maintained during a specious present through the self-activation originating from the archetypal core; thus it has a larger temporal span than the individual landscapes of the intentional coregulators. The retrospection process allows to reactivate ('intentionally' or not) events of the recent past, which are not observable in the individual landscapes, either because they have already faded from them, or even because they occurred at a

lower level. The prospection process makes use of the greater stability of the global landscape to select more adapted and/or complex procedures, possibly extending on a longer period, even planning on the long-term..

To explain the retrospection process, let us suppose that an arousing event S, say a fracture, occurs at t in the landscape of at least one of the intentional coregulators ("flashlight" in Baars' theater), say CR. For instance, S can be an unusual sound caused by the fall of an object in another room. As said above, it causes a surge of attention, which extends the t-activated domain D_t. The cooperation of the intentional coregulators is strengthened, and together they form the global landscape GL_t. The agents of CR have thus access to more information for identifying the nature of the original event and its possible causes. They can select (as their procedure) to operate an 'abduction' process (in the sense of Pierce, 1903) to recall similar past events and what had been their causes; it is done through a series of loops in D_t, among them archetypal loops. (Cf. Figure 18)

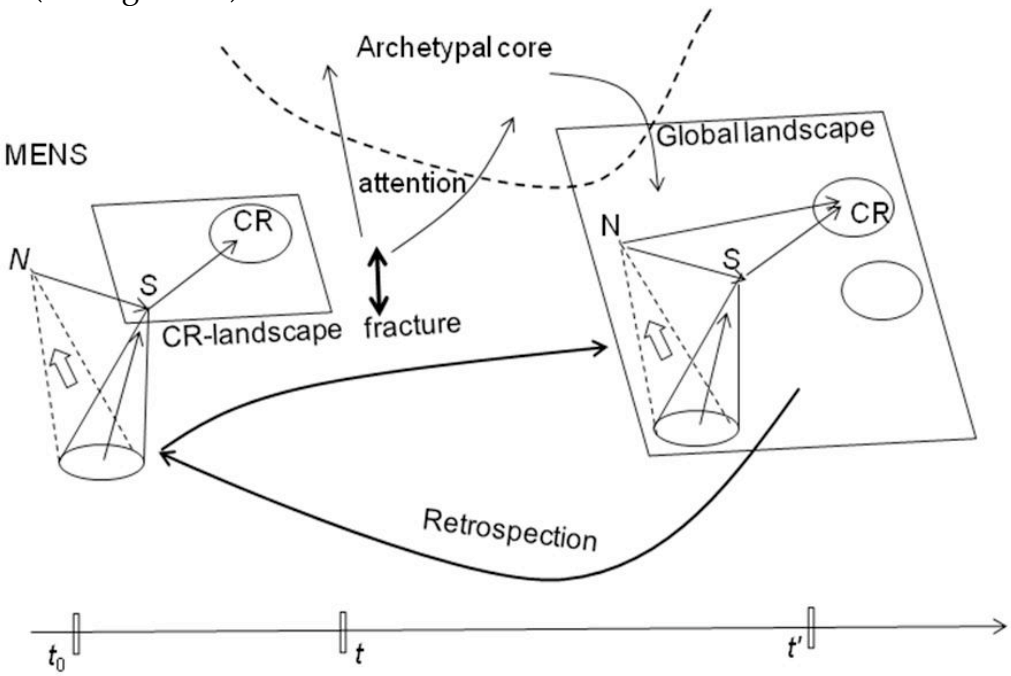

FIGURE 18. *N* activates *S* at t_0 which causes a fracture to an intentional coregulator CR by appearing in its landscape at *t*. In response there is a surge of attention which propagates through the archetypal core, increases the *t*-activated domain and leads to the formation of a global landscape. It allows a retrospection process recalling various components of *S* (possibly through the unfolding of ramifications of *S*), hence also their binding *N*. Thus the link from *N* to S becomes observable in the global landscape which unites and extends the landscapes of the various intentional coregulators.

First they try to recall as many as possible characteristics of the event; some of them were too weakly activated to be observable in the initial landscape of CR, but the increase of activation makes them observable in the global landscape. For instance, a retrospective analysis of the traces left by the sound S gives cues on the direction from where it came and its auditory characteristics. Have such characteristics already been associated to a sound in the past? This question re-activates a search for records of events having caused such a sound; it is done via loops in the activated domain, in particular archetypal

loops, all reflected in the global landscape. If this search recalls a unique record with the same characteristics, it is probably the cause of the sound, and the search stops. If several possible records are activated, a new search begins to trace back other weaker characteristics of the sound, thus refining its probable nature; and so on up to the retrieval of a unique possible cause *N*; for instance the fall of a book in the living room. If no similar event had occurred in the past, more or less different events can be recalled for, by comparison with them, trying to discover (or rather 'reconstitute') the real cause of S.

Anyway, the retrospection allows for a search in the past and on various levels, including lower levels inside the activated domain, unfolding ramifications, activating one instance of a concept or another. In the above example, the retrospection is directed by an intentional coregulator. However, it is not always the case. An example is given by the priming effect, which may orient the prospection process as follows. An object, say an apple, is subliminally presented to a subject so that its record has only a briefly activated perspective for a lower coregulator. Soon after (during the specious present), it is followed by the full presentation of a set of objects, among them an orange, and the subject is asked to select one of these objects. Which one will he select? Experiments prove that it is the orange. This selection is explained by the fact that the record of the apple is still weakly activated when the cognitive effort of the subject causes a surge of attention; the record of the apple is then reactivated and enters in the *t*-animated domain, where it recalls the concept 'fruit'. On the other hand, the view of the orange also activates the concept 'fruit', causing a shift between the two instances, and a stronger activation of the orange with respect to the other objects in the global landscape, whence the selection of the orange by the subject.

While the retrospection process is oriented toward the past, the prospection process makes use of the global landscape to select procedures extending on the long-term future. This landscape gives a space where to 'virtually' select a procedure, evaluate its probable results as they have been recorded from former similar situations; and the process can be iterated during the specious present; it allows selecting a sequence of procedures to be successively realized. For instance, we can plan series of actions in advance by anticipating their results; however the anticipation relies on our former experiences and there is always the risk of fracture because of a non-anticipated change in the context.

A long term procedure can also be selected under the form of a sequence of procedures to be alternatively commanded by several coregulators, each one depending on the anticipated result of the preceding one (as it is retained in the activated domain). Or procedures can be imposed to lower coregulators who command them while intentional coregulators perform another one; for instance, speaking while driving a car. In this case, the higher procedure can be interrupted by a fracture caused by the lower coregulators (a road obstacle forces the attention of the driver). We speak of "consciousness spikes".

7.6 Consciousness and thought

Thought appears as series of mental images activated through a succession of intertwined retrospection and prospection processes; they have another dimension than a simple film, being colored by the temporal dimension of the self. Indeed they rely (through the global landscape) on the archetypal core which conjugates in the specious present a sketch of the past, keeping trace of the successive consciousness spikes, attributing a kind of instantaneous semantic which helps as referent if the intentional coregulators select to initiate a retrospection process. Let us give an illustrative example in which the several operations are described both in usual language, and in terms of the global landscape.

7.6.1 The situation

Collect of information in the landscapes of several lower coregulators while an intentional coregulator CR pursues a specific procedure:

The sight errs. The 'eye' records several images of objects, of color, of light, of shapes. Simultaneously the mind performs another task, for instance looks for the advent of something of interest (a prey or a predator for an animal, an object of curiosity or study for a man…).

7.6.2 Fracture

New stimuli at time *t* cause a fracture for lower coregulators and activate a pragmatic semantic classification, while CR pursues its current procedure.

A spot in the sky appears as a vague form, not very definite. It can be semantically associated with many things:

- undefined without interest (indistinct spot, fly or other insect,...),
- not very distinct nor interesting (very far small bird, flying leaf, airplane),
- more determined and interesting, but still imprecise (distant bird).

In each of these three cases, a pragmatic classification slightly activates the concepts of insect, small bird, leaf, more defined bird, plane, but differentially depending on the observer being an animal or a man (plane only for the later).

Some semantic circuits are more activated, depending on the interest of the observer, whether he looks for anything new, or for a special kind of things. In the first case, his attention will converge on the spot. In the second case, only spots similar to the object of his search will arouse his attention. The surge of attention at the time *t* produces a consciousness spike, which allows for a synchronous oscillation between the circuits pre-activated by the procedure of CR and the circuits which have stored in the memory the several attributes of the spot. The flight of a plane is far off and linear, with a deep sound, the flight of a bird of prey is curved, slow and easy, alternately coming and going away with small cries, the flight of a small bird will be sinuous, hopping, with specific sounds, the flight of an insect is uncertain, near the grass, and so on. The

phenomena may be a fugitive, brief awareness, or prolonged, leading to a fracture.

7.6.3 The response

A fracture for CR extends the *t*-activated domain and leads to the formation of the global landscape; an iterative retrospection-prospection process is started for identifying the spot by recovering its different attributes.

In the case of a fracture, a virtual momentarily autonomous internal world is formed, (the global landscape) which takes hold of all the capacity of attention, memorization and observation of the observer; the retrospection permits multiple comparisons:

- between the spot at successive instants t, $t+1$, $t+2$, … $t+f$,
- between the spot at successive instants and the circuits pre-activated by the search of the observer,
- between the results of the above comparisons and the pre-activated circuits.

It constitutes an iterated reflex loop of the type: "classification, comparison, perceptive memories of the stimuli, their semantic extensions and corresponding procedures, classification, comparison….", each newly recalled attribute leading to new possible causes, and thence to more adequate procedures to respond. For instance the man thinks:

"It is a bird, similar to a bird of prey, even if it is small, it might be a small bird with the flight of a bird of prey, some swifts have a light resemblance with a bird of prey, but his flight is too slow for a swift; it is a bird of prey like the one I saw some days ago above the pasture; I will photograph it"

The procedure for identifying the spot has first activated the general concept of bird, then its attributes are refined: small, with the flight of a bird of prey, whence two possibilities: swift or bird of prey. A new retrospection permits to retrieve some data: the flight is slow, which causes a fracture in the virtual landscape corresponding to the choice 'swift'; thus this choice is excluded. It remains to confirm the choice 'bird of prey' by searching for instances of this concept, thus reactivating a recently activated instance (a kind of priming effect), whence the final identification of a bird of prey and the next procedure: to photograph it.

8 Discussion

We propose MENS as a theory of mind, in which an algebra of mental objects emerges from the functioning of the neural system. It accounts for the development of a hierarchy of mental objects of increasing complexity by an iterative process, based on the two main operations that man can perform: binding mental objects into a more complex one, and classifying mental objects with the formation of concepts. The mental objects and processes are modeled by category-neurons, constructed from the neuronal level up, through a sequence

of complexification processes. The construction is done in the frame of category theory, which gives a rigorous description of the binding and classifying processes (formation of a colimit or of a projective limit). Philosophically, MENS amounts to an emergentist reductionism (Bunge, 1979).

The cat-neurons are iteratively constructed: a cat-neuron of a given level is constructed as the binding of a pattern P of cat-neurons of strictly lower levels, and it takes its own identity as an independent component of MENS. It may have or later acquire other decompositions than P, possibly non-interconnected with P. The cat-neuron has several ramifications obtained by descending the levels down to the neuron level, and its activation consists in the unfolding of one of them, corresponding to the activation of a synchronous hyper-assembly (or assembly of assemblies... of assemblies) of neurons.

8.1 The brain-mind problem

What is the correlation between a mental state (modeled by a cat-neuron) and a brain state?

For a cat-neuron of level 1, the correlation is given by the fact that the cat-neuron binds an assembly of neurons, and conversely is activated by the synchronous activation of this assembly; however even at this level the correlation is non-univocal, since the cat-neuron may bind several non-interconnected assemblies of neurons. This multiplicity (or "degeneracy") is a consequence of the degeneracy of the neuronal coding (Edelman, 1989); it is at the root of the emergence of complex links and, thanks to them, of the emergence of cat-neurons of increasing complexity order.

A mental object modeled by such a cat-neuron "supervenes" on physical brain processes via the stepwise construction of a ramification from the neuron level up; later it will cause physical brain states through the unfolding of this ramification down to the neuron level, leading to a synchronous hyper-assembly of neurons. However, as we have explained in Section 4.4, this unfolding is intricate, necessitating a stepwise construction accounting for emergent properties at each step; and it is multiple (or 'degenerate') since a cat-neuron may have several non-interconnected ramifications. As Kim (1998) has explained, this "multiple realizability" (in his terms) makes mental causation possible while preserving the physical closure of the world.

8.2 Development of higher mental processes

The dynamics of MENS is modulated by the cooperation/competition between a net of internal regulation organs, the coregulators. Each coregulator forms its own landscape where it selects a procedure; the objectives of the various procedures participate in the interplay among procedures, an equilibration process leading to the operative procedure whose objectives (formation of the binding of some patterns and of the classifier of others, possible elimination of some cat-neurons) will be carried out via a (mixed) complexification process. In the interplay the multiplicity of ramifications of a cat-neuron gives much latitude to try to make coherent the procedures of the

various coregulators. It also allows the development of a memory whose records are not rigid, but flexible enough to adapt to progressive changes.

Higher animals have a supplementary capacity: classifying their records, and formalizing such a class by a concept. It leads to the development of a semantic memory. A concept has several instances and can be activated by anyone of them, with a possible shift between concepts. They develop the archetypal core, a personal memory integrating the persistent experiences of any nature; as it merges the past and the present in a dynamic way, it is at the basis of the self. This core has an internal organization which allows for its self-activation via loops of strong and fast links, namely the links in fans. Its activation spreads to the cat-neurons in its halo, and possibly to their ramifications, forming a large 'activated domain' which persists for a long time. It allows the formation of a global landscape by a net of higher coregulators, the intentional coregulators in the halo of the archetypal core; these coregulators receive feedbacks from lower coregulators evaluating the homeostatic drives and hedonic states; and they cooperate in the global landscape.

Conscious processes rely on the global landscape which is formed following an arousing event which activates the archetypal core and extends the activated domain. We have characterized consciousness by two more or less intermingled temporal processes: the retrospection to retrieve the possible cause of the arousing event, and the prospection for selecting long term procedures. The global landscape reflects our conscious experiences. In the global landscape an object is not apprehended as such, but by the intermediary of an activated perspective; it gives an internal perception of the object, different from that an external observer would have. Could this difference be at the origin of the qualia, thus giving an approach to the "hard problem" (Chalmers, 1996)?

The great stability of the global landscape and its very progressive change over time, with overlapping successive global landscapes, can explain the development of self-consciousness: the occurrence of fractures reveals the existence of constraints and, by opposition, leads to the differentiation of the self. For man, language allows developing a more elaborate thought, allowing for extended communication with others at the basis of education, higher learning and culture.

Animals with a nervous system are able to develop at least a primary consciousness (Edelman, 1989). Consciousness extends for higher animals; they can even acquire self-consciousness and, we suppose, have some kind of thought. The usefulness of the temporal dimension of consciousness for the well-being of the animal may explain the development of consciousness through natural selection.

8.3 Possible developments and generalizations.

MENS proposes essentially a *qualitative* model for a theory of mind, even if the energetic and temporal constraints (via the strengths and propagation delays of the links) play an important role in the development of the memory (via

Hebb rule), the temporal constraints of the coregulators, the selection of procedures and the conscious processes. It would be interesting to make it more 'computable', in particular:

- to find some general rules for the selection process (perhaps using cohomological operations, as suggested by R. Guitart, 2009);
- to develop simulations; this is presently tried by Monteiro *et al*. 2009, using the model of Izhikevich *et al.* (2004) for neural systems.

Several generalizations are possible. In MENS, the binding operation (and its opposite, classifying), is essential since it is iteratively applied from the neuron level up to construct cat-neurons modeling mental objects of increasing complexity order; we have attributed their emergence to the multiple realizability of the binding: the same cat-neuron is the binding of several non-interconnected patterns of strictly lower level cat-neurons. The binding operation has been modeled by the categorical colimit operation which has the advantage to allow for an explicit description, via the (mixed) complexification process, of the 'good' links between cat-neurons, making possible the iteration of the process. In some cases it could be interesting to somewhat 'refine' it along one of the following ways:

- The categories could be equipped with a supplementary structure, for instance a topology accounting for the topography of the brain, and making rigorous the geometric metaphor of the archetypal core as gluing together Möbius bands (Section 7.2). The complexification process extends to categories equipped with a compatible enough structure (Ehresmann, 1967), so that our model could easily be translated if we replace the categories by, for instance, topological categories or multiple categories. This last case has been suggested by Brown (2003) (cf. also Changeux and Connes, 1989).
- Colimits could be replaced by "local colimits" (Ehresmann, 2002), or more generally by Baas "hyperstructures" (Baas, 1997; Baas, Ehresmann and Vanbremeersch, 2004), or in the multiple categories case above, by lax colimits. However, to be able to iterate the process, it would be necessary to find precise constructions generalizing the complexification process, and this raises difficult problems.

Appendix: Mathematical definitions

A.1. Categories

For a general theory of categories we refer to Mac Lane's 1971 book. Here we just recall the definitions used in this article.

A (multi-)graph consists of a set of objects (its vertices), and a set of oriented edges between them, represented by arrows f: N → N'. There can exist several parallel edges from N to N'.

A category K is defined as a graph equipped with an internal (partial) composition law associating to the pair of 2 consecutive arrows f: N → N' and g: N'

→ N", a 'composite' arrow fg: N → N", this composition being associative; moreover each object N has an 'identity' id_N: N → N. The arrows are called morphisms or, more simply, links.

A (partial) functor from K to K' is a homomorphism of graphs from (a sub-category of) K to K' which respects the composition and the identities.

If K is a category and K' a sub-category, a *reflection* of an object N of K in K' is an object N' of K' with a morphism d: N → N' such that any other morphism f from N to an object in K' factors in a unique way as $f = df'$ with f' in K'.

A.2. Evolutive Systems (Ehresmann and Vanbremeersch, 1987)

An Evolutive System K consists of the following items:

- a timescale T (finite or infinite subset of the real numbers) modeling its lifetime;
- for each t in T, a category K_t representing the configuration of the system at t;
- for each $t < t'$ in T, a partial functor $K_{t,t'}$ from K_t to $K_{t'}$, called *transition*, which represents the change of configuration from t to t'; we suppose that, for $t < t' < t''$ in T, the transition $K_{t,t''}$ is the composite of $K_{t,t'}$ and $K_{t,'t''}$.

A *component* N of **K** is defined as a maximal family (N_t), indexed by an interval T_N of T, where N_t is an object of K_t and $N_{t'}$ is the image of N_t by the transition from t to t'; the links between components are defined similarly.

For each interval U of T, the components N of **K** such that U is contained in T_N and their links form a category $\mathbf{K}_U$. These categories on the different U form a sheaf of categories on T. When we speak of the colimit of a pattern of components, it is computed in one of these categories.

An evolutive sub-system **K'** of **K** is an evolutive system whose timescale T' is a sub-set of T, its configuration categories $K'_{t'}$ being sub-categories of $K_{t'}$ and its transitions restrictions of those of **K.**

A.3. Colimits (or binding)

Let K be a category. A *pattern* (often called a *diagram*) of objects P in the category consists of a family $(N_i)_{i\varepsilon I}$ of objects N_i and some distinguished links x: $N_i \rightarrow N_j$ between them (thus defining a homomorphism of a graph G to K, the set of vertices of G being I). A *collective link* (or cone) from P to an object N' is a family $(f_i: N_i \rightarrow N')_{i\varepsilon I}$ of links correlated by the distinguished links of P, *i.e.*, for each x: $N_i \rightarrow N_j$ in P, we have $xf_j = f_i$.

A pattern P admits the object cP as a *colimit* (or inductive limit, Kan, 1958) if there is a collective link (c_i) from P to cP such that each collective link (f_i) from P to any N' *binds* into a unique link f from cP to N' satisfying the equations $f_i = c_i f$ for each i. In this case, we also say that P admits cP as its *binding*, and that P

is a *decomposition* of cP. A pattern may have no colimit; if it exists, the colimit is unique (up to an isomorphism).

Two patterns are homologous if there is a 1-1 correspondence between their collective links to any object N'. In this case, either they both have the same colimit, or none of them has a colimit.

A.4. Simple and complex links

Let P and P' be two patterns in the category K. A cluster from P to P' is a maximal family π of morphisms from objects of P to objects of P' such that:

- For each object P_i of P there is at least one g in π from P_i to an object of P', and if there are several such morphisms, they are correlated by a zig-zag of distinguished links of P'.
- The composite of a morphism in π with a distinguished link of P (on the left) or of P' (on the right) is also in the cluster.

If π is a cluster from P to P' and if P and P' admit colimits cP and cP' respectively, there is a unique link cπ from cP to cP' binding the collective link (gc'_j) from P to cP', where g varies in π and (c'_j) is the collective link from P' to cP'; it is called the (P, P')-*simple link* binding π. A (P, P')-simple link might not be (Q, Q')-simple for other decompositions Q of cP and Q' of cP'.

Two decompositions P and Q of the object N are *interconnected* if the identity of N is a (P, Q)-simple link or a (Q, P)-simple link. Otherwise, we say that P and Q are *non-interconnected* and the passage from P to Q is called a *complex switch.*

Dually P admits a *projective limit* (or *classifier*) C in K if C is a colimit of the pattern opposite to P in the category K^{op} opposite to K (obtained by changing the direction of its morphisms). A *pro-cluster* from P to P' is a cluster from P' to P in K^{op}; if P and P' have classifiers, the pro-cluster is 'classified' by a (P, P')-simple link from C to C'; complex links are obtained by composing simple links classifying non-adjacent pro-clusters.

A.5. Hierarchical categories

A category K is hierarchical if its objects are divided up into in a finite number of complexity levels so that an object N of level *n+1* is the colimit of at least one pattern of objects of strictly lower levels. If N admits several such non-interconnected decompositions, it is said to be *n*-multifold. If K has multifold objects, we say that K satisfies the multiplicity principle. In this case there are complex links obtained by composing simple links binding non-adjacent clusters, with complex switches between the different decompositions of the intermediate multifold objects. The complexity order of an object N of level *n+1* is the smallest k such that N is the colimit of a pattern of objects of levels strictly lower than *k*; we have $k \leq n$ and we have given conditions for having $k < n$. (Cf. Ehresmann and Vanbremeersch, 1987.)

An evolutive system is hierarchical if the configuration categories are hierarchical and the transitions respect the levels.

A.6. The complexification process

A pattern P of objects in a category K may have no colimit. In this case the category K can be extended into a larger category in which P acquires a colimit. This is the basis of the complexification process.

Given a category K, a procedure (or 'option' in Ehresmann and Vanbremeersch, 2007) is a list of objectives for modifying K by means of some of the following actions:

- Binding a set B of patterns P of K: if P has a colimit in K, this colimit should be preserved, and if P has no colimit, a new object should be added to become its colimit.
- Eliminating a set S of objects, possibly thus dissociating some bindings.
- Adding a set E of external objects.

The complexification of K with respect to this procedure (Ehresmann and Vanbremeersch, 1987) is a category K' which is a universal solution of the problem of realizing the objectives of the procedure Pr. For an explicit construction of K', we refer to our 2007 book (Chapter 4). The objects of K' are the objects of K except those of S, the objects of E, and, for each P in B which has no colimit in K a new object cP which becomes its colimit in K'; if two patterns in B are homologous, the same cP is chosen for binding them. The links between two added objects cP and cP' are the (P, P')-simple links binding clusters from P to P', and complex links composing simple links. While a (P, P')-simple link only depends on the 'local' interactions between P and P', the complex links depend on the whole structure of the initial category, and they represent properties 'emerging' in its complexification.

A *mixed procedure* is a procedure whose objectives also require classifying a set of patterns Q by adding a new object which becomes the projective limit, or classifier, of Q. The corresponding *mixed complexification* is also described in Ehresmann and Vanbremeersch (2007, Chapter 4).

An important theorem on complexifications (Ehresmann and Vanbremeersch, 1996, 2002) asserts that: *If a category* K *satisfies the multiplicity principle, then so does any of its complexifications*. In this case, we have proved that an iteration of the complexification process leads to the emergence of a whole hierarchy of objects with strictly increasing complexity orders.

A.7. Grothendieck topologies

A *sieve* on an object N of the category K is a family of morphisms f_i: $N_i \rightarrow N$, closed by composition with morphisms to the N_i's. A *Grothendieck topology* (Grothendieck and Verdier, 1972) on K associates to each object N of K a class of sieves on N (called its *covering sieves*), so that N acts as an open set of a

topological space and the covering sieves on N as the open coverings of N. With this topology, K becomes a *site*.

A Grothendieck topology on K may be generated by the data, for each object N of K, of a set F(N) of morphisms f_i: $N_i \rightarrow N$, called a *covering family* of N. It is obtained by taking first as covering sieves on N the sieve F*(N) generated by F(N) and the sieve I*(N) generated by the identity of N (its elements are all the morphisms with codomain N); and then constructing all the sieves $g^*(R)$ where g is a morphism from N' to N, R a covering sieve on N', and $g^*(R)$ has for elements the composites of the elements of R with g.

References

Baars, B. J., 1997, In the theatre of consciousness: The workspace of the mind, Oxford University Press. Oxford.

Baas, N.A., Ehresmann, A.C. and Vanbremeersch, J.-P., 2004, Hyperstructures and Memory Evolutive Systems, Int. J. of Gen. Syst. 33 (5), 553-568.

Barlow, H.B., 1972, Single units and sensation: A neuron doctrine for perceptual psychology, Perception 1, 371-394

Borsuk, K., 1975, Theory of shape, Monografie Mat. 59, Warsow.

Brown, R.and Porter, T., 2003, Category theory and higher dimensional algebra: potential descriptive tools in Neuroscience, in Proc. Intern. Conf. on theoretical Neurobiology (Ed. Singh), NBRC, New Delhi, 62-79.

Bunge, M., 1979, Treatise on Basic Philosophy, Vol. 4, Reidel, Dordrecht.

Chalmers, D., 1996, The Conscious Mind, Oxford University Press, Oxford.

Changeux, J.-P., 1983, L'homme neuronal, Fayard, Paris.

Changeux, J.-P. and Connes, A. 1989, Matière à pensée, Fayard, Paris.

Collins, J., Imhoff, T. and Grigg, P., 1996, Noise-enhanced tactile sensation, Nature 383, 770.

Crick, F., 1994, The Astonishing Hypothesis, Macmillan Publishing Company, New York.

Damasio, A., 1999, The Feeling of What Happens: Body and Emotion in the Making of Consciousness, Harcourt Brace, New York.

Dehaene, S., Kerszberg, M. and Changeux, J.-P., 1998, A neuronal model of a global workspace in effortful cognitive tasks, Proc. Natl. Acad. Sc. USA 95, 14529-14534.

Dehaene, S., Sergent, C. and Changeux, J.-P., 2003, A neuronal network model linking subjective reports and objective physiological data during conscious perception, Proc. Natl. Acad. Sc. USA 100, 8520-8525.

Dennett, D., 1990, La stratégie de l'interprète, NRF Gallimard, Paris.

Edelman, G.M., 1989, The remembered Present, Basic Books, New York.

Edelman, G.M. and Gally, J.A., 2001, Degeneracy and complexity in biological systems, Proc. Natl. Acad. Sci. USA 98, 13763-13768.

Ehresmann, A.C., 2002, Localization of universal problems. Local colimits, Applied Categorical Structures 10, 157-172.

Ehresmann, A.C. and Vanbremeersch J.-P., 1987, Hierarchical Evolutive Systems: A mathematical model for complex systems, Bull. of Math. Bio. 49 (1), 13-50.

Ehresmann, A.C. and Vanbremeersch J.-P., 1992, Semantics and Communication for Memory Evolutive Systems, in Proc. 6th Intern. Conf. on Systems Research (Ed. Lasker), University of Windsor.

Ehresmann, A.C. and Vanbremeersch J.-P., 1996, Multiplicity Principle and emergence in MES, Journal of Systems Analysis, Modelling, Simulation 26, 81-117.

Ehresmann, A.C. and Vanbremeersch J.-P., 1999, Online URL:

http://pagesperso-orange.fr/vbm-ehr

Ehresmann, A.C. and Vanbremeersch J.-P., 2002, Emergence Processes up to Consciousness Using the Multiplicity Principle and Quantum Physics, A.I.P. Conference Proceedings 627 (CASYS, 2001; Ed. D. Dubois), 221-233.

Ehresmann, A.C. and Vanbremeersch J.-P., 2007, Memory Evolutive Systems: Hierarchy, Emergence, Cognition, Elsevier, Amsterdam.

Ehresmann, C., 1967, Sur l'existence de structures libres et de foncteurs adjoints, Cahiers Top. et Géom. Diff. IX, 133-180; reprinted in Charles Ehresmann: Œuvres complètes et commentées, Part IV (Ed. A.C. Ehresmann), 1983, Amiens, 117-264.

Eilenberg, S. and Mac Lane, S., 1945, General theory of natural equivalences, Trans. Am. Math. Soc. 58, 231-294.

Engert, F. and Bonhoeffer, T., 1997, Synapse specificity of long-term potentiation breaks down at short distances, Nature 388, 279-282.

Fisahn, A., Pike, F.G., Buhl, E.H. and Paulsen, O., 1998, Cholinergic induction of network oscillations at 40 Hz in the hippocampus in vitro, Nature 394, 186-189.

Fodor, J.A., 1983, The modularity of Mind, MIT Press, Cambridge.

Frey, U. and Morris R., 1997, Synaptic tagging and long-term potentiation, Nature 385, 533-536.

Grothendieck A. and Verdier J.I., 1972, Théorie des topos, SGA 4. Springer Lecture Notes in Math. 269-270.

Guitart, R., 2009, How to compute Sense?, Axiomathes (to appear), Springer.

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C.J., Van J. Wedeen and Sporns, O., 2008, Mapping the Structural Core of Human Cerebral Cortex, PLoS Biology 6, Issue 7, 1479-1493. Online: www.plosbiology.org

Hebb, D. O., 1949, The organization of behaviour, Wiley, New York.

Hopfield, J. J., 1982, Neural networks and physical systems, Proc. Natl. Acad. Sci. USA 79, 2554-2558.

Hubel, D.H. and Wiesel, T.N., 1962, Receptive fields..., J. Physio. 160 (1), 106-154.

Izhikevich, E.M., Gally, J.A. and Edelman, G.J., 2004, Spike-timing Dynamics of Neuronal Groups, Cerebral Cortex 14, N 8, Oxford University Press.

James, W., 1890, Principles of Psychology, H. Holt and C°, New York.

Kan, D. M., 1958, Adjoint Functors, Trans. Am. Math. Soc. 89, 294-329.

Kim, J., 1998, Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation, M.I.T. Press. Cambridge, Massachusetts.

Laborit, H., 1983, La Colornbe Assassinée, Grasset, Paris.

Levin, J. and Miller, J., 1996, Broadband neural encoding in the cricket cercal sensory system enhanced by stochastic resonance, Nature 380, 165-168.

Mac Lane, S., 1971, Categories for the working mathematician, Springer.

Malsburg (von der), C., 1995, Binding in models of perception and brain function, Current Opinions in Neurobiology 5, 520-526.

Malsburg C. (von der) and Bienenstock E., 1986, Statistical coding and short-term synaptic plasticity, in Disordered systems and biological organization, NATO ASI Series 20, Springer, 247-272.

Merleau-Ponty, M., 1945, Phénoménologie de la perception, Ed. Gallimard, Paris.

Miltner, W., Braun, C., Arnold, M., Witte, H. and Taub, E., 1999, Nature 397, 434-436.

Minsky, M., 1986, The society of mind, Simon and Schuster, New York.

Monteiro, J., Ribeiro, J.H., Kogler, J.E. and Netto, M.L, 2009, On building a Memory Evolutive System for application to learning and cognition modeling, in Brain Inspired Cognitive Systems (Ed. Cutsuridis, V., Hussain, A. , Barros, A.K. and Aleksander, I., Springer.

Morin, E., 1977, La Méthode, Editions Seuil, Paris.

O'Keefe, J. and Dostrovsky, J., 1971, The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat, Brain Research 34, 171-175.

Piaget, J., 1940, Le développement mental de l'enfant, in Six études de psychologie, Paris.

Pierce, C.S., 1903, Abduction and induction, in Philosophical writings of Pierce (Buchler, J., Ed.), Dover Publications, New York, 150-156.

Rodriguez, E., George N., Lachaux J.-P., Martinerie, J. Renault, B. and Varela F., 1999, Perception's shadow: long-distance synchronization of humain brain activity, Nature 397, 430-433.

Rosch, E. 1973, Natural categories, Cognitive psychology 4, 328-350.

Russell, B., 1971, La méthode scientifique en Philosophie, Payot, Paris.

Ryan, A. 2007, Emergence is coupled to scope, not to level, Complexity 13, Issue 2, 67-77.

Stryker, M.P., 1989, Is grand-mother an oscillation? Nature 339-351.

Usher, M. and Donnelly, N., 1998, Visual synchrony affects binding and segmentation in perception, Nature 394, 179-182.

Varela, F.J. 1989, Autonomie et connaissance, Editions du Seuil, Paris.

Wallace, R., 2004, Consciousness, cognition and the hierarchy of contexts: extending the global neuronal workspace, Online URl: http://cogprints.org/3677

Wehr, M. and Laurent, G., 1996, Odour encoding by temporal sequences of firing in oscillating neural assemblies, Nature 384, 162-166.

Wiesenfeld, K. and Moss, F., 1995, Stochastic resonance and the benefits of noise..., Nature 373, 33-36.

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

Zhang, Li, Tao H., Holt C., Harris, W. and Poo M., 1998, A critical window for cooperation and competition among developing retinotectal synapses, Nature 395, 37-43.

The Unbearable Heaviness of Being in Phenomenologist AI

Jaime Gómez and Ricardo Sanz

Autonomous Systems Laboratory, Universidad Politécnica de Madrid

Abstract

The aim of this paper is to pin down the misuse of Heidegger's philosophal insights within the discipline of artificial intelligence (AI) and robotics. In this paper we argue that a central thesis of phenomenology, in Husserl's words, "putting the world between brackets", has led to a positioning in embodied AI that deeply neglects fundamental representational aspects that are totally necessary for the purpose of building a theory of cognition. The unification of representational and being-in-the-world aspects, are necesary for the explanation and realization of complex consciousness phenomenon in a cognizer, both animal and mechanic. The emphasis on the self (post-cognitivists), on the being (phenomenologists), as well as the Being by Heidegger's followers, has contributed interesting insights concerning the puzzle of cognition and consciousness. However, has neglected the necessity and even denied the possibility to provide a scientific theory of cognition.

On the other hand, the phenomenologist's separation of the world into two different ones, the scientific and objective world, and that of our common and lived experience is untenable. The claim that any scientific-theoretical world must find its foundation in the so called live world is ill-founded. In this paper we will propose the basis of a theoretical framework where only one world —with entities and processes— exists and can be known to a certain degree by the cognitive system. This calls for a unified vision of both ontology and epistemology.

1 The Phenomenological Bias

1.1 The object/subject problem revisited

Phenomenology arose out the necessity to surmount the difficulties posed by the dichotomic vision established in Idealist and Materialist philosophies. Apparently, at the core of this dichotomic philosophical approach lurks a paradox pointed out by Husserl: "How is it possible that myself, as a transcendental ego, builds-up the world, being at the same time a human ego inside the world ?". But, where is the paradox? We can't really see it.

The agent is in the world and builds a world of its own, but there is no such paradox. Assuming that for a finite agent it is impossible to give a causal explanation for every fact in the world this is not, in any case, due to a world's opacity to the cognitive capabilites, but to the fact that we are limited cognitive agents inserted in the same reality we want to know. We are part of the world and situated in it. Therefore we can perceive the world only partially.

The world we build and the world we live in are not identical, but closely bound by what Rosen's called the modeling relation (1). This closeness being of evolutionary survival value.

The phenomenologist's approach is obviously excesively biased towards the experiencing agent. This bias has been inherited by robocists and other AI scholars as a reaction to the perceived failures of GOFAI (6). It has been used as a starting point for further development of common-sense centered theories and other naïve conceptions of peception and cognition.

In Husserl's philosophy (3) the object appears as essentially determined by the structure of thinking itself. The world is placed between brackets and the focus is put on the Cogito in the Cartesian's Cogito ergo sum, and objectivity is not longer on the consciousness side.

Husserl pretends to arrive at the essence of things from the experiencer[1] point of view. To that end, phenomenology proposes a method called transcendental reduction (epoché) to get to the essence of the objects, hence bracketing the assumption of the existence of an external world. So, access to the real being of the things may only be achieved by the transcendental reduction process grounded in the experiencing self.

The direct economic approach from engineering is necessarily closer to a Humean theory of the self. Hume rejects the object-subject dichotomy, eliminating the self as a knower. Hume's claim, unlike other empiricists like Locke or Berkeley, is in a sense more ontological than epistemological, because he does not have to posit the object of the knower, instead he just describes and analyzes a group of entities called perceptions. The self would be just that succession of related ideas and impressions (perceptions in Hume's words) of which the agent has an intimate memory2. This interpretation of the self, as a connected succession of perceptions, will be taken afterwards by other authors (e.g. William James).

1.2 Two kinds of beings for two kinds of worlds

In Husserl's philosophy, a distinction between the world and the everyday world (*Lebenswelt*) is established. This is a logical consequence of his tenets: if the cognitive agent is who *rises* the world depending on the agent's attitude, the world could be configurated in a different manner.

Here, there is an implicit criticism to the scientific method. In Husserl's view, the scientific method would be just one attitude, valuable to understand the world explained by physics, but not the correct one to unveil the everyday world (*Lebenswelt*). This claim, that is, the inescapable distinction between the external reality and the reality perceived by the cognitive agent,

[1] Then phenomenology becomes the discipline that investigates the essential nature of the world.

[2] If we eliminate, as Humes does, the epistemologic concept of knower, we do it too for the antinomy between unknown reality and known reality. Hume erases the trascendence in the cognitive agent, transcendence that by other means will be emphasized in Phenomenology, with the harmful consequences that will be shown next.

animal or robot, has been repeated as a totem by continental philosophers and some AI and roboticist scholars of postmodernist vein.

We fully agree with the analysis that there are different attitudes and that we perceive things, categorize items or infer new sentences, in part motivated and shaped by our current attitude. But the distinction of worlds as a consequence of the attitude, vanishes when we define the concepts in a rigorous manner. Attitudes are structured frames or theories that can be eventually formalized, and might not be confounded with intentionality, which is, as Brentano pointed out, the focus of consciousness.

Intentionality and attention are radically different things, the former is the power of minds to be about or to stand for things, and guiding the behavior, or said à la Dennett "an active engagement with the real world"; and the last is a more complex understanding of objects and process that frames the intentionality of the cognitive agent.

The question about the existence of two worlds —or two thousand worlds— appears promptly. This degeneration3 in the use of the word "worlds´´, is in part motivated by the mistake which considers thought and word as the same thing. Obviously language is an important high order cognitive ability, whose fundamental function is to share mental states, that is, as a means to vehiculize, to make one's thoughts public. However, inferring from that that there is an ontologic equivalence between mental concepts and the words that, denotate them, in order to make accessible to the linguistic level, is totally wrong4.

The distinction between the world explained by the physics and the everyday world (*Lebenswelt*) does not correspond to any scientific reason but is a sign of obscurantist or at best, lazy thinking. The construction of the everyday world, different to the world of the physics, is not justified. There is only one world, whose entities and process are known to a certain degree, both to scientists and cognitive agents. Our duty as scientists is to explain this world, its phenomenon and entities, by means of laws and causal theories either deterministic or probabilistic or a mixture of both.

2 Heideggerian AI. The being in the world

Husserl's program is indeed deeply epistemologic, but this is not the case for Heidegger, so keen to many post-modern roboticists. For Heidegger, Ontology is possible only as a kind of Phenomenology. We can obtain the structures of the being only by means of the way they manifest themselves as phenomenon. Heidegger's is primordialy concerned with the pre-conceptual under-

[3] Gerald Edelman (12) uses the same term -degenerate- to explain consciousness, "neural groups whose degenerate responses can, by selection accommodate the open-ended richness of environmental input, history, and individual variation".

[4] The falsity of the ontological equivalence between thinking and speaking is easily demostrated: not all the concepts are linguistic concepts. This confussion was exemplary described by *the first* Wittgenstein: "the limits of my language mean the limits of my world´´.

standing of Being (*Dassein*) like a protoconsciousness, already socialized. But, explaining conciousness in terms of *Dasein ignotum per ignotius*.

Heideggerian philosophy rejects the apparent Cartesian isolation of the epistemological subject. There is never an isolated ”I” given without the world, rather any ontology is only conceived as the ontology of a subject. Being-in-theworld is the mode of being a cognitive agent immersed, not just in interactions, but in couplings with surrounding entities.

This metaphysics differentiates two kinds of beings, the readiness-to-hand and unreadiness-to-hand, the former is the being when we are using it and the second, when we contemplate it[5].

This analysis is fundamentally based in the perceptual and motor interaction with equipments. The habitual example of the hammer, which has two different modes of being -a hammer hammering a nail, or a hammer in a drawer. This offers an extremely basic categorization (maybe that is the reason why it has some followers in AI) that is also extremely limited, because it is focused only on tools. It looks like Heidegger's phobia of technology[6] gives his system a kind of hand made or medieval touch in his philosophy.

What this approach seems to provide, and to our understanding the central reason for its luring capability, is that it seems to offer an explanation for the apparent failure of GOFAI and a potential alternative to explore in the implementation of cognitive architectures. Agents, Heidegger's followers say, do not need representation, but rather continuous sensory-motor immersion in its reality. The aphorism “the map is not the territory”(13) became the motto of the situated robotics movement7. This immersion in the world seemingly offers a solution to the so called frame problem. If the agent uses the world as its own map it is no longer necessary to keep in sync the world and mental representation.

The agent captures reality in the form of patterns (see Figure 2) or in the words of Agre these representations ”designate, not a particular object in the world, but rather a role that an object might play in a certain time-extended pattern of interaction between an agent and its environment´´(14).

[5] Another Heideggerian, J.P. Sartre distinguishes between etre-en-soi and etre-pour-soi

[6] ”When man reveals that which presences, he merely responds to the call of unconcealment even when he contradicts it. Thus when man, investigating, observing, ensnares nature as an area of his own conceiving, he has already been claimed by a way of revealing that challenges him to approach nature as an object of research, until even the object disappears into the objectlessness of standingreserve. Modern technology as an ordering revealing is, then, no merely human doing´´.(10)

[7] Curiously enough, some argue for this approach being non-externalist in the sense of Clark cognitive externalism.

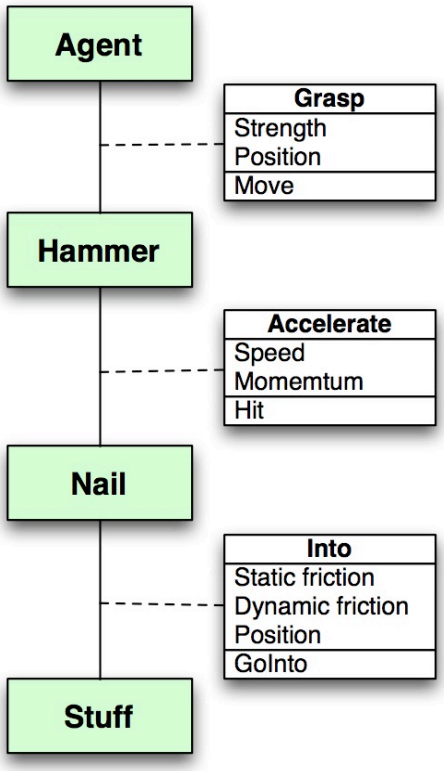

Figure 1: The hammer, the nail and the stuff constitute a pattern.

Of course, this Heideggerian conceptual system for beings is far too simple to give clear responses to other kind of concepts like the abstract or the simulated ones.

The epistemic Husserlian program anticipates the frames theory developed by Minsky with his concept of Noema: a symbolic description of the anticipated features and values of an object, a sort of inner horizon of expectations that permit the structure of incoming data, conforming the context of the object. Heidegger criticizes this enterprise of determining the inner horizon as insufficeint to give an account of the context, because the necessary condition to determine it, is to consider as a whole the cultural practises. Therefore, the relevant characteristics which define the context are always already contextualized in a cultural and historical background.

Paraphrasing Heidegger we can say that "[Agents] are already always in a situation". But H.L. Dreyfus—a Berkeley professor and Heideggerian reference in the AI world— claims, in an opossing line, that a robot, even counting on all the possible knowledge it would get from the outside, would not be in any situation, the robot being a decontextualized entity 8.

[8] Heideggerian AI arises out from the frame problem. However it does not provide any solution to the problem, not even any useful insight; but it is a pernicious influence for AI and robotics. Indeed as Dreyfus points out, Heideggerian and positive theories are a *contradictio in terminis*.

But the Heideggerian analysis of AI is useful in the sense that raises some critical issues concerning the kind of control architecture that a real-world cognitive agent should have —including the representational aspects they would abhor. This analysis does not exclude the possibility to formally describe the situation and hence derive representations for it. Heideggerians opposing representation-based architectures and modular structures go indeed too far in their analyses of the limitations. For example, their case for coupling vs input/output interactions seem to ignore the trivial fact that any interaction —whether input or output— is indeed bidirectional except in degenerate cases, because the labeling input and output is plainly arbitrary and is in the eye of the beholder.

The thesis defended in this paper is pretty far from this anthropomorphist view. We stress again, that the big mistake is in giving to the mental phenomenon a condition of ontologic difference with respect to the external phenomenon, driving the theorist to ascribe ascientific assumptions and intuitionist theories.

One clear example of this is when Heidegger claims that the mental model of a human of the world is the world itself (cf Korzybski before). Were this the case, any two agents navigating the world would be similarly proficient. But it is obvious that humans, unlike robots or cockroaches, have a mental model of the world that is more acute —ideally isomorphic to a certain extent— that is good enough to permit the human race to survive. We can not say the same of the Heideggerian robots like Brooksian insects or of Cog, the failed humanoid. But we can say something about of cockroaches, their maximal survability being the reason for the mystifying power of bioinspiration, and it is that in a direct human-cockroach confrontation for an ecological niche all we know what would happen.

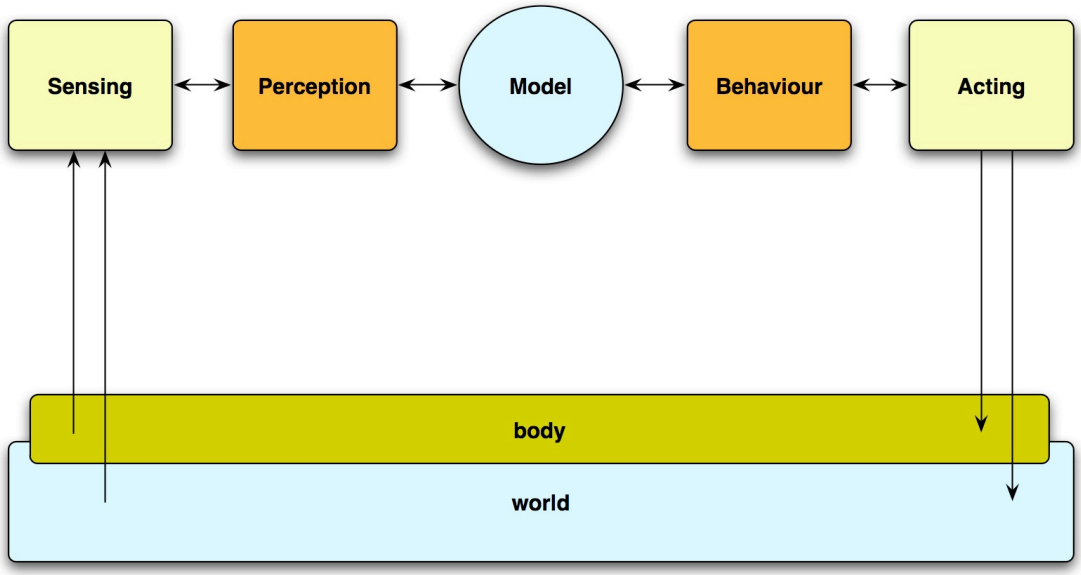

Figure 2: The simple vision of the epistemic —model-based— control loop.

It is a logical absurdity the claim that the mental model of the world of the cognitive agent is the external world; it can be suggestive as a poetic figure, but no scientific model or theory can accept an ontologic falsity as that as a

valid proposition9. We claim that there is not any unsurmountable obstacle in formally defining a context for the everyday action. The focus must be put in the theory, which is operating as *cache memory* when we categorize or define concepts, we call this theory Legality. This is done in the context of the realization of an epistemic control loop, where a model of the surrounding world is used by the agent in the performance of its dwelling (See Figure 2).

3 The Embodied Cognition or the Being with Flesh

All these efforts are very valuable but, from our systemic perspective, all these fleshists —Heidegger, Marleau-Ponty, van Gelder, Lakoff, Dreyfus, etc— put too much flesh in the dish of cognition.

Marleau-Ponty (15) gives a sound account that supersedes the dichotomy subject(knower) --object(knowing), formulating the circularity in the perception-action loop. The animal is moved to action in order to acquire and maintain an optimal perceptual grip on what is significant to him in the world10; in other words, the body evolves in a pathway of permitted states defined by a net of basin of attractors ,which lead the body to move towards an optimal grip.

The introduction by Marleau-Ponty of the body and the perception action loop in his cognitive theory is consistent with the naturalized studies of consciousness, and has set the basis of embodied cognition theories, which are biologically inspired, where the mental phenomena are studied not as personal feelings but as a natural phenomenon. The body (coper) interacts with the environment in such a way as to cope with an environment organized in terms of that organism's need to find its way around.

So for Marleau-Ponty, the body is not just the physical space occupied by the thinking agent, but the necessary instrument to cope best coping with the environment, and to that end, the body moves towards its equilibrium. But once achieved, the coper can not stop there because the environment continues sending solicitations to be interpreted by the coper, in order to get a new best coupling or equilibrium between coper and environment. It is interesting to consider, at this point, the analysis done in neuroscience —and consequently in neuro-inspired robotics— in terms of learning stimlus-response and action-outcome pairs. The question of causality lurks here and is strongly related with Merleau-Ponty's concept of solicitation.

Marleau-Ponty reduces or explains cognition based just on the perceptive process; it looks like the body is the magic key, which explains and obtains all the meanings. [11]

---

[9] Heidegger here is totally coherent because Heidegger himself is ascientific.

[10] This is also the central tennet of W.T. Powers perceptual control theory (16).

[11] Admitting the importance of mirror neurons discovery in motor verbs, we can not construct a
global theory of knowledge just with bodily metaphors, flesh is not enough we need the

Van Gelder considers that the external world is too complex to possibly get a representation of it, and argues that it is cognition that enables the agent to cope successfully with the world."The post-Cartesian agent manages to cope with the world without necessarily representing it. ...the internal operation of a system interacting with an external world can be so subtle and complex as to defy description in representational terms, or in other words, cognition can transcend representation" (17). Obviously this is only true if representations are to be universal and not action-oriented. It is clear that representation complexity can be reduced without much performance sacrifice for concrete tasks. The tradeoff between the complexity of the representation and the competence it offers is resolved in evolutionary economic terms.

Even if van Gelder is using the term cognition in a wider sense as the act of knowing or, as an emergent property of the cognitive agent, representation can not be excluded from cognition, van Gelder eliminates the representational power of the agent in cognition, and puts in its place the notion of coupling. Indeed coupled system performance —e.g. in terms of agent survival— is the result of an isomorphic representation of the world by the agent (more on this later). However, van Gelder suggests that cognition must be untangled from representation except for sophisticated cases involving representation such as breakdowns, problem solving, and abstract thought; but such phenomenon are best understood as emerging from a dynamic substrate, rather than as constituting the basic level of cognitive performance".

But we think that the coupling part in the dynamic information processing, realized by the agent in a dynamic environment, is not the appropriate alternative to the representation of what the Heideggerians call "the everyday world". Van Gelder is missing the point. There is not any justification to separate cognition and representation, both are inherently informational processes or products of such processes; and on the other hand, when he points out that thinking an abstract thought is a phenomenon better understood as emergent, not only he is not not saying anything of any value about such a phenomenon but he is also suggesting a sort of emergentist inexplicablity.

No matter what the emergent properties are, they must occur following laws, as do all the other phenomena happening in the world12. Denial of this is pure obscurantism, an attitude incompatible with the scientific stance.

4 A proposal: Systemic-Explanation

When we observe a Heideggerian robot trying to avoid a non trivial obstacle (see for example (20)) we certainly know that this is not what we see from an animal not much more sophisticated than an insect or what we would expect from the machines of the future. Higher animals do have cognitive capabilities

bones, the skeleton! Maybe too much importance is given to the body(11).

[12] It is quite probable that the so called emergent phenomena is just massive non-linearity, to be explained in the future using thoeries like nonlinear thermodynamics, chaos theory, etc.

that surpass what the *ready-to-hand* and *present-at-hand* ontologies make possible. Deep representations and representation-based behavior engines lie at the core of this capability. For us, something is a representation of another something if it contains/captures some aspect of this second something.

In a sense, the whole issue of anti-representationism seems absurd from our perspective. What a sensor does is re-present in a different value space the value of a certain magnitude. So, from this perspective, if there is a sensor there is a representation. Elephants don't play chess but they necessarily represent the light in the sky, the water that they drink or the sound of their youngsters.

Beyond the concreteness and atomicity of such representations, can you imagine going back home by means of being-in-the-world?. That would take time, too much time indeed for an evolutionarily viable system.

The impression that we get from the Heideggerians is that they see representation in the simplest GOFAI sense of collections of atomic predicates. Obviously this representation is untenable as a substrate for cognitive behavior in a world for the simple reason that these representations can not represent relevant aspects of the world; fundamentally those related with dynamical-structural aspects of the world.

Heideggerians realize this fact and their reaction is the rejection of the representation as such —and its associated sense of separation between agent and world— to embrace a holistic approach: the agent can't be separated from the world and it must be its own representation. What they should reject is not representation but the kind of representation that is not effective for the particular class of world that the agent in interacting with.

Systems theorists describe systems as a collection of things and a relation between them:

$$S = (T, R)$$

In the system we are interested in, the things T included by the agent and its surrounding reality. Heideggerians aptly see that a collection of representations of the states of the things is not enough to capture the dynamics of the agent-world system. But they fail when they revert into strictly focusing just into the relational aspects R. There is no system without the relation and there is no system without the things. Both parts are necessary to understand the dynamics and hence necessary to master to make a living in that context. While centered on social studies, the article of Mario Bunge (8) is extremely clarifying in this aspect.

So, what a perfect cognitive system must do is to perfectly represent the whole system S = (T, R) in its mind in order to maximize its performance. Obviously, perfection in representation is not possible (this is van Gelder's argument) but thank God it is not necessary for a real agent. What an agent actually needs is a sufficiently good representation of S —we call this a model— to get a sufficiently good outcome from its use. Fortunately, simpler models can be qualitatively equivalent to detailed ones in a certain region of their state space. This

is what makes driving cars on roads possible, or the use of computers without being a computer scientist, or what enables cooking without being an histologist, a chemical engineer and perception psychologist at the same time.

Quoting Edward Box we can say "Essentially, all models are wrong, but some are useful". Cognitive agents just exploit useful models. But having a model is not a question of contingency but of necessity. There is no alternative other than internalizing a model to be effective. As Conant and Ashby demonstrate (19) every good regulator must contain a model of the system it is controlling, or, put into the words of cognitive science, the agent must represent the world to dwell in it. This has strong implications: if an agent is successful in a certain world, it is because it is driven by a model of that world.

This does not mean that we can open the agent and read in its mind, the structure of the world —reading the model– because the model can be collapsed with the perceptual or behavioral systems or with both (see figure fig:epistemic). For example the hammering model of Figure 2 can be collapsed with behavioral subsystem so a hammering order will directly map into the motor action of the agent for a concrete hammer, a particular type of nail and a class of *stuff* to nail the nail into. These embodied realizations of the hammering agent are less effective for a different hammer, different nail or a different *stuff* —the things— for for a different grasp or static friction coefficient of —the relations. But a non-embodied, cognitive agent, can appropriately reason in those circumstances.

We may then question what is the adaptive value of embodiment. The answer is clear and well known in engineering: there are tradeoffs that define families of control structures for the available niches; speed vs cost, robustness vs variety, size vs growth, etc. Embodiment is just an economic, effective solution for certain operational niches.

But we shall remember the fact that embodiment sacrifices behavioral flexibility and that in conditions of no restrictions the pure disembodied agent is maximally performant.

We may wonder what the theoretical substrate is, that enables the construction and exploitation of effective model-based representations. The deep insight is that models do have morphic relations with the modeled. This means that entailments in the modeled —e.g. causal entailments in a physical system— can be mapped into logical entailments in the model, and logical entailments in the model can be mapped back to the modeled system. So we can use the model to reason about the modeled —e.g. to drive part of the world to a certain state or to get some qualia for the agent.

This relation between systems (Figure 4) is called the modeling relation by Rosen (1) and to our understanding captures the very nature of cognition: minds can be put into congruence with the world.

5 Conclusions

Trying to overcome Descartes, the Phenomenologists —from Husserl to Varela passing through Heidegger— have proposed something worse than the Cartesianism, the invention of transcendental entities hardly justifiable in the modern scientific paradigm to give substance to an impoverished relational model of the system composed by an agent and its environment. What Husserl calls the Ego, Heidegger calls the *Dassein*; these are more or less suggestive metaphors or rhetorical pictures insufficient in obtaining a scientific representation of the mental state.

We must strive to find the general conditions of possibility for the mental phenomenon. Naturalizing minds up to the level of consciousness is a long term project but scientifically falsable, technically sound and in any case better than the simple option of the phenomenologists.

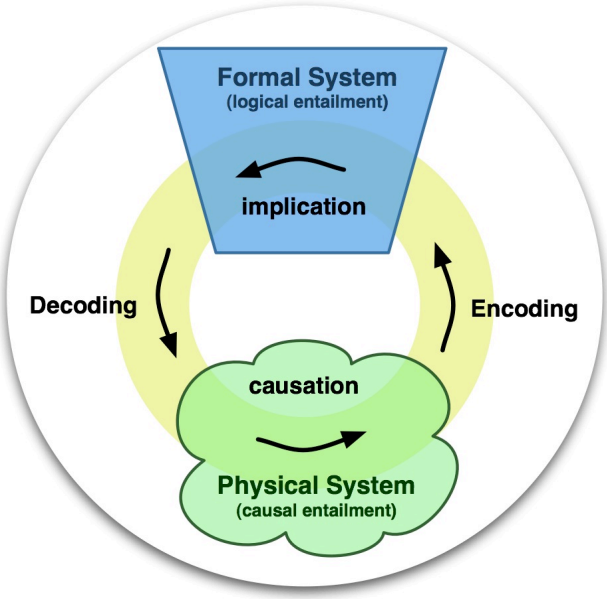

Figure 3: The Rosen's modelling relation captures the basic tenets of cognition as model-based interaction with the world.

Anti-reductionism is usually mislead, and the distinction between ontologic and epistemic reductionism must be known. The first is a reductionism of type "A is B", being A and B predicates (i.e: A neural process is a mental process), and the later is "B is explained in terms of A" (i.e: In a depressive state the concentration of serotonine is low).

The nature is structured in levels, the postulation of the *Dassein* is a consequence of the incapability to appreciate this fact. The everyday world is the same world of the books of physics. Indeed Newtonian mechanics can be written in Einstein equations. It is a question of granularity (norms and theories, the legality in Petitot terms) and not of non- measurability or a-representability. Models do not only have resolution levels but qualitatively hierarchical morphisms. The external world exists independently of the subject and the real processes and entities belonging to the world can be described and explained objectively. On the opposite side of this view are those who claim that there only exist, the appareances perceived by the subject; but even the extreme phenomenist take for granted the reality, independently of

what he is observing, he assumes the reality of what he is observing and also the reality and of himself as an observer of the phenomenon. In conclusion, it is impossible to avoid being a realist and it is nonsense to be an anti-representationinst. This is how we-are-in-the-world.

References

[1] Rosen, R. (1991). Life Itself: A Comprehensive Inquiry into the Nature, Origin, and Fabrication of Life. Columbia University Press.

[2] Husserl, E. (2008) Cartesian Meditations, Springer.

[3] Smith, B. and Smith, D.W. (eds) (1995) The Cambridge Companion to Husserl. CAmbridge University Press.

[4] Baars, B.J.(1997) In the Theater of Consciousness: TheWorkspace of the Mind, Oxford University Press.

[5] Box, G.E.P. and Draper, N.R. (1987). Empirical Model-Building and Response Surfaces, p. 424, Wiley.

[6] Dreyfus, H.L. (2007) Why Heideggerian AI failed and how fixing it would require making it more Heideggerian, Artificial Intelligence, Volume 171, Issue 18.

[7] Jean Petitot http://www.crea.polytechnique.fr/JeanPetitot/home.html

[8] Bunge, M. (2000) Systemism: the alternative to individualism and holism, Journal of Socio-Economics 29, pp. 147-157.

[9] Mario Bunge, "La investigación científica ". Ediciones Ariel. Barcelona, 1969

[10] Martin Heidegger, in The Question Concerning Technology and Other Essays, trans. William Lovitt (New York: Harper and Row, 1977

[11] Rizzolatti G, Fogassi L, Gallese V. , Cortical mechanisms subserving object grasping and action recognition: A new view on the cortical motor functions. In: Gazzaniga MS, editor. The new cognitive neurosciences, 2nd ed Cambridge (Massachusetts): MIT Press, 2000

[12] Edelman, G.M. (2005).Wider Than the Sky: The Phenomenal Gift of Consciousness. Yale University Press.

[13] Korzybski, A. (1931) "A Non-Aristotelian System and its Necessity for Rigour in Mathematics and Physics," Meeting of the American Association for the Advancement of Science, December 28, 1931.

[14] Agre,P.E. (1997) Computation and Human Experience (Learning in Doing: Social, Cognitive & Computational Perspectives). Cambridge University Press.

[15] Merleau-Ponty, M. (1984) The Structure of Behavior, Duquesne University Press; New Edition.

[16] Powers,W.T. (1989) Living Control Systems: Selected Papers. Control Systems Group.

[17] Van Gelder,T. (1997) Dynamics and cognition, in Mind Design II, John Haugeland, Ed., A Bradford Book. The MIT Press, pp. 439-448.

[18] Conant, R. C. (1969). The information transfer required in regulatory processes. IEEE Transactions on Systems Science and Cybernetics, 5(4):334-338.

[19] Conant, R. C. and Ashby,W. R. (1970). Every good regulator of a system must be a model of that system. International Journal of Systems Science, 1(2):89-97.

[20] Arkin, R.C. (1998) Behavior-based Robotics.The MIT Press.

Pragmatics and Its Implications for Multiagent Systems

Tariq Samad

Honeywell Automation and Control Solutions

Abstract

Pragmatics is the subfield of linguistics that is concerned with the study of language use in real-world contexts. Coordination among humans, for both simple and complex tasks, relies heavily on linguistic pragmatics and its sophisticated interplay of communication, reasoning, and knowledge. Although specific developments in pragmatics have inspired research efforts in multiagent systems, the importance of this linguistics discipline as a whole to the engineering of intelligence and autonomy is rarely emphasized. Relevant topics in pragmatics that are reviewed include deixis, conversational implicature, speech acts, and conversational structure. The article illustrates how incorporating pragmatics can permit more secure and succinct communication and more robust, efficient, and effective coordination, with attendant benefits for multiagent system performance. Examples of hypothetical pragmatics-endowed teams of autonomous unmanned aerial vehicles (UAVs), robots, and intelligent software agents are outlined. Some preliminary remarks toward formalizing pragmatics are offered.

What, reduced to their simplest reciprocal form, were
Bloom's thoughts about Stephen's thoughts about
Bloom and Bloom's thoughts about Stephen's
thoughts about Bloom's thoughts about Stephen?
—James Joyce, *Ulysses*

1 Introduction

Language is a distinctively human phenomenon, and perhaps it is the single most obvious manifestation of our superior intelligence as a species. Any attempt toward the development of mind theory, then, must encompass the human language faculty. But the implication goes beyond theoretical research. If we are to ultimately realize mindlike engineered intelligent systems, language in the sense of a communication competency that incorporates the sophistication and complexity that natural languages provide for human interlocutors—must be engineered too.

Language is used—indeed, is essential—not just for self-expression in humans but for effective coordination and collaboration among us. Whether in social networks, corporate organizations, or military hierarchies, and whether in

one-on-one engagements, coordination within small physically collocated groups, or large-scale collaborations spanning remote participants, people work together for personal and collective purposes, relying on shared linguistic competencies to convey a variety of kinds of content and for a variety of practical purposes. It is through language, typically, that we issue commands, communicate objectives, provide status reports, offer feedback, and share our individual knowledge and opinions. Furthermore, in noncooperative, or not entirely cooperative, situations, language is also employed for negotiation, deception, and rhetoric.

The study of language structure and use is a multifaceted field. Three areas of inquiry are often distinguished, two of which are especially well-established. The best-known area is *syntax*, which is concerned with the composition of lexical items such as words, in particular to form complete and correct sentences. Syntactical theories are in the form of generalizations, formally expressed, that hope to capture regularities implicit in specific languages, or in human language in general. The concept of grammar, for example, is a syntactic concept. At least in its stereotypical definition, syntax is not concerned with meaning per se. Sentences can be grammatically correct yet meaningless, as exemplified by the Chomskyan composition, "Colorless green ideas sleep furiously" (Chomsky, 1957; but see Pereira, 2000).

In contrast to syntax, the study of what sentences mean, as distinct from what makes them grammatically correct, falls under the topic of *semantics*. In most cases, linguistic semantics focuses on literal and sentential meaning—on how meanings of words combine to form the meaning of the sentence. Extralinguistic factors, such as the time and place of the utterance or shared knowledge about the world, as well as the broader linguistic context, such as the previous communication between speakers, are excluded.

Both syntax and semantics have had an impact on technology. Computer languages and compiler theory owe a substantial debt to Chomsky's early, revolutionary theories of syntactical structures, and natural language understanding—the field of artificial intelligence concerned with the development of computer programs that allow humans to interact with machines using a natural language such as English—has relied extensively on linguistic theories of syntax and semantics.

But our use of language goes well beyond syntax and semantics. Consider even the simple question:

Can you pass me the salt shaker? (1)

Said at the dinner table, this question is not intended to be interpreted literally. A purely semantic analysis would suggest that only a linguistic response ("yes") is required. After all, if the speaker really wanted the salt at that moment, would it not be more succinct and direct to say

Pass me the salt now.

instead?

Or consider the following exchange, perhaps from a breakfast table of a bygone era (Levinson, 1983):

A: Do you know what time it is? (2)
B: Well, the milkman just came.

Once again, it's clear from this example that surface structure and literal meaning is insufficient for understanding language use. The situational context, shared knowledge, awareness of others' intents and beliefs, all are involved.

The subfield of linguistics that is specifically concerned with language use in real-world contexts and with how people manage to effectively and efficiently convey information and coordinate their activities through language is *pragmatics*. Although it has a long and distinguished intellectual heritage, pragmatics as a linguistic discipline in itself is relatively new (Clark, Eschholz, and Rosa 1981). The research literature in the field is more descriptive than rigorous. There is no shortage of insights and examples that hint at deeper models and theories, or concepts and generalizations that suggest explanations for observed phenomena, but the rigor that is associated with syntax and semantics is lacking.

And perhaps for good reason… The very characteristics that privilege pragmatics—its coupling of the mental with the physical, the individual with the social—create a disciplinary challenge. Unlike syntax and semantics, pragmatics cannot be meaningfully studied under a "closed world" assumption—pragmatics is not about language as a human competency in itself but about language use by humans within the context of their interactions with other people and with the physical world at large.

My objective in this paper is to raise awareness of pragmatics with specific attention to its relevance and importance for mind theory and its future technological implementations. Understanding communication and coordination among humans is a worthwhile goal in itself and one that can be well served by taking pragmatics from descriptive theory to realizable technology. In addition, I claim that pragmatics will be essential for attaining the vision of machines that exhibit humanlike levels of intelligence and autonomy. This vision, of course, continues to motivate research in numerous disciplines of science and technology. Seeking inspiration from human cognition for intelligent systems is not new; expert systems and fuzzy logic are a result of this inspiration and have been integrated in computer-based engineered systems with success in practical applications.

An overview of pragmatics, covering its major topics, appears in the next section. Since rigorous formulations are largely absent in the field, the discussion is unavoidably intuitive rather than formal—its purpose is to motivate further, focused research rather than to present fully developed results.

Another lacuna in pragmatics is a pragmatic orientation! The issue of operational benefits of pragmatic devices for engineered systems remains largely unarticulated—and, without an explication of these benefits, the relevance of pragmatics as a broad-based technology will remain in question. It may be hypothesized, for example, that people rely on pragmatics for reasons that are deemed irrelevant for intelligent systems of our devising. Without taking a position on whether or not compelling reasons exist to endow systems that we have the liberty to design for our practical purposes with such human characteristics as politeness, shyness, and social mores, in Section 3 I suggest a number of (other) practical benefits that pragmatic devices can help realize for multiagent systems—whether human or engineered or in symbiotic integration.

Section 4 sketches three conceptual examples of multiagent systems endowed with pragmatics capabilities. The examples include robotic teams and software agents. In each case scenarios are outlined to suggest how pragmatics, developed to the point of mature technology, could result in improved performance of complex engineered systems—and in the interim could help understand human coordination in some cases.

As might be considered befitting a proposed subject within an emerging field—i.e., mind theory—the discussion in this article is speculative and anecdotal. The article does not present specific technical results but is intended to highlight the role of pragmatics in human interaction and to motivate the need for, and approaches to, interweaving knowledge, reasoning, and communication in multiagent systems. If theory is to proceed ultimately to realization, however, pragmatics will need a formal, rigorous treatment. Some preliminary remarks related to formalizing pragmatics are included in Section 5.

Some points of clarification before concluding this introduction… First, pragmatics is considered a subfield of linguistics and the vast majority of the work in the field is concerned with language use. However, people often use language in concert with other communication cues and sometimes such cues can substitute for linguistic expressions. Thus the use of gesture and, analogously from the "receiver's" side, the observation and interpretation of visual signals, is often subsumed within pragmatics.

On a related note, I follow linguistics terminology in referring to speakers, hearers, and utterances. But these terms should be understood, as they are in linguistic pragmatics, in a more general sense. Thus they cover analogous terms in written communications and, as mentioned above, gestures and other visual signals (modalities such as touch can be covered too by extension).

Finally, the examples used in this article are all in English. As might be expected, languages across human societies differ markedly in how they manifest many areas of pragmatics. Readers interested in cross-language comparisons are referred to Huang (2007).

2 Selected Topics in Pragmatics

I discuss five subareas of linguistic pragmatics in this section: speech acts, deixis, presuppositions, conversational implicature, and conversational structure. Outside of linguistics, pragmatics is often identified with a subset of these subareas, often just the first of them. Only capsule summaries are provided; comprehensive descriptions would require a much more extensive treatment than is possible here.

2.1 Speech Acts

Things people say are said for some purpose. Speech act theory studies how, in the words of John Austin, the philosopher of language who initiated the modern investigation of the concept, "by saying something we do something" (Austin, 1955).

Often what we do through an utterance is purely communicative—for example, we inform the hearer about an observation, we ask a question, we greet or bid farewell. The intent or effect isn't necessarily apparent from the form or literal meaning, as already seen in (1) above, which is uttered as, and recognized as, a request despite its question structure.

A particularly interesting category of speech acts consists of those in which the effect of the utterance is to change the state of the world, in some localized sense. Austin gives the following examples of what he termed "performative sentences":

> "I do" (said in a marriage ceremony)
> "I name this ship the *Queen Elizabeth*" (said when smashing the bottle against the stern of the ship)
> "I give and bequeath my wristwatch to my brother" (written in a will)
> "I bet you sixpence it will rain tomorrow"

In these cases, the saying or writing of a few words, in the appropriate circumstances, has extralinguistic, extramental consequences, viz., a legal marriage, the naming of a ship, the (future, contingent) transfer of property, and a contingent financial transaction, respectively.

Research in speech acts is concerned with issues such as categorization of these acts, whether or not the concept can be reduced to syntax and semantics, and on determining "felicity conditions" under which speech acts can be effectively performed. Searle (1969) identifies and exemplifies three such conditions. For the example of an order, felicity conditions include that the speaker is in a position of authority over the hearer (a "preparatory condition"), that the speaker wants the ordered act to be performed (a "sincerity condition"), and that the speaker intends the utterance as an attempt to get the hearer to do the act (an "essential condition").

As with most areas of pragmatics, theoretical developments have not progressed to the point of broadly accepted, rigorous formulations or models,

although for intimations toward this objective see Jurafsky (2004), where algorithms and computational models for the interpretation of speech acts are presented and discussed. Speech act theory has also influenced the development of agent communication languages, as discussed in Dignum and Greaves (2000).

2.2 Deixis

In Levinson's (1983) definition, "deixis concerns the ways in which languages encode or grammaticalize features of the context of utterance or speech event, and thus also concerns ways in which the interpretation of utterances depends on the analysis of that context of utterance." Deictic references to person, place, and time are commonplace in normal language use. Persons engaged in the utterance can be referred to directly (e.g., through the use of first/speaker or second/hearer person pronouns) and indirectly (the third person). Place deixis in its simplest manifestation distinguishes between locations that are close to the speaker ("here" or "this") and locations that are distant ("there" or "that"). Temporal references are expressed (in English) both with adverbs ("now," "then," "yesterday") and through tense markings. Variations, extensions, and alternative expressions abound, however. Just to note one, instantaneous time in the hearer's frame can be referenced through the timing of speech ("At the count of three, …") or even in text, when there may be a temporal separation between writer and reader ("As soon as you read this, …").

The last example is not just a temporal reference, it is a deictic reference to the text itself (as in "The last example ..."!). Textual references can be self-referential—a philosophically and logically interesting topic in its own right as evidenced by the extensive literature on the liar paradox, such as the sentence,

> *This sentence is false.*

Deixis can also connect language with gesture, image, and other extralinguistic phenomena—thus when I point my index finger somewhere and say, "Go there!", I am using "there" to refer to a specific location that can only be identified by observing the visual cue. Self-reference can play here too—cf. Magritte's "This is not a Pipe" painting (and Foucault, 1982).

For an overview of recent research related to deixis, see Akman and Bazzanella (2003) and the associated special issue of the *Journal of Pragmatics*. Another category of deixis that has been studied is social deixis; see (Huang, 2007; p. 163) for a discussion that highlights how social status and relationships are encoded in different languages.

2.3 Conversational Implicature

As virtually any interpersonal linguistic exchange will demonstrate, the conventional use of language in natural contexts relies extensively on the reasoning capabilities of participants. Conversational implicature refers to how we manage to say, and understand, more than is literally said, often by relying on

what may seem common sense. Sentence (2) in the introduction is one example. See Figure 1 for another (also referring to a bygone era).

How and why do implicatures work? Grice's "maxims of conversation" were the first, and continue to be the best known, attempt toward a systematic analysis (Grice, 1989). The maxims are four in number and are as follows:

The maxim of Quality:
- Do not say what you believe to be false.
- Do not say that for which you lack adequate evidence.

The maxim of Quantity:
- Make your contribution as informative as is required for the current purposes of the exchange.
- Do not make your contribution more informative than is required.

The maxim of Relevance:
- Be relevant.

The maxim of Manner:
- Avoid obscurity of expression.
- Avoid ambiguity.
- Be brief.
- Be orderly.

Figure 1. Conversational implicature in action (Meehan, 2004; reproduced with permission of the artist).

These maxims are assumed by hearers to be in effect. When the literal meaning appears to violate a maxim hearers assume that a nonliteral interpretation is required. Conversely, speakers know that hearers are assuming these maxims hold and can therefore proffer indirect utterances. The Gricean maxims are not rigorous or foolproof rules but they help explicate how nonliterally intended utterances work. In the comic strip above, Rolf's ultimate aha (or "aaah…") comes from realizing why his partner posed the question—superficially the question appears irrelevant. (And the partner can pose the question rather than directly suggesting that Rolf wash because she predicts Rolf's reasoning.)

Conversational implicature arises from speakers and hearers relying on models of their interlocutors. These models can in fact be recursive, as expressed by the *Ulysses* quote at the beginning of this article. Such recursion leads to interesting puzzles in mathematical logic that bear strongly on the potential

for conversational implicature to effect sophisticated distributed reasoning. Consider the "muddy children" puzzle (Fagin et al., 1995): n children are playing together and $k > 1$ of them have a dab of mud on their foreheads. A child does not know if he or she is muddy but each can see the foreheads of all other children. An adult tells the group something that each child already knows: "At least one of you has mud on your forehead." The adult now proceeds to repeatedly ask the question: "Does any of you know that you have mud on your forehead?" If the children are perfectly rational and intelligent, they will all say "No" for the first $k - 1$ times the question is asked and all and only the children with muddy foreheads will answer "Yes" on the kth asking.

2.4 Conversational Structure and Discourse Analysis

Much recent research in pragmatics focuses on the structure of discourse and conversation (Gee, 2005). Work in this area is empirically driven, with significant effort put toward documenting and analyzing conversations in real-world settings. The discourse-level scope taken means that many other topics in pragmatics come into play here as well. Research in conversational structure analysis attempts to answer questions such as: Can generalizations be derived that can explain how people converse? Are there different types of conversation that, for example, follow different patterns? How does relative social standing influence discourse? What cues, linguistic and otherwise, help with discourse transitions and tags?

As an example of raw material for conversational structure analysis, reproduced in Figure 2 is a near-verbatim transcript of a series of short cellular telephone conversations overheard at an airport lounge over the course of about a half hour. The context is a professional one and the transcript illustrates a coordination activity among four people (only one of whom is heard here).

Alice?
How are you? Tired? A bit tired.
I just picked up your message. I just landed.
Bob left me a message saying he wouldn't be able to make the call but he didn't tell me what time it was.
Would Charlie...?
Should I call you on your desk phone? Ok. Splendid. Thanks, Alice.

Hi Alice.
Oh... still tired ...
Yah, just landed a few minutes.
Ok, you know the easiest thing would be to do tomorrow morning ... if that's possible.
Ok
Yah, because we have this crisis in Andover so the easiest thing ... would be to go down there.
I could meet you for breakfast. The whole morning would be free. Actually,...
Ok ... aah ... ok ... ok
Yes ... ok ... ok ... I'm just gonna stay here in this airport for a little while. Not that I

don't want to go over this afternoon. If he can't see us tomorrow morning.
Oh I've got a nice crash place here. Yeah.
I don't know whether Bob can join us tomorrow morning either. His daughter's in the hospital ... yeah, sounds pretty serious. She's five years old. I don't even want to harass him; I'll leave him a message.
Me too. But preference is tomorrow.
Ok, thanks Alice. Thanks. Bye.

Davis here, Bob. I'll try your cell phone. I just spoke to Kathy. I could meet you for lunch around noon. A great pleasure as always. I'll call your cell phone.

Davis here, Bob. Alice … I've talked to Alice. I'm just going to hang out here at the airport. Have wireless access here. She's going to call the guy about meeting tomorrow. I'll be able to see you for lunch. Alright, see you then.

Dan Davis.
Ok, Alice.
Yes. Hi Charlie.
Yes I am ... and ... I have a pretty tight schedule.
The best time for me to come up would be tomorrow morning. I don't know if that's convenient for you.
Well.. that's kind of up to you. I know Alice has a little bit of a conflict later in the morning.
Ok.
Alright. Thanks Charlie. Look forward to seeing you tomorrow. Byebye.

Yes. It's been a while.... which one is it?
Geez, ohh, ok. I ... I gave her a number of those.
I can picture that door. I can picture freeway. And I can picture everything.
Yes...
Yes, just go all the way, yeh.
Okay. Nine a.m. Splendid. Alright, well, let me know if anything changes, otherwise I'll see you tomorrow morning.
I will... I will do that, yeh. Ok thanks Alice. Bye.

Figure 2. A near-verbatim transcript of a cellular telephone call overheard in an airport lounge.

The greeting-content-closing structure is evident in all calls, with the purposeful part of the content sometimes prefaced by social niceties depending on the relationship between the parties involved and their recent interactions. In the calls, as distinct from the voicemails being left, frequent acknowledgements are interjected when the other person is talking. An "okay" or "alright" or similar token is typical before a conversation is concluded; these words serve as "discourse markers." (For a computational connection, Zufferey and Popescu-Belis [2004] use decision-tree classifiers to automatically identify when lexical tokens mark discourse transitions, exemplified by the ambiguous word "like" and "well.")

It is also apparent from the transcript, as indeed it is in hearing language use in any natural context, that syntax and grammar rules are not rigidly followed. Incomplete sentences and elisions are ubiquitous.

2.5 Presuppositions

In pragmatics, presupposition is a technical term that can be contrasted with implication or entailment. "John is a boy" semantically entails "John is male." But to take a well-known sentence from Bertrand Russell, "The present king of France is bald," incorporates the pragmatic presupposition that (the speaker believes that) France has a king. One test suggested for discriminating between entailments and presuppositions is that the latter remain constant under negation whereas the former do not. So (another well-known example), both "John has stopped beating his wife" and "John has not stopped beating his wife" presuppose that John was beating his wife—the basis of the Groucho Marx question, "Have you stopped beating your wife?" (either a yes or no answer damns the respondent).

Some linguists have suggested that presuppositions can be adequately handled by semantics alone—that the utterance provides all the context necessary for determining whether a clause is presupposed or not, at least under certain conditions. Often this context is based on certain trigger words. "Stopped" above is one example; "know" is another—when used in the second or third person, "know" even in negation indicates that the speaker believes the predicated assertion as being true. "John doesn't know that Jill is an engineer" presupposes that Jill is an engineer. "Know" can be distinguished from "belief" in this respect.

As counterpoint to the semanticists (but see Hegarty, 1992), Levinson (1983) gives the example of "before" usually presupposing the phrase it heads:

Sue cried before she finished her thesis. (3)

This presupposes that Sue did in fact finish her thesis. But now consider

Sue died before she finished her thesis.

So our knowledge about what happens to people when they die (i.e., they are no longer able to do things) overcomes the presupposition potential of "before" as exemplified by (3) and hence the extra-utterance context of world knowledge is required to correctly reject the presupposition.

3 Why Pragmatics?

Natural (or human) language is the principal means of human communication because it has evolved, with all its complexity, to satisfy our individual and societal needs. But what specific practical benefits, if any, accrue from the incorporation of pragmatics mechanisms in human discourse? Is pragmatics pervasive just because it is an unavoidable side-effect of human cognitive limitations, emotional makeup, or social circumstances? If our interest in the

topic extends beyond linguistics or cognitive science and toward engineering applications, the potential practical improvements that engineered multiagent systems could achieve if endowed with pragmatic mechanisms—the complexity of which will certainly be substantial—must be articulated.

I am not aware of any systematic attempt to answer the above questions—as a linguistics specialization, the study of pragmatics emphasizes description, not justification outside the human-discourse context. Hence this section, where I briefly note some advantages that pragmatics brings for systems, whether human or otherwise.

3.1 Distributed coordination

Human teams are models of distributed coordination. In small and large groups, we manage to accomplish complex tasks and objectives. Teams today can be geographically dispersed to the point of having members who know each other only through e-mails and conference calls throughout a project's duration. Yet, our coordination is usually robust to individual failures, situational changes, and other complications.

Organizations differ tremendously in their scale and structure, but even in rigidly hierarchical ones, the intelligence and autonomy embodied in every individual have the potential to positively affect organizational performance. In principle these very characteristics pose a substantial coordination challenge, but, in part through pragmatics, we are able to exploit them to advantage. In well-performing organizations, people tend to know not only what to communicate, but when to communicate, whom to communicate to, and how to communicate. Models of others' goals, giving feedback during planning and execution in the right form, and contextualizing content are all essential. For example, in global organizations, locational and temporal deictic devices are adapted accordingly. "Today" and "tomorrow" can be ambiguous. The lack of affect in digital communication is compensated for through a myriad of adaptations, such as emoticons, resulting in a "pragmatics of computer-mediated communications (CMC)" (Herring, Stein, and Virtanen, 2009). Through pragmatics (and other processes) the potential chaos of multiple autonomous entities doing their own thing is harnessed for overall benefit.

From a team performance point of view, the contrast between distributed coordination and centralized decision making is instructive. If a single agent can observe global state and rapidly compute and communicate optimized actions for all other agents, repeating the process as rapidly as required for the application, the overall system can attain high levels of performance with little need for sophisticated and subtle communication among agents. It is when the centralized approach is untenable —because of the scale of the system or the complexity of the problem or some other reason—that individual agents need to have significant levels of autonomy. The cues and mechanisms characteristic of pragmatics are required then. Individual agents can determine actions that are appropriate for the team, not just themselves, without centralized processing and decision making. Models of the team and team members,

even recursive/reflexive models, are essential at local levels if distributed coordination is to be effective.

3.2 Coordination efficiency

Use of pragmatics enables dramatic reduction in communication requirements compared to schemes where agents do not have the ability to process messages in context. Conversational implicature and deixis allow detailed exchange of information with limited communication bandwidth—the receiver of the communication can overcome the lack of explicit detail by applying its knowledge, by reasoning, and by referring to the context.

Deictic devices reduce communication bandwidth through anaphoric usage—enabling people, places, actions, and things previously mentioned to be referenced again in short-hand—and also through references to the discourse or conversation itself, or parts thereof. When a participant in a meeting says,

> Well, on balance, I think I'd like to recommend we go with the first option we discussed.

she relies on the shared recent context to keep from repeating what could be a long text.

Pragmatics allows coordination efficiency through conversational implicature as well. Agents can rely on their models of others instead of relying solely on communication to short-circuit negotiation or coordination iterations or to provide information succinctly. "Well, the milkman just came" (Sentence 2) can be understood as shorthand for a much longer statement: "I do not know the time exactly, but I know that you know what time the milkman usually comes. I also know that the milkman just came, so by giving you the information that he just came I know you will have information relevant to knowing the time. And even if you already know that the milkman has come, my reminding you of this fact in response to your question may help you arrive at an answer to your question." Note that the fact that the milkman came only provides a rough lower bound on the time. Note also that the speaker may not know what time the milkman comes, but as long as he believes that his partner knows this time the statement is informative.

The efficiency benefit becomes even more important in time-critical situations, and gesture (and/or touch) may supplement language in the interests of rapid response. Sports are illuminating in this regard. American football is especially intricate. Several distinct information channels are in constant use throughout the game. Coaching staff in a booth, who have a panoramic view of the field and access to video replays, are in constant communication with the coach and his assistants on the sideline. The sideline coaches communicate with the players before each play. For example, for the team on offense, the coach is in wireless contact with the quarterback and hand signals are used as well. In the huddle before each play the quarterback communicates the called play to his teammates. When the huddle disbands and before the ball is snapped, the quarterback relies on verbal codes and gesture to communicate

options and for timing. Once the ball is snapped and the play is underway some verbal communication may still occur and in addition visual signals may be used. Similar coordination activities occur on the defense side as well. 40 seconds are allowed between the end of one play and the snap of the ball of the next so each communication cycle must be completed within this deadline. Natural language, customized verbal codes, and gesture are woven together to ensure that the team can communicate appropriate information across the multiple channels and individuals involved within this duration while being responsive to the dynamic context, especially observations of the defensive side.

3.3 Security

Security is not typically the reason behind the use of pragmatics and it is typically not an important consideration in normal conversation, but exceptions are not hard to find. Overhearing a business conversation, or even a telephone call, in an obviously public setting readily provides examples of security through pragmatics:

> I'm in an airplane and can't really talk … I'm on my way for that thing we discussed yesterday … yes, yes… that company, yes… that one issue got resolved … the lawyers earned their keep. Okay, bye.

Or consider the simple parental strategy of saying

> Let's not go for i-c-e c-r-e-a-m.

In both these cases, the interpretation of a message requires knowledge that the receiver holds but a third-party observer does not. But messages do not even need to be cryptic by design; we often rely on shared context in normal conversation to an extent that at least casual eavesdroppers gain little specific information:

> Alright, see you at the usual time and place, next time.

Security can be considered a secondary benefit of pragmatics for everyday human interactions. For some engineering applications, however, security of communications is critical for success. The pragmatic mechanisms that people use, deliberately or by happenstance, can provide additional security in some cases. Even if encryption is compromised and message content revealed, key information may be inaccessible unless some part of the receiver's state is shared.

3.4 Clarity and Disambiguation

Ambiguity is pervasive in natural languages, as perhaps it must be in all conceivable languages that hope to serve the complex communication needs of intelligent agents. Discussions of ambiguity usually focus on its syntactic and semantic varieties, where polysemy and multiple parse trees can result in multiple interpretations for a sentence. Hence (another Groucho Marx attribu-

tion) “Time flies like an arrow. Fruit flies like a banana.” In fact the first sentence of this pair has several readings. The extrasentential context is necessary for disambiguation.

The role of pragmatics for disambiguation can be much more subtle. For example, the use of conventions such as in speech acts can help to disambiguate intentions. Thus the performative utterance “I promise to repair the problem tomorrow” is a social contract. The hearer knows that the speaker (if acting in good faith) has made a commitment, perhaps as much of a commitment as the speaker can make in the hearer’s mind. Thus until the promise is given (the word “promise” itself is often but not always required), the speaker can press for greater commitment, and once it is given both parties understand the import. From the point of view of an intelligent agent negotiating with another for a commitment, the expression of a “promise” (however realized in the language that the agents communicate in) may be a signal, at least from the side of the agent pressing for commitment, of a satisfactory conclusion. The importance attached to words like “sorry” and “apologize” in many human social situations can also be seen as a desire to seek clarity on the feelings and sense of contrition of the speaker.

Indeed, recognizing the speech act that is being performed by a speaker is crucial to clarifying the intent of the utterance. Cues, such as key words or phrases, given certain contexts (i.e., felicity conditions), allow humans to effectively communicate. Thus when a person in a position of authority says, “Can you bring me X?” the hearer may understand the question as a command, whereas from a peer the utterance would more likely be interpreted as a request.

Understanding communicative intent will be no less crucial for artificial agents. Performatives such as “promise” may seem like a special case, but speech acts in general are pervasive. Artificial agents as well as humans need to know when an exchange is being initiated or terminated, when a question is being asked or answered, when a presumed useful fact is being conveyed, when a request for an action is being made, when an order is being given or acknowledged, when a third party is being introduced, and so on.

4 Examples and Applications

Autonomous agents (consider humans) interpret received messages in context. The effects of these messages on the receiving agents’ states and behaviors depend on these interpretations. Agents can communicate efficiently and effectively by taking into account this added layer of processing. Thus the intended effect of a message may not be directly reflected in its semantics. As observed above, communication can thereby be more succinct, more secure, and more effective, and overall team operation will have attendant benefits.

In this section I outline three pragmatics-enabled coordination scenarios. The examples do not cover all the pragmatics topics presented earlier; the first two are mostly focused on implicature and the associated models required, whereas the last example refers to speech acts and deixis as well. The exam-

ples are also diverse in their domains, suggesting the broad reach of pragmatics as technology. In keeping with the spirit of this article, the presentation is intuitive rather than formal. An abbreviated version of the first example appears in Samad, Bay, and Godbole (2007).

4.1 Example 1

What an agent says or does in a team context will depend not only on what it believes. It may also need to take into account (its knowledge of) its team members' beliefs about its knowledge or beliefs.

To illustrate the point, let A and B be two autonomous unmanned aerial vehicles (UAVs): A issues a command to B to "Go to the target." Assume that A believes that the target is at x and that A believes that B believes that the target is at x. We can represent the latter belief of A as

$$\mathcal{B}_{\mathrm{A}} [\, \mathcal{B}_{\mathrm{B}} (\text{“target at } x\text{”}) \,]. \tag{4}$$

On the basis of this belief, A may think it knows where B will go. However, solely on the basis of (4), A cannot know if B will know that its (B's) destination is known to A. For this latter belief (about B's belief about A's state of knowledge) to hold, A's belief state must also include the following:

$$\mathcal{B}_{\mathrm{A}} [\, \mathcal{B}_{\mathrm{B}} [\, \mathcal{B}_{\mathrm{A}} [\, \mathcal{B}_{\mathrm{B}} (\text{“target at } x\text{”}) \,] \,] \,].$$

(That is, A must believe that B believes that A believes that B believes the target is at x.)

On the other hand, if instead of the above,

$$\mathcal{B}_{\mathrm{A}} [\, \mathcal{B}_{\mathrm{B}} [\, \mathcal{B}_{\mathrm{A}} [\, \neg\, \mathcal{B}_{\mathrm{B}} (\text{“target at } x\text{”}) \,] \,] \,], \tag{5}$$

then A will believe that B thinks that A thinks that B is not headed to x (even though because of (4) A in fact does think that B believes the target is at x). If A is correct in its belief about B's beliefs about its (A's) beliefs per (5), B may in fact not head to the target even though both A and B believe the target is at x—even if B knows the target is at x it could decide that it was not A's intent to direct it to that location.

Thus the representation of one agent's beliefs by another, where that belief representation incorporates (beliefs regarding) the other agent's beliefs … and so on recursively … can influence coordination strategies. Mutual and reciprocal knowledge and belief have many further implications—examples can be developed that require arbitrarily deep recursions, such as in the case of the muddy children puzzle discussed in Section 2.3.

Given the convoluted thought processes illustrated above, it may not occasion much surprise to see how a reasoning agent may be able to achieve its objectives more simply when its follower is relatively naïve. In the above scenario, if A believes that B thinks there is a target at x, but B maintains no beliefs about A (and A is aware of B's naivety in this regard), then A can give the

command "Go to the target" to B without having to contemplate mutual and reciprocal beliefs. On the other hand, the shared complexity brings many benefits. For example, if B believes differently than what it thinks A believes about it, the two may, through informed negotiation and communication, jointly arrive at a common belief on the target location—a belief that is the better supported of A's and B's beliefs.

The case of noncooperative interaction is interesting as well. If two agents are competing, and A thinks that B's objectives, situation, or beliefs are different than what they really are, B is at an advantage. Furthermore, an agent can intentionally mislead its competitor. Hence the notion and effectiveness of a bluff, which is essentially an attempt to influence the other into holding an incorrect belief about one's situation or belief.

Some work in discrete-event control systems shows how reasoning, communication, and reciprocal belief can result in effective coordination and is notable here. Ricker and Rudie (2003) incorporate inferencing capabilities in knowledge-based systems, thereby extending the range of decision-making problems that can be solved. The key is that agents make inferences not just about their own capabilities but about the capabilities of other agents—an agent's decision can be based on what another agent is expected to decide.

4.2 Example 2

Pragmatics is relevant even when speech or other explicit communication modalities are not involved. Assume a team of three autonomous robots, with A as the leader and B and C both following A's trajectory. A has two objectives in determining its motion: the team should stay together and it should reach the destination before a deadline. B and C strive to maintain a given distance d from A. No explicit communication is possible (perhaps for stealth reasons) but each robot can observe the others (and each knows that the others can observe it). Unknown to A and C, B is short of fuel. We also assume (realistically) that the range of a robot increases with reduced speed and (futuristically) that the robots are aware of this dependence.

The baseline scenario is as follows. A proceeds as fast as it can. If this speed is greater than what B or C are capable of, one or the other will start to lag. A can observe the increasing separation and reduce its speed. With appropriate control logic, after a transient period the three proceed in formation at the maximum speed that all can achieve. However, after some time B's fuel is exhausted and it stops before the team can reach its goal.

In an alternative scenario, B knows that A's objectives include keeping the team together. Knowing that it will not be able to reach the destination at the current speed, B slows down (to a speed computed to permit it to reach the destination, with a safety margin). A predictably slows down too in the interests of maintaining the formation, and, following its lead, so does C. The full team manages to reach the destination, later than A and C could have reached it by themselves but (so we assume) prior to the deadline.

In another alternative scenario, A may determine that at the reduced speed the estimated time of arrival (ETA) is too late and that reaching the destination in time is a higher priority objective than ensuring that multiple robots reach together. In this case A may allow B to lag and A and C could arrive at the goal in time.

These anthropomorphic explanations could also be realized more simply, by hardwired rules. Thus A could be following rules such as (roughly):

R1: If separation from a follower is increasing, reduce speed
R2: If ETA > t_{max}, increase speed

with a prioritization of R2 > R1.

But the more general and robust, if computationally more challenging, solution is to have the agents reason on the basis of their knowledge of their own and their partners' objectives. By simulating the effect of a hypothesized slowdown, B could predict that A will slow down too, and if the execution of this plan isn't successful it would realize—since it knows that A observed its speed reduction and deliberately ignored it—that its tactic was being overridden by its leader.

It is notable too that in the pragmatics-assisted scenarios, communication requirements are reduced, relative to a fully centralized architecture, in terms of both bandwidth and the type of information communicated—i.e., communication of B's fuel state is not required. In addition, the distributed approach can more readily be extended to multiple agents.

Although the pragmatics connection is not noted by them, Papageorgiou and Cofer (2003) show how leader-follower configurations can communicate trajectory intent in situations in which the communication and/or computation bandwidth is restricted. The scheme described is inspired in part by how human pilots coordinate formation flight when under stealth communication constraints.

4.3 Example 3

After the above examples focusing on mobile agents and physical coordination, for a final example I turn to coordination of information sharing and decision making. Consider organizational decision making, as represented by, say, a conference call in a corporate setting. Such a meeting may seem essentially human in nature today, but we can attempt to abstract away from the essentially human aspects toward a hypothetical future where intelligent artificial agents, either among themselves or jointly with human agents, engage in negotiation and decision making.

The meeting may have been called by a project leader to present the results and recommendations of a project in an area of emerging and/or strategic importance. The participants might be (possibly avatars—in the software agents sense—of) heads of business divisions, leaders of marketing and tech-

nology functions, representatives from government relations groups within the company, and selected project workers.

Numerous sorts of activities take place in such meetings. Data is presented—often the first major item on the agenda is a presentation. Presenters of data will provide their own interpretations which may be disputed by others. Implications for the company are usually a major topic of discussion. These can relate to technological, business, legal, legislative, and other positions. The discussion can include the competitive landscape and near- and long-term impacts, for example. Usually such meetings are expected to result in actions. To this end, recommendations can be presented by the project leader for discussion and approval. These recommendations may require reprioritization of investments or additional tasks for the project team or others.

Several typical activities take place during the course of such meetings. These activities are sometimes explicitly demarcated but at other times participants must rely on contextual cues. Different participants have different responsibilities or expectations on them for the different activities. Table 1 attempts a characterization of some of the roles performed in such business meetings. The contributions and roles indicated are understood as such by speakers in forming their utterances and also understood as such by hearers. Most of these utterances can be considered speech acts within the specific context of the meeting.

A few additional remarks related to the structure and procedures of such meetings.

If the meeting extends beyond its scheduled time and the senior-most participants (e.g., business leaders) remain on the call, other participants are usually expected to stay on as well—their conflicting commitments will be considered of lower priority unless critical.

The level of project support staff will make a significant difference in their level of participation. More senior participants (relative to the organizer and project lead) who are involved in an advisory capacity in the project may serve as champions during the meeting. Relatively junior staff will take a more passive role.

An explicit go/no-go from the senior-most person within an organization is usually a terminal statement for that group. Elaboration and clarification may be requested and offered but once a decision is communicated (as distinct from intimations toward a decision) it is construed as final for that meeting.

Participants need to know whether a question or a point raised is implicitly addressed to a particular person or function. This will often not be obvious from the surface form of the utterance (i.e., a person may not be addressed directly). For example, if an utterance refers to a particular organizational function participants associated with that function are the implicit addressees.

Often conference calls will start with some people not having joined in. Whether a late joiner is briefed or not on the status of the discussion to that point depends on his/her/its organizational role in particular.

The conference call setting, increasingly popular in today's global organizations, imposes constraints that are not found in face-to-face meetings. Most obviously, gesture and facial expression are no longer part of the pragmatics vocabulary. Furthermore, the lack, except where special tools are deployed, of a whiteboard or large-format notepad implies that there is no commonly viewable record of points and issues.

I have limited this example to one kind of corporate meeting with participants in particular roles. Participant roles and meeting procedures, along with linguistic and other markers, will be different for other types of meetings.

To reiterate, the point of this example is to suggest that effective participation in business decision making requires knowledge of and sensitivity to the pragmatics exercised in this environment. Company staff acquire the required expertise through experience and coaching; it is rarely articulated explicitly. How software agents will gain the appropriate competency is an open question, but without suitably sophisticated pragmatics the roles for intelligent systems will always be circumscribed.

Table 1. Participants and roles for a corporate meeting.

	Meeting organizer and project leader	***Project support staff***	***Functional leaders***	***Business leaders***
Roll call—pro forma	*Initiator*	*May introduce themselves if not known to some participants*	*May introduce themselves if not known to some participants*	*No introduction typically needed or done*
Agenda and objectives—typically brief	*Initiator*	*No comments expected*	*No comments expected*	*May raise other related topics*
Presentation—questions and comments usually allowed during its course	*May cover entirely or may present jointly with project support staff*	*May help with formal presentation or provide clarifications and comments as needed*	*Clarification questions for project team*	*Clarification questions—can be asked of project team or functional leaders*
Recommendations and actions discussion—usually overlaps with Q&A	*Initiator*	*Clarifications and comments*	*Opinions and views; may make alternative recommendations (may or may not be aligned with project team's)*	*May take ownership and lead if supportive of proposed effort; serve as final authority in cases of indecision or conflict*
Time management	*Responsible for keeping meeting on time*	*Expected to be cognizant of time constraints*	*Expected to be cognizant of time constraints*	*Can cut meeting short or propose extension*
Summary and conclusion	*Initiator*	*May support if needed*	*Agreement*	*Agreement; may also initiate*

Although the focus here has been on the coordination procedures and cues involved, pragmatics is also relevant to negotiation and discussion among intelligent agents, human or otherwise. See Rahwan (2005) and citations therein for argumentation approaches in multi-agent systems research that bear directly on this example.

5 Toward Formalizing Pragmatics

The lack of formal treatment of pragmatic phenomena noted above has a principled heritage. Logicians and semanticists concerned with linguistic meaning, such as Carnap, Russell, and the early Wittgenstein, focused their attention on a limited aspect of language where progress seemed feasible—the meaning of sentences in isolation of the linguistic or extra-linguistic context. As Habermas puts it, "On [Carnap's] view, the pragmatics of language is not determined by a general system of reconstructible rules in such a way that it could be opened up to conceptual analysis like syntax and semantics" (Habermas, 1998; p. 108). The subsequent development of pragmatics can be seen as a reaction to formal semantics, with an emphasis on empirical studies and informal analyses exemplified by the later Wittgenstein and Austin. The connection with the origins of pragmatics, especially the semiotics of C.S. Peirce with its emphasis on systematic analysis and formal methodology, became largely of historical interest. Formal analysis has started a comeback with Searle and Habermas (Cooke, 1998), but in several respects pragmatics remains a poster child of "post-analytic philosophy" (Rajchman and West, 1985).

Philosophical reflections aside, realizing practical benefits from pragmatics requires formalization and rigor. Bringing the full complexity of human language use within a mathematical framework is a far-horizon ambition but we can essay steps toward that goal. In this spirit, this section offers some brief preliminary and speculative remarks.

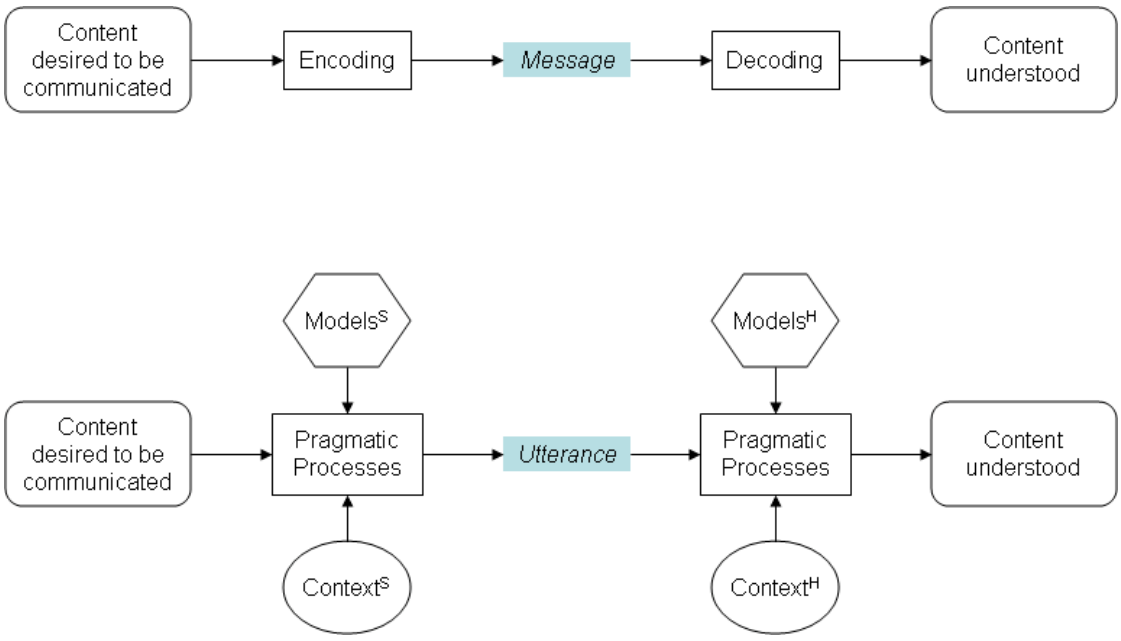

Figure 3. Communication without (top) and with (bottom) pragmatics.

5.1 Pragmatics—Implications for Intelligent Systems Architecture

We are used to thinking of communication in technological systems in terms of encoding and decoding messages. Complex algorithms may be used for

these processes but they are architecturally simple input/output blocks. Incorporating pragmatics, however, means that the "content" of the message must be processed, and this processing must be informed by contextual knowledge and pragmatic models, before the message/utterance is issued. Similarly, the receiver must also process the utterance in a context-sensitive and knowledge-based manner to understand the intent of the communication. Fig. 3 illustrates this fundamental enhancement required in the system architecture for pragmatics. (Conventional encoding and decoding of the message may still be required in the pragmatics-enabled system, prior to the issuing or processing of the utterance, and is not shown in the figure.)

As implied in the figure, models and context are private to the speaker (S) and hearer (H). Yet communication with pragmatics is only possible because of the considerable overlap between the speaker's and hearer's versions. How we capture and represent context and pragmatic models (and how these should be best represented for artificial agents) remain open research questions.

5.2 Formalizing a pragmatics of politeness

Different surface forms can be used to convey the same content. The work of Brown and Levinson (1987) in the pragmatics of politeness offers an interesting insight into how and why more-or-less polite versions of requests are used (Fig. 4). In the example illustrated, the speaker is asking the hearer for money. Brown and Levinson define the concept of "face-threatening acts" (FTA) to order the options for formulating a request. In airing a request, a speaker has to consider the prospect that the request will be denied; the notion of an FTA indicates the potential of embarrassment or awkwardness for the speaker if her request is denied. At one extreme ("off-record") the speaker has plausible deniability that she was even intending to ask for money—the indirectness of the request is a face-saving tactic. At the other extreme (on-record, without redress), there is no ambiguity about the speaker's intentions and hence the greatest threat to her "face."

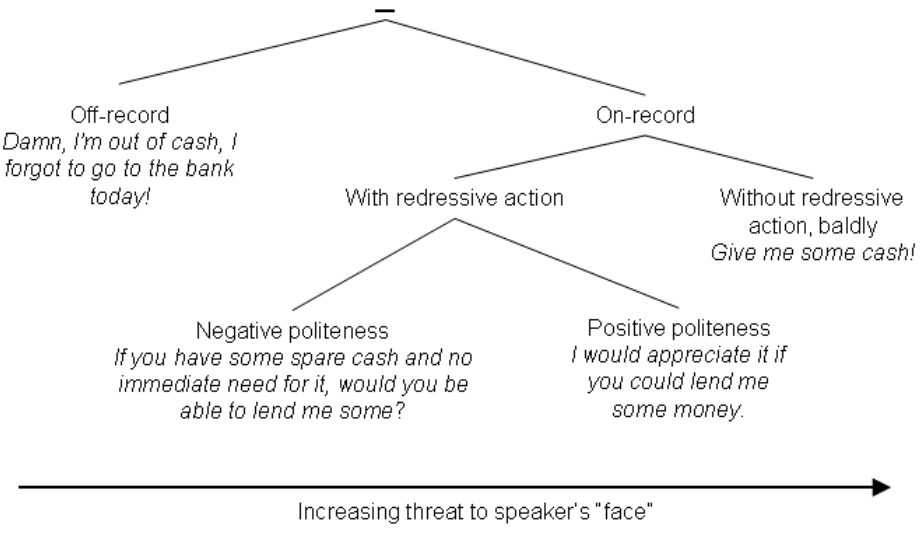

Figure 4. Strategies and examples for face-threatening acts (Brown and Levinson, 1987).

The choice of which request formulation to use in a particular situation will depend on contextual factors such as the relative social status of the hearer to the speaker and the speaker's assessment of the likelihood that the request will be granted. Direct requests are expected in socially intimate contexts; exposing oneself to the risk of an FTA is a sign of closeness—the adoption of an indirect stance will often be seen as excessively and unnecessarily formal between friends. On the other hand, a bald request to someone not close to the speaker is risky from both sides—embarrassment could result for both parties if the hearer is not in a position to accede to the request.

5.3 Modeling Conversational Implicature

As a final example, let us consider how a sentence such as "Well, the milkman just came." may be processed in the discourse context of the exchange (2) earlier. The first step is realizing that, taken literally, the response is not responsive to the question asked. Expectations of relevance would then come into play, possibly triggered by cues such as discourse markers (the initial "well" in this case). The hearer would then engage in a reasoning process to attempt to match the literal content to the information requested. Fig. 5 charts the steps and decisions involved.

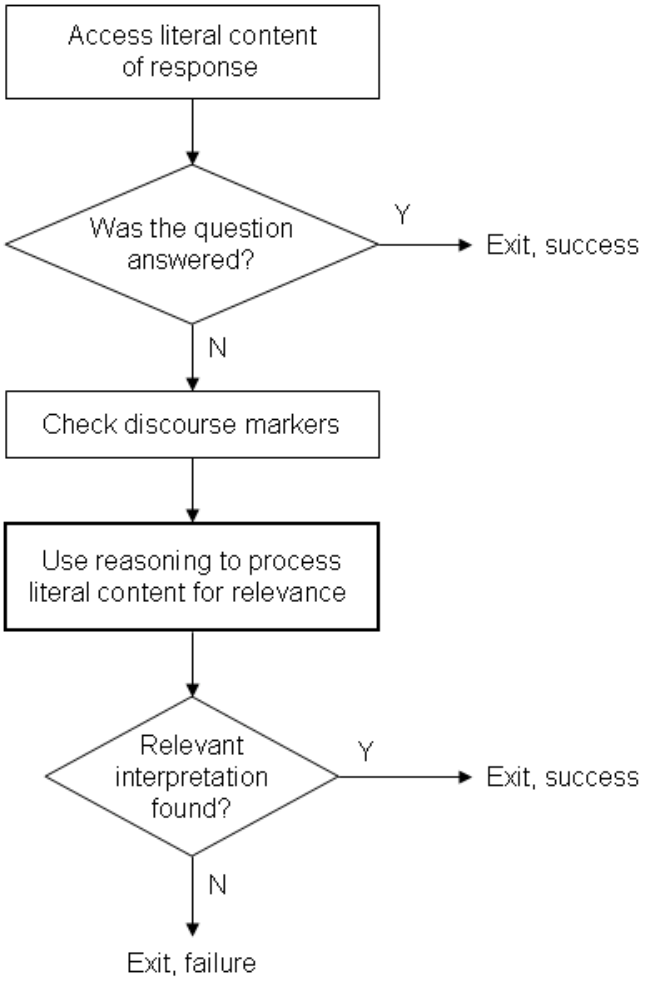

Figure 5. A flowchart for conversational implicature.

This analysis is admittedly superficial. The reasoning required to extract the answer to the question from the response is deep and interesting. Articulating and elaborating the representations, models, and algorithms enabling conversational implicature is a particularly promising topic for research. Aspects have been explored in the commonsense reasoning literature in artificial intelligence; see (Morgenstern, 2006) for an overview of some texts.

6 Conclusions

The vision of machines that exhibit humanlike levels of intelligence and autonomy continues to motivate research in numerous disciplines of science and technology. This research often looks to nature and biology for inspiration—a linkage with a productive record: developments in areas like artificial intelligence, cognitive science, and intelligent control have produced innovations such as expert systems, genetic algorithms, neural networks, fuzzy logic, and swarm optimization. In these cases and others, concepts from the natural world have been adapted for computer-based engineering systems with significant successes in practical applications.

In this article I have suggested an area of investigation that is based on a distinctive facet of human cognition. Doubtless any aspect of language or linguistics can tell us something about mind, but pragmatics can arguably claim special privilege. It is pragmatics, after all, that, more than any other area of linguistics, connects human language with human activities in the real, social world.

Although specific aspects of pragmatics have been explored for multiagent systems in the software engineering and artificial intelligence communities, the broader synergistic prospects have not been emphasized. As I have attempted to highlight, pragmatics should be considered a core discipline for multiagent systems and other applications where complex coordination among distributed entities is required. "Mind theory" is an appropriate rubric under which these broader scientific and technological objectives can be pursued.

References

Akman, V. and C. Bazzanella (2003). "The complexity of context: guest editors' introduction." *Journal of Pragmatics*, vol. 35, pp. 321-329.

Brown, P. and S. Levinson (1987). *Politeness. Some Universals of Language Use*. Cambridge, U.K.: Cambridge University Press.

Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.

Clark, V.P., P.A. Eschholz, and A.F. Rosa (eds.) (1981). *Language: Introductory Readings*, 3rd edition. New York: St. Martin's Press.

Cooke, M. (1998). "Introduction." In Habermas (1998).

Dignum, F.P.M. and M. Greaves (2000). *Issues in Agent Communication.* Springer.

Fagin, R. et al. (1995). *Reasoning About Knowledge*. Cambridge, MA: MIT Press.

Foucault, M. (1982). *This is not a Pipe.* Translated and edited by J. Harkness. Berkeley, CA: University of California Press.

Gee, J.P. (2005). *An Introduction to Discourse Analysis: Theory and Method, 2nd edition*. Routledge.

Grice, P. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Habermas, J. (1998). *On the Pragmatics of Communication*, M. Cooke (ed.). Cambridge, Mass.: MIT Press.

Hegarty, M.V. (1992). *Adjunct Extractions and Chain Configurations*. Ph.D. thesis, Massachusetts Institute of Technology.

Herring, S.C., D. Stein, and T. Virtanen (eds.) (2009). *Handbook of the Pragmatics of CMC*. Berlin: Mouton de Gruyter. (To appear)

Huang, Y. (2007). *Pragmatics*. Oxford, U.K.: Oxford University Press.

Jurafsky, D. (2004). "Pragmatics and computational linguistics." In *Handbook of Pragmatics*, L.R. Horn and G. Ward (eds.), Oxford, U.K: Blackwell.

Levinson, S.C. (1983). *Pragmatics*. Cambridge, U.K.: Cambridge University Press.

Meehan, K. (2004). *Meehan's Streak*. 3 Sept. 2004.

Morganstern, L. (2006). "Knowledge representation and commonsense reasoning: Reviews of four books." *Artificial Intelligence*, vol. 170, pp.. 1239 – 1250.

Papageorgiou, G. and D. Cofer (2003). "Coordinated control of uninhabited air vehicles with communication and processing power limitations," paper AIAA-2003-5662, AIAA Guidance, Navigation, and Control Conference and Exhibit, Austin, Texas, Aug. 11-14.

Pereira, F. (2000), "Formal grammar and information theory: together again?" *Philosophical Transactions of the Royal Society*, vol. 358, no. 1769, pp. 1239–1253.

Rahwan, I. (2005). "Guest editorial: Argumentation in multi-agent systems." *Autonomous Agents and Multi-Agent Systems*, vol. 11, no. 2, pp 115 - 125

Rajchman, J. and C. West (1985). *Post-Analytic Philosophy*. New York: Columbia University Press.

Ricker, L. and K. Rudie (2003). "Knowledge is a terrible thing to waste: Using inference in discrete-event control problems." *Proc. American Control Conference*.

Samad, T., J. Bay, and D. Godbole (2007). "Network-centric systems for military operations in urban terrain: the role of UAVs." *Proc. of the IEEE*, vol. 95, no. 1, pp. 92-107.

Searle, J. (1969). *Speech Acts*. Cambridge, U.K.: Cambridge University Press.

Zufferey, S. and A. Popescu-Belis (2004). "Towards automatic identification of discourse markers in dialogs: the case of like." *Proc. SIG-dial*, pp. 63 – 71.

Mimetic Minds as Semiotic Minds. How Hybrid Humans Make Up Distributed Cognitive Systems

Lorenzo Magnani

Department of Philosophy and Computational Philosophy Laboratory, University of Pavia, Pavia, Italy

Abstract

The main thesis of this article is that the externalization/disembodiment of mind is a significant cognitive perspective able to unveil some basic features of abduction and creative/hypothetical thinking: its success in explaining the semiotic interplay between internal and external representations (mimetic and creative) is evident. Following Peirce's semiotics, the interplay between internal and external representation can be depicted taking advantage of what I call semiotic brains. They are brains that make up a series of signs and that are engaged in making or manifesting or reacting to a series of signs. Through this semiotic activity they are at the same time engaged in "being minds" and thus in thinking intelligently. An important effect of this semiotic brain activity is a continuous process of disembodiment of mind that exhibits a new cognitive perspective on the mechanisms underlying the semiotic emergence of meaning processes. Language itself can be seen as a mediating "ultimate artifact": from this perspective the brain would merely be a pattern completing device while language would be considered an external resource/tool which is adaptively – through Darwinian evolution – fitted to the human brain helping and supporting it to enhance its cognitive capacities. I will describe the centrality to semiotic cognitive information processes of the disembodiment of mind from the point of view of the cognitive interplay between internal and external representations, both mimetic and creative, where the problem of the continuous interaction between on-line – like in the case of manipulative abduction – and off-line (for example in inner rehearsal) intelligence can properly be addressed. I consider this interplay critical in analyzing the relation between meaningful semiotic internal resources and devices and their dynamical interactions with the externalized semiotic materiality already stored in the environment. This materiality plays a specific role in the interplay due to the fact that it exhibits (and operates through) its own cognitive constraints. Hence, minds are "extended" and artificial in themselves. The example of elementary geometry will also be examined, where many external things, usually inert from the semiotic point of view, can be transformed into what I have called "epistemic mediators" (cf. [Magnani, 2001a]) that then give rise – for instance in the case of scientific reasoning – to new signs, new chances for "interpretants", and thus to new interpretations.

1 Mimetic and Creative Representations

Brains organize themselves through a semiotic activity that is reified in the external environment and then re-projected and reinterpreted through new

configurations of neural networks and chemical processes. This process, which [Mithen, 1999] called 'disembodiment of mind", can nicely account for low-level semiotic processes of meaning creation, bringing up the question of how could higher-level processes be comprised and how would they interact with lower-level ones.

1.1 External and Internal Representations

We can account for this process of disembodiment from an impressive cognitive point of view.

I maintain that representations are external and internal. We can say that

- external representations are formed by external materials that express (through reification) concepts and problems already stored in the brain or that do not have a natural home in it;
- internalized representations are internal re-projections, a kind of re-capitulations (learning), of external representations in terms of neural patterns of activation in the brain. They can sometimes be "internally" manipulated like external objects and can originate new internal re-constructed representations through the neural activity of transformation and integration.

This process explains why human beings seem to perform both computations of a connectionist type[13] such as the ones involving representations as

- (I Level) patterns of neural activation that arise as the result of the interaction between body and environment (and suitably shaped by the evolution and the individual history): pattern completion or image recognition,[14]

and computations that use representations as

- (II Level) derived combinatorial syntax and semantics dynamically shaped by the various external representations and reasoning devices found or constructed in the environment (for example geometrical diagrams); they are neurologically represented contingently as pattern of neural activations that "sometimes" tend to become stabilized

[13] Here the reference to the word "connectionism" is used on the plausible assumption that all mental representations are brain structures: verbal and the full range of sensory representations are neural structures endowed with their chemical functioning (neurotransmitters and hormones) and electrical activity (neurons fire and provide electrical inputs to other neurons). In this sense we can reconceptualize cognition neurologically: for example the solution of a problem can be seen as a process in which one neural structure representing an explanatory target generates another neural structure that constitutes a hypothesis for the solution.

[14] Clark, adopting a connectionist perspective, maintains that human brain is essentially a device for pattern-association, pattern-completion, and pattern-manipulation.

structures and to fix and so to permanently belong to the I Level above.

The I Level originates those sensations (they constitute a kind of "face" we think the world has), that provide room for the II Level to reflect the structure of the environment, and, most important, that can follow the computations suggested by these external structures. It is clear we can now conclude that the growth of the brain and especially the synaptic and dendritic growth are profoundly determined by the environment.

When the fixation is reached the patterns of neural activation no longer need a direct stimulus from the environment for their construction. In a certain sense they can be viewed as fixed internal records of external structures that can exist also in the absence of such external structures. These patterns of neural activation that constitute the I Level Representations always keep record of the experience that generated them and, thus, always carry the II Level Representation associated to them, even if in a different form, the form of memory and not the form of a vivid sensorial experience.

Now, the human agent, via neural mechanisms, can retrieve these II Level Representations and use them as internal representations or use parts of them to construct new internal representations very different from the ones stored in memory (cf. also [Gatti and Magnani, 2005]).[15]

1.2 Language as the Ultimate Artifact

The example of recent cognitive theories concerning natural language is particularly useful to illustrate the interplay between external and internal representations. Following [Clark, 1997, p. 218] language is an "ultimate artifact". In this perspective brain is just a pattern completing device (as I have illustrated introducing the I level in the previous subsection), while language is an external resource/tool which is adaptively – along the Darwinian evolution – fitted to the human brain, helping and supporting it to enhance its cognitive capacities [Wheeler, 2004]. Language is culturally passed from one generation to the next and is thus learnt again and again just through exposure to a sample of it, and then suitably generalized.[16] It is not only an important part of the

[15] The role of external representations has already been stressed in some central traditions of cognitive science and artificial intelligence, from the area of distributed and embodied cognition and of robotics [Brooks, 1991, Clark, 2003, Zhang, 1997] to the area of active vision and perception [Gibson, 1979, Thomas, 1999]. I also think this discussion about external and internal representations can be used to extend and enhance the Representational Redescription model introduced by [Karmiloff-Smith, 1992], which accounts for how these levels of representation are generated in the infant mind. [Sterelny, 2004] lists some of the most important results we can obtain thanks to external representations: they 1) ease memory burdens, 2) transform difficult cognitive problems into easier perceptual problems, 3) transform difficult perceptual problems into easier ones, 4) transform difficult learning problems into easier ones, 5) engineer workspaces to complete tasks more rapidly and reliably.

[16] On the importance of arbitrariness in natural languages cf. [Gasser, 2004]. Research acknowledging the fact that language understanding cannot be performed through the

cognitive niche built by human beings a long time ago, but it also formed a permanent artificial environment that in turn created a further selective pressure in evolution, in the co-evolutionary interplay between genes and culture: in a recent article [Clark, 2006, p. 370] himself acknowledges that "language is a self-constructed cognitive niche" consisting of structures that "combine with appropriate culturally transmitted practices to enhance problem-solving".

Exactly like hammers and PCs are fitted to the human brain and to the structure and capacities of human hands, language is a medium of communication and information and it "[…] alters the nature of the computational tasks involved in various kinds of problem solving" that affect human beings (and their brains) [Clark, 1997, p. 193]. It is said that language scaffolds cognition for the mind [Clowes and Morse, 2005]. Basically, language is for Clark a cognitive tool that facilitates thought and cognition through 1) memory augmentation, 2) environmental simplification, 3) coordination of activities through control of attention and resource allocation, 4) the activity of transcending path-dependent learning (the learning of linguistic organisms is not constrained by complicated cognitive paths that are circumvented thanks to language), 5) control loops (that act for our future behavior: for example writing plans difficult to keep in one's head), 6) data manipulation and representation [Bermúdez, 2003, p. 151]. From this perspective there is no innate domain-specific language processing system, – like for example the one maintained by [Chomsky, 1986] and language does not deeply alter the "basic modes of representation and computation" of the brain [Clark, 1997, p. 198]. [17]

The acquisition of language is a kind of reprogramming of the computational resources of the human brain in such a way that "[…] our innate pattern-completing neural architecture comes to simulate a kind of logic-like serial processing device" [Wheeler, 2004, p. 696], without a substantial modification of the brain's processing architecture. Just like diagrams can help us in many cognitive tasks and especially in mathematical reasoning language helps in various cognitive tasks, for instance as a sensory relay in human communication (and in other various simulations of basic psychic endowments), when writing in notebooks, building databases, organizing actions and plans, creating narratives and theories, etc. Moreover, language helps us in a more internal modality, such as in self-directed speech (silent, in auditory imagery, or aloud), when for example we repeat some instructions to ourselves.

---

manipulations of arbitrary symbols alone, but has to be based on the body interaction with the environment is described in [Glenberg and Kaschak, 2003, Zwaan, 2004]. In this perspective language acquisition and meaning comprehension are partly achieved through the same simulative structures used to plan and guide action [Svensson and Ziemke, 2004].

[17] On the received view on language, the so-called "language myth" cf. [Love, 2004], who discusses Clark's rejection of the idea that natural languages are codes and usefully analyzes some aspects of Saussure's and Harris' perspectives. A computational framework for studying the emergence of language and communication, which sees language as a heterogeneous set of artifacts implicated in cultural and cognitive activities is presented in [Cangelosi, 2007], also taking into account social, sensorimotor, and neural capabilities of cognitive agents.

In Clark's words, "[…] exposure to, or rehearsal [of spoken and written language, through visual, auditory, and haptic sensorial systems] […] always activates or otherwise exploits many other kinds of internal representational or cognitive resources" that are able "to provide a new kind of cognitive niche whose features and properties complement, but do not need to replicate the basic modes of operation and representation of the biological brain [Clark, 2006, pp. 370–371]. Various experiments provide evidence that the adoption of language (and symbols) would favor the de-coupling of the cognitive agent from the "immediate pull of the encountered scene" and would provide a "new realm of perceptible objects" which simplify certain kinds of attentional, reasoning, and learning tasks (ibid.). For example, in the case of the use of external linguistic tags or symbols (for example numbers), the brain is enabled – by re-presenting them when needed – to solve problems previously seen as puzzling. Studies on writing as thinking show how their coupling involves a kind of reciprocal influence, where inner and outer features have a causal influence on one another which is occurring over time [Harris, 1989, Menary, 2007]: "The restructuring of thought which writing introduces depends upon prising open a conceptual gap between sentence and utterance. […] Writing is crucial here because autoglottic inquiry presupposes the validity of unsponsored language. Utterances are automatically sponsored by those who utter them, even if they merely repeat what has been said before. Sentences by contrast, have no sponsors: they are autoglottic abstractions. The Aristotelian syllogism like the Buddhist panchakarani, presupposes writing" [Harris, 1989, p. 104].

Language would stabilize and discipline (or "anchor", Clark says) intrinsically fluid and context-sensitive modes of thought and reason:[18] one of the fruitful qualities of connectionist or artificial neural-network models is their capability and their need to be stabilized. Moreover, words act on mental off-line inner states affecting not only other internally represented words but also many other model-based and sensorimotor representations and modes, between and within humans: "Words and sentences act as artificial input signals, often (as in self-directed inner speech) entirely self-generated, that nudge fluid natural systems of encoding and representation along reliable and useful trajectories", where a "a semi-anarchic parallel organization of competing elements" (a metaphor taken from [Dennett, 1991]) is at play and explains the origin of language. These elements take control at different times in a distributed structure informed by "a wealth of options involving intermediate grades of intelligent and semi-intelligent orchestration, and of hierarchical and semi-hierarchical control" [Clark, 2006, p. 372].[19]

---

[18] I agree with [Bermúdez, 2003, p. 155] who, speaking of Clark's approach says: "His view, I suspect, is that the environmental simplification that language provides applies to a perceived environment that is already parsed into objects or objects like entities" (on prelinguistic reification in animal cognition cf. chapter five, [Magnani, Forthcoming], on the role of spatial cognition in reification cf. chapter four, same book.

[19] The never-ending problem of the role of language in "necessarily" rendering thoughts possible, and even its role in any form of conceptual thinking, or at least in the mere acquisition of thoughts or in scaffolding them and in making communication,

Quoting Clark, we have stressed above that "exposure to, or rehearsal of [spoken and written language, through visual, auditory, and haptic sensorial systems] […] always activates or otherwise exploits many other kinds of internal representational or cognitive resources" that are able "to provide a new kind of cognitive niche whose features and properties complement, but do not need to replicate the basic modes of operation and representation of the biological brain" [Clark, 2006, pp. 370–371]. The semantic approach to language can take advantage of this perspective in a more traditional framework that does not take into account the concept of cognitive niche, but is oriented towards a dynamic systems framework: [Logan, 2006, p. 153] nicely expresses an analogous consideration. A word is "a strange attractor for all the percepts associated with the concept represented by that word", and a concept can be characterized like an "artificial or virtual percept". Instead of "bringing the mountain or the percept of the mountain directly to the mind the word brings the mind to the mountain through the concept of the mountain" so accessing and capturing suitable memories. In the terms of dynamic systems approach [Logan, 2006, p. 155] "An attractor is a trajectory in phase space towards which all of the trajectories of a non-linear dynamic system are attracted. The meaning of the word [as an attractor] being uttered does not belong simply to the individual but to the community to which the individual belongs […] and emerges in the context in which it is being used". The variability of the context explains that "The attractor is a strange attractor because the meaning of a word never exactly repeats itself" for instance because of the variability of the constraints imposed by the medium at hand.[20] According to the theory of dissipative systems [Prigogine and Stengers, 1984], spoken and syntaclilized language and abstract conceptual thinking can be seen as having emerged at exactly the same time as the "bifurcation" of the brain which shifted from the concrete percept-based thinking of prelingual hominids to that of the fully fledged human species, Homo sapiens sapiens, providing an example of both punctuated equilibrium and a new order coming out of a chaotic linear system. Of course Homo sapiens sapiens vestigially retains the perceptual-oriented features of hominid brains.

In tune with this dynamic approach to semantics are the considerations made in terms of the catastrophe theory: at the level of human individuals we can hypothesize that there exists an "[…] isomorphism between the mental mechanisms which ensure the stability of a concept Q, and the physical and material mechanisms which ensure the stability of the actual object K repre-

---

consciousness and mind-reading possible, is extensively treated in [Carruthers, 2002]. Coherently with Clark's contention which I have just described, the author maintains that language is the medium for "non-domain specific thinking", which fulfils the role of integrating the outputs of a variety of domain-specific conceptual faculties (or "central-cognitive quasi-modules"). The opinion that embodiment in cognitive science undervalues concepts such as convention/norms, representation, and consciousness, as essential properties of language, is provided by [Zlatev, 2007].

[20] Details on the concept of attractor are given in chapter four and and in chapter eight of [Magnani, Forthcoming].

sented by Q" [Thom, 1980, p. 248]. Here the semantic depth of a concept is characterized by the time taken by the mental mechanisms of analysis to reduce this concept to its representative sign. The more complex the concept is, the more its stability needs regulator mechanisms, the greater is its semantic density to an actual object, as obviously happens in the case of nouns which refer to a substance: "The supreme prize is handed to animate beings, and most likely to man. An animal to live must periodically resort to a whole spectrum of activities: eating, sleeping, moving, … etc. To these fundamental physiological activities are added (for man) mental activities almost as indispensable to the meaning of being human: speaking, thinking, believing, …etc., which constitute a form of regulation which superimposes itself at the beginning and on the presupposed" [Thom, 1980, p. 248].

In the following section I will illustrate some fundamental aspects of the interplay above in the light of basic semiotic aspects of general abductive reasoning.

2 Model-Based Abduction and Semiosis beyond Peirce

I think there are two basic kinds of external representations active in the process of externalization of the mind: creative and mimetic. Mimetic external representations mirror concepts and problems that are already represented in the brain and need to be enhanced, solved, further complicated, etc. so they sometimes can creatively give rise to new concepts and meanings. In the examples I will illustrate in the following sections it will be clear how for instance a mimetic geometric representation can become creative and give rise to new meanings and ideas in the hybrid interplay between brains and suitable cognitive environments, as "cognitive niches"[21] that consequently are appropriately reshaped.

What exactly is model-based abduction from a philosophical point of view?[22] I have already said that Peirce stated that all thinking is in signs, and signs can be icons, indices, or symbols and that all inference is a form of sign activity, where the word sign includes "feeling, image, conception, and other representation" [Peirce, 1931-1958, 5.283] (for details cf. [Kruijff, 2005]), and, in Kantian words, all synthetic forms of cognition. In this light it can be maintained that a considerable part of the creative meaning processes is model-based. Moreover, a considerable part of meaning creation processes (not only in science) occurs in the middle of a relationship between brains and external objects and tools that have received cognitive and/or epistemological delegations (cf. the previous section and the following subsection).

---

[21] This expression, Clark used in the different framework of the cognitive analysis of language appears very appropriate also in this context [Pinker, 2003].

[22] Abductive cognition can be glossed as "inference to hypotheses (both explanatory and not-explanatory)". For example, when scientists decide what is the best possible explanation for a set of observed phenomena, they are performing abductive inference. I have extensively studied the problem of abductive inference and cognition in [Magnani, 2001a, Magnani, Forthcoming].

Following this Peircean perspective about inference I think it is extremely useful from a cognitive point of view to consider the concept of reasoning in a very broad way (cf. also [Brent, 2000, p. 8]). We have three cases:

1. reasoning can be fully conscious and typical of high-level worked-out ways of inferring, like in the case of scientists' and professionals' performances;
2. reasoning can be "acritical" [Peirce, 1931-1958, 5.108], which includes every day inferences in conversation and in various ordinary patterns of thinking;
3. reasoning can resort to "operations of the mind which are logically analogous to inference excepting only that they are unconscious and therefore uncontrollable and therefore not subject to logical criticism" [Peirce, 1931-1958, 5.108].

Immediately Peirce adds a note to the third case "But that makes all the difference in the world; for inference is essentially deliberate, and self-controlled. Any operation which cannot be controlled, any conclusion which is not abandoned, not merely as soon as criticism has pronounced against it, but in the very act of pronouncing that decree, is not of the nature of rational inference – is not reasoning" (ibid.).

As Colapietro clearly states [Colapietro, 2000, p. 140], it seems that for Peirce human beings semiotically involve unwitting trials and unconscious processes. Moreover, it seems clear that unconscious thought can be in some sense considered "inference", even if not rational; indeed, Peirce says, it is not reasoning. Peirce further indicates that there are in human beings multiple trains of thought at once but only a small fraction of them is conscious, nevertheless the prominence in consciousness of one train of thought is not to be interpreted an interruption of other ones.

In this Peircean perspective, which I adopt in this article, where inferential aspects of thinking dominate, there is no intuition, in an anti-Cartesian way. We know all important facts about ourselves in an inferential abductive way:

[…] we first form a definite idea of ourselves as a hypothesis to provide a place in which our errors and other people's perceptions of us can happen. Furthermore, this hypothesis is constructed from our knowledge of "outward" physical facts, such things as the sounds we speak and the bodily movements we make, that Peirce calls signs [Brent, 2000, p. 8].

Recognizing in a series of material, physical events, that they make up a series of signs, is to know the existence of a "mind" (or of a group of minds) and to be absorbed in making, manifesting, or reacting to a series of signs is to be absorbed in "being a mind". "[…] all thinking is dialogic in form" [Peirce, 1931-1958, 6.338], both at the intrasubjective[23] and intersubjective level, so that

[23] "One's thoughts are what he is 'saying to himself', that is saying to that other self that is just coming to life in the flow of time. When one reasons, it that critical self that

we see ourselves exactly as others see us, or see them exactly as they see themselves, and we see ourselves through our own speech and other interpretable behaviors, just others see us and themselves in the same way, in the commonality of the whole process [Brent, 2000, p. 10].

As I will better explain later on in the following sections, in this perspective minds are material like brains, in so far as they consist in intertwined internal and external semiotic processes: Peirce clearly anticipated the "extended mind" hypothesis maintaining that "[…] the psychologists undertake to locate various mental powers in the brain; and above all consider it as quite certain that the faculty of language resides in a certain lobe; but I believe it comes decidedly nearer the truth (though not really true) that language resides in the tongue. In my opinion it is much more true that the thoughts of a living writer are in any printed copy of his book than they are in his brain" [Peirce, 1931-1958, 7.364].

2.1 Man is an External Sign

Peirce's semiotic motto "man is an external sign" is very clear about the materiality of mind and about the fact that the conscious self is a cluster actively embodied of flowing intelligible signs:[24]

> It is sufficient to say that there is no element whatever of man's consciousness which has not something corresponding to it in the word; and the reason is obvious. It is that the word or sign which man uses is the man himself. For, as the fact that every thought is a sign, taken in conjunction with the fact that life is a train of thoughts, proves that man is a sign; so, that every thought is an external sign, proves that man is an external sign. That is to say, the man and the external sign are identical, in the same sense in which the words homo and man are identical. Thus my language is the sum total of myself; for the man is the thought [Peirce, 1931-1958, 5.314].

It is by way of signs that we ourselves are semiotic processes – for example a more or less coherent cluster of narratives. If all thinking is in signs it is not true that thoughts are in us because we are in thoughts.[25]

The systemic perspective of the catastrophe theory also stresses the role of signs in their creation of semiotic brains. In the structure of signs (as potential messages for humans) there is always a kind of dynamic instability, which

---

one is trying to persuade: and all thought whatsoever is a sign, and is mostly in the nature of language" [Peirce, 1931-1958, 5.421].

[24] Consciousness arises as "a sort of public spirit among the nerve cells" [Peirce, 1931-1958, 1.354]. The contemporary researcher on consciousness Donald fully acknowledges the "materiality of mind" [Donald, 2001, pp. 96-99].

[25] It is similar to the situation of the dreamer who is so deeply involved in the dream (we say, "she is lost in her dreams") that she does not feel she is in the dream.

renders them less probable than naturally created forms: "The imprint of a finger on the sand, the tracing of a stylet on clay, are so many naturally fragile marks of man's deliberate acts" [Thom, 1980, p. 284]. Nevertheless, on being perceived by human organisms – which consequently also "become" semiotic processes – these unstable structures return to a normal stability, and in so doing they activate semantic values, generating – mentally – the content signified by the message.

I think it is at this point clear the Peircean claim that all thinking is in signs, and signs can be icons, indices, or symbols and that, moreover, all inference is a form of sign activity, where the word sign includes feeling, image, conception, and other representation. The model-based aspects of human cognition are central, given the central role played for example by signs like images and feeling in the inferential activity "[…] man is a sign developing according to the laws of inference. […] the entire phenomenal manifestation of mind is a sign resulting from inference" [Peirce, 1931-1958, 5.312 and 5.313].

Moreover, the "person-sign" is future-conditional, that is not fully formed in the present but depending on the future destiny of the concrete semiotic activity (future thoughts and experience of the community) in which she will be involved. If Peirce maintains that when we think we appear as a sign [Peirce, 1931-1958, 5.283] and, moreover, that everything is present to us is a phenomenal manifestation of ourselves, then feelings, images, diagrams, conceptions, schemata, and other representations are phenomenal manifestations that become available for interpretations and thus are guiding our actions in a positive or negative way. They become signs when we think and interpret them. It is well-known that for Peirce all semiotic experience – and thus abduction – is also providing a guide for action. Indeed the whole function of thought is to produce habits of action.[26]

Let us summarize some basic semiotic ideas that will be of help in the further clarification of the cognitive and computational features of model-based and manipulative abduction. One of the central property of signs is their reinterpretability. This occurs in a social process where signs are referred to material objects.

As is well-known, for Peirce iconic signs are based on similarity alone, the psychoanalytic patient who thought he was masturbating when piloting the plane interpreted the cloche as an extension of his body, and an iconic sign of the penis; an ape may serve as an icon of a human.[27] Indexical signs are based

[26] On this issue cf. for example the contributions contained in a recent special issue of the journal Semiotica devoted to abduction [Queiroz and Merrell, 2005].

[27] Iconic signs preserve the relational structure governing their objects. This fact does not always have to be interpreted as a mirror-like resemblance, it can be seen as a "relation of reason" [Peirce, 1931-1958, 1.369] with the object. Rather, the structural relation would be better and more generally grasped through the mathematical notion of homomorphism – between icons and icons and their referents, as already indicated by [Barwise and Etchemendy, 1990, Stenning, 2000], and recently stressed by [Ambrosio, 2007]. A general homomorphic relationship would also be more satisfactory to account

on contiguity and dynamic relation to the object, a sign which refers to an object that it denotes by virtue of being "really affected" by that object: a certain grimace indicates the presence of pain, the rise of the column of mercury in a thermometer is a sign of a rise in temperature, indexical signs are also the footprints in the sand or a rap on the door. Consequently we can say indexical signs "point". A symbol refers to an artificial or conventional ("by virtue of a law") interpretation of a sign, the sign ∞ used by mathematicians would be an example of Peirce's notion of symbol, almost all words in language, except for occasional onomatopoeic qualities, are symbols in this sense, associated with referents in a wholly arbitrary manner.[28]

We have to immediately note that from the semiotic point of view feelings too are signs that are subject to semiotic interpretations at different levels of complexity. Peirce considered feelings elementary phenomena of mind, comprising all that is immediately present, such as pain, sadness, cheerfulness. He believes that a feeling is a state of mind possessing its own living qualities independent of any other state of the mind. Neither icon, index, nor symbol actually functions as a sign until it is interpreted and recognized in a semiotic activity and code. To make an example, it is the evolutionary kinship that makes the ape an icon of the man, in itself the similarity of the two animals does not mean anything.

Where cognition is merely possible, sign action, or semiosis, is working. Knowledge is surely inferential as well as abduction, that like any inference requires three elements: a sign, the object signified, and the interpretant. Everywhere "A signifies B to C".

There is a continuous activity of interpretation and part of this activity – as we will see – is abductive. The Peircean notion of interpretant plays the role of explaining the activity of interpretation that is occurring in semiosis. The interpretant does not necessarily refer to an actual person or mind, an actual interpreter. For instance the communication to be found in a beehive[29] where the bees are able to communicate with the others by means of signs is an example of a kind of "mindless" triadic semiosis: indeed we recognize that a sign has been interpreted not because we have observed a mental action but by observing another material sign. To make another example, the person recognizing the thermometer as a thermometer is an interpretant, as she generates in her brain a thought. In this case the process is conscious, but also unconscious or emotional interpretants are widespread. Again, a person points (index) up at the sky and his companion looks up (interpretant) to see the object of the sign. Someone else might call out "What do you see up there?" that is also another interpretant of the original sign. As noted by Brent "For Peirce, any appropriate response to a sign is acting as another sign of the object originally signified. A sunflower following the sun across the sky with

---

for the case in which the manipulation of diagrams is able to creatively convey new information and chances, like in the case of algebraic representations.

[28] On the role of symbols in mathematical abduction cf. [Heeffer, 2007].
[29] This kind of communication is studied in [Monekosso et al., 2004].

its face is also an interpretant. Peirce uses the word interpretant to stand for any such development of a given sign" [Brent, 2000, p. 12].

Semiosis is in itself a dynamic and interactive process that happens in time and presupposes the notions of environment and agents. As anything can be seen as a sign, the collection of potential signs may encompass virtually everything available within the agent, including all data gathered by its sensors. In the context of the science of complexity semiosis can be depicted as an emergent property of a semiotic system: emergent properties constitute a certain class of higher-level properties, related in a certain way to the microstructure of a class of system, that thus become able to produce, transmit, receive, compute, and interpret signs of different kinds. In this last sense they are more than simple reactive systems which in principle are not able to use something as a sign for something else [Gomes et al., 2000, Loula et al., Forthcoming]. It has to be stressed that semiotic systems are obviously materially embodied because they can be only realized through physical implementation.

Finally, an interpretant may be the thought of another person, but may as well be simply the further thought of the first person, for example in a soliloquy the succeeding thought is the interpretant of the preceding thought so that an interpretant is both the interpretant of the thought that precedes it and the object of the interpretant thought that succeeds it. In soliloquy sign, object, and interpretant are all present in the single train of thought.

Interpretants, mediating between signs[30] and their objects have three distinct levels in hierarchy: feelings, actions, and concepts or habits (that is various generalities as responses to a sign). They are the effect of a sign process. The interpretant produced by the sign can lead to a feeling (emotional interpretant), or to a muscular or mental effort, that is to a kind of action – energetic interpretant (not only outward, bodily action, but also purely inward exertions like those "mental soliloquies strutting and fretting on the stage of imagination" – [Colapietro, 2000, p. 142]. Finally, when it is related to the abstract meaning of the sign, the interpretant is called logical,[31] as a generalization requiring the use of verbal symbols. It is a further development of semiosis in the hierarchy of iconic, enactive, and symbolic communication: in short, it is "an interpreting thought", related for instance not only to the intellectual activity but also to initiate the ethical action in so far as a "modification of a person's tendencies toward action" [Peirce, 1931-1958, 5.476].

[30] It has to be noted that for Peirce no sign is so general that it cannot be amended, hence all general signs are to an extent incomplete. Consequently, a sign holds the chance of taking any particular feature previously unknown to its interpreters, many of these new features remaining inconsistent with other possibilities.

[31] The logical interpretant is not "logical" in the sense in which deductive reasoning is studied by a discipline called "logic", but rather because it attributes a further meaning to the emotion or to the mental effort that preceded it by providing a conceptual representation of that effort.

The logical interpretants are able to translate percepts, emotions, unconscious needs, and experience needs, and so to mediate their meanings to arrive to provisional stabilities. They can lead to relatively stable cognitive or intellectual habits and belief changes as self-controlled achievements like many abductive conceptual results, that Peirce considers the most advanced form of semiosis and the ultimate outcome of a sign. Indeed abduction – hypothesis – is the first step toward the formation of cognitive habits: "[…] every concept, every general proposition of the great edifice of science, first came to us as a conjecture. These ideas are the first logical interpretants of the phenomena that suggested them, and which, as suggesting them, are signs" [Peirce, 1931-1958, 5.480].[32]

Ortogonal to the classification of interpretants as emotional, energetic, and logical is the alternate classification given by Peirce: interpretants can also be immediate, dynamic, and normal. Some interpreters consider this classification a different way of expressing the first one. It is sufficient to note this classification can be useful in studying the formation of a subclass of debilitating and facilitating psychic habits [Colapietro, 2000, pp. 144–146] and, I would add, of certain reasoning devices that are used by human agents.[33] Colapietro proposes the concept of quasi-final interpretants – as related to the Peircean normal interpretants – as "[…] effective in the minimal sense that they allow the conflict-ridden organism to escape being paralyzed agent: they permit the body-ego to continue its ongoing negotiations with these conflicting demands, even if only in a precarious and even debilitating manner. In brief, they permit the body-ego to go on" [Colapietro, 2000, p. 146]. For instance there are some sedimented unsconscious reactions of this type in immediate puzzling environments – later on useless and stultifying in wider settings – but there also is the recurrent reflective and – provisionally – productive use of fallacious ways of reasoning like hasty generalizations and other arguments [Woods, 2004].

Some Peircean words about instinctual beliefs are very interesting and can be stressed to further comprehend the character of the unconscious reactions above: "[…] our indubitable beliefs refer to a somewhat primitive mode of life" [Peirce, 1931-1958, 5.511] but it seems their authority is limited to that domain "While they never become dubitable in so far as our mode or life remains that of somewhat primitive man, yet as we develop degrees of self-control unknown to that man, occasions of actions arise in relation to which

[32] Habits also appear in organic and inorganic matter: "Empirically, we find that some plants take habits. The stream of water that wears a bed for itself is forming a habit" [Peirce, 1931-1958, 5.492]. In human beings, it has to be stressed that Peirce's habit is not a purely mental, rational, or intellectual result of the semiotic process, but it is a mental representation that is always connected to the somatic and motor level, and thus constitutively embodied. On the abductive creative formation of habit as typical of self-organizing dynamic systems and processes, cf. [Gonzalez and Haselager, 2005].

[33] On the role of agency in distributed cognitive systems cf. also [Giere, 2006]. I have illustrated the role of these kinds of – more or less conscious – reasoning processes in real "human-agents", as contrasted with the abstract templates of thinking as crystallized and stabilized in the so-called "ideal logical agents" in [Magnani and Belli, 2006].

the original beliefs, if stretched to cover them, have no sufficient authority" (ibid.).

2.2 Cultured Unconscious and External/Internal Representations

In the perspective of the disembodiment of mind we can also understand how both modern human beings and externalized culture contain within them "implicit" traces of each of the previous stages of cognitive evolution. The first case of externalized distributed culture is evident: remains, buildings, manuscripts, and so on, are fragments of ancient "cognitive niches" from which we can retrieve cultural knowledge.

In the second case it can be hypothesized that much of what Freud attributes to the unconscious is truly unconscious only in the cultural sense of the word, that is formed by "things that are not expressed or are repressed at the level of culture". It has to be acknowledged that in recent cognitive science, and in the sense I have attributed to it in the previous subsections on "man as an external sign" the unconscious is a solipsistic notion, not a cultural one and concerns a part of human mind that is a priori outside the reach of consciousness, a golem, an "automaton world of instincts and zombies", like Donald eloquently says. An example is object vision: "It serves up all the richness of the three-dimensional visual world of awareness, gratis and fully formed. But we can never gain access to the mysterious region of mind that delivers such images. It lies on the other side of cognition, permanently outside the purview of consciousness" [Donald, 2001, pp. 286-287].

In the case of psychoanalysis unconscious is constructed by drives, intuitions, and representations that are shaped by the brain/culture symbiosis and interplay and so are not a priori inaccessible to awareness. It is interesting to remember that Jung has also hypothesized the existence of a collective unconscious, that is that part of individual unconscious we would share with others humans, shaped by the evolution of the above interplay, which hard-wired in it archetypes, also very ancient, that still would act in our present behavior. An example can be the "scapegoat" mechanism, typical of ancient groups and societies, where a paroxysm of violence would tend to focus on an arbitrary victim and a unanimous antipathy generated by "mimetic envy" would grow against him. The brutal elimination of the victim would reduce the appetite for violence that possessed everyone a moment before, and leaves the group suddenly appeased and calm so granting the equilibrium of the related social organization (for us repugnant, but not less useful for that societies for this reason).[34]

Like Girard [1986] says, and many researchers maintain, this kind of archaic brutal behavior, fruit of a conscious (at that time) cultural religious invention of our ancestors is still present in civilized human conduct in rich countries, it is almost always implicit and unconscious, for example in racist and mobbing behaviors. Given the fact that these kinds of behavior are widespread and

[34] On this archaic mechanism and its effect in the violence that characterizes modern societies cf. [Girard, 1977, Girard, 1986].

partially unconsciously performed it is easy to understand how they can be implicitly "learned" during infancy and then implicitly "pre-wired" by the individual in that cultured unconscious we humans collectively share with others. The result is that they are there, available in our minds/brains, to be picked up and executed – paradoxically, given the fact we are often convinced we are meant to be civil modern human beings – as archaic forms of "social" behavior.

2.3 Duties, Abductions, and Habits

The Peircean theory of "habits" can help us understand duties as imposed on ourselves from a philosophical, evolutionary, and pragmatic viewpoint, a conception I consider to be in tune with the idea of abduction that I am proposing in this article: as I contended above, all semiotic experience – and thus abduction – also provides a guide for action. For example, the logical interpretant, as a hypothetical fruit of abductive thinking requiring the use of verbal symbols is in itself "an interpreting thought", related for instance not only to the intellectual activity but also to initiate the ethical action in so far as a "modification of a person's tendencies toward action" [Peirce, 1931-1958, 5.476]. Indeed the whole function of thought is to produce habits of action, Peirce says that "[…] conduct controlled by ethical reason tends toward fixing certain habits of conduct, the nature of which […] does not depend upon any accidental circumstances, and in that sense may be said to be destined" [Peirce, 1931-1958, 5.430]. This philosophical attitude "[…] does not make the summum bonum to consist in action, but makes it to consist in that process of evolution whereby the existent comes more and more to embody those generals which […] [are] destined, which is what we strive to express in calling them reasonable" [Peirce, 1931-1958, 5.433]. This process, Peirce adds, is related to our "capacity of learning": increasing our "knowledge" will occur through time and generations, "by virtue of man's capacity of learning, and by experience continually pouring over him" [Peirce, 1931-1958, 5.402 n. 2]. It is in this process of anthroposemiosis that civilization moves toward clearer understanding and greater reason. It is in this process of anthroposemiosis, Peirce maintains, that we build highly beneficial habits that help us to acquire "ethical propensities".

Not only abductions, but also reiterations originate ethical habits as logical interpretants, and in this case the interplay between internal and external representations is still fundamental, related to the exercise of rational self-control and self-reproach guilt feelings which can be further strengthened by direct commands to oneself: "Reiterations in the inner world – fancied reiterations – if well-intensified by direct effort, produce habits, just as do reiterations in the outer world; and these habits will have power to influence actual behaviour in the outer world; especially, if each reiteration be accompanied by a peculiar strong effort that is usually likened to issuing a command to one's future self " [Peirce, 1931-1958, 5.487]. Moreover, reiterations originate habits both through imaginary and actual exertions[35] – for example repeated outward actions – but also in a hybrid way, in the suitable combination of the two).

[35] "[…] every sane person lives in a double world, the outer and the inner world, the world of percepts and the world of fancies" [Peirce, 1931-1958, 5.487].

Moreover, it may be useful to recall here what Peirce says about instinctual beliefs, we have already quoted above: "our indubitable beliefs refer to a somewhat primitive mode of life" [Peirce, 1931-1958, 5.511], but their authority is limited to such a primitive sphere. "While they never become dubitable in so far as our mode of life remains that of somewhat primitive man, yet as we develop degrees of self-control unknown to that man, occasions of action arise in relation to which the original beliefs, if stretched to cover them, have no sufficient authority." (ibid.) The problem Peirce touches on here relates to the role of emotions in ethical reasoning: I agree with him that is it is only in a constrained and educated – not primitive – way that emotions like love, compassion, and good will, for example, can guide us "morally.[36]

Anyway, a link between ethical rules and conventions and drives and instincts can be hypothesized at a more basic level, as Damasio contends in the framework of a neurological perspective: "Although such conventions and rules need be transmitted only through education and socialization, from generation to generation, I suspect that the neural representations of the wisdom they embody, and of the means to implement that wisdom, are inextricably linked to the neural representations of innate regulatory biological processes" [Damasio, 1994, p. 125]. Of course in this perspective drives and instincts have to be considered not only innate but also acquired, like in the case of educated emotions (cf. [Moorjani, 2000, p. 116]).

Natural entities exhibit different habits and various degrees, ways, and speeds with which they abandon old habits and adopt (or integrate the old ones with) new ones. Peirce says "The highest quality of mind involves greatest readiness to take habits, and a great readiness to lose them" [Peirce, 1931-1958, 6.613]. Colapietro observes that "[…] this capacity entails a measure of consciousness below that of the most acute sensations (e.g., intense pleasure or pain) but above that of our quasi-automatic reactions resulting from the unimpeded operation of effective habits in familiar circumstances" [Colapietro, 2000, p. 139]. In this sense inanimate matter is more reluctant than – for example – brains, to lose old habits and assume new ones, but it is absolutely not exempt from habit-change. We must not forget that for Peirce there is a real cosmic tendency to acquire novel dispositions that is extremely strong in well encephalized human beings.[37]

---

[36] I defended this perspective in a recent book [Magnani, 2007, chapter six].

[37] The idea of morality as "habit" – originated through the long negotiation between instinctual impulses and the inescapable pressure of cultural practices – is also supported by James Q. Wilson in a strict Darwinian framework: "I am not trying to discover 'facts' that will prove 'values'; I am endeavoring to uncover the evolutionary, developmental, and cultural origins of our moral habits and our moral sense." He also argues for a biological counterpart that would facilitate the formation of these habits. He continues "But in discovering these origins, I suspect we will encounter uniformities; and by revealing uniformities, I think that we can better appreciate what is general, non-arbitrary, and emotionally compelling about human nature" [Wilson, 1993, p. 26].

The previous two sections have introduced to both the interplay between internal and external representations and to some basic semiotic aspects of abductive reasoning: the following sections will take advantage of this background. I will describe how the interplay of signs, objects, and interpretants is working in important aspects of abductive reasoning. Of course model-based cognition acquires its peculiar creative relevance when embedded in abductive processes. I will show some examples of model-based inferences. It is well known the importance Peirce ascribed to diagrammatic thinking (a kind of iconic thinking), as shown by his discovery of the powerful system of predicate logic based on diagrams or "existential graphs". As I have already stressed, Peirce considers inferential any cognitive activity whatever, not only conscious abstract thought; he also includes perceptual knowledge and subconscious cognitive activity. For instance in subconscious mental activities visual representations play an immediate role [Queiroz and Merrell, 2005].

3 Constructing Meaning through Mimetic and Creative External Objects

3.1 Constructing Meaning through Manipulative Abduction

Manipulative abduction occurs when many external things, usually inert from the semiotic point of view, can be transformed into what I have called, in the case of scientific reasoning, "epistemic mediators" [Magnani, 2001a] that give rise to new signs, new chances for interpretants, and new interpretations.

We can cognitively account for this process of externalization[38] taking advantage of the concept of manipulative abduction (cf. [Magnani, 2001a, chapter three].) It happens when we are thinking through doing and not only, in a pragmatic sense, about doing. It happens, for instance, when we are creating geometry constructing and manipulating an external suitably realized icon like a triangle looking for new meaningful features of it, like in the case given by Kant in the "Transcendental Doctrine of Method" (cf. [Magnani, 2001b]). It refers to an extra–theoretical behavior that aims at creating communicable accounts of new experiences to integrate them into previously existing systems of experimental and linguistic (semantic) practices. [Gooding, 1990] refers to this kind of concrete manipulative reasoning when he illustrates the role in science of the so-called "construals" that embody tacit inferences in procedures that are often apparatus and machine based. I have described them in [Magnani, 2001a, chapter three].

It is difficult to establish a list of invariant behaviors that are able to describe manipulative abduction in science.[39] Even if abduction operates, like Peirce says, according to the aesthetic process of musement: "a certain agreeable occupation of the mind" [Peirce, 1992-1998, II, p. 436] which must follow "the

[38] I have illustrated above in this article a significant contribution to the comprehension of this process in terms of the so–called "disembodiment of the mind".

[39] A list is provided in [Magnani, 2001a, chapter three].

very law of liberty" [Peirce, 1931-1958, 6.458], as I have already illustrated above, the expert manipulation of objects in a highly semiotically constrained experimental environment certainly implies the application of old and new templates of behavior that exhibit some regularities.[40] The activity of building construals is highly conjectural and not necessarily or immediately explanatory: these templates are hypotheses of behavior (creative or already cognitively present in the scientist's mind-body system, and sometimes already applied) that abductively enable a kind of epistemic "doing". Hence, some templates of action and manipulation can be selected in the set of the ones available and pre-stored, others have to be created for the first time to perform the most interesting creative cognitive accomplishments of manipulative abduction.

3.2 Manipulating Meanings through External Semiotic Anchors

If the structures of the environment play such an important role in shaping our semiotic representations and, hence, our cognitive processes, we can expect that physical manipulations of the environment receive a cognitive relevance.

Several authors have pointed out the role that physical actions can have at a cognitive level. In this sense Kirsh and Maglio [1994] distinguish actions into two categories, namely pragmatic actions and epistemic actions. Pragmatic actions are the actions that an agent performs in the environment in order to bring itself physically closer to a goal. In this case the action modifies the environment so that the latter acquires a configuration that helps the agent to reach a goal which is understood as physical, that is, as a desired state of affairs. Epistemic actions are the actions that an agent performs in a semiotic environment in order to discharge the mind of a cognitive load or to extract information that is hidden or that would be very hard to obtain only by internal computation.

In this section I want to focus specifically on the relationship that can exist between manipulations of the environment and representations. In particular, I want to examine whether external manipulations can be considered as means to construct external representations.

If a manipulative action performed upon the environment is devoted to create a configuration of signs that carries relevant information, that action will well be able to be considered as a cognitive semiotic process and the configuration of elements it creates will well be able to be considered an external representation. In this case, we can really speak of an embodied cognitive process in which an action constructs an external representation by means of manipulation. We define cognitive manipulating as any manipulation of the environ-

[40] It is simple to explain why abduction works according to musement. This is the general attitude we adopt when we are wondering about the beauty and the harmony of universes and their connections [Peirce, 1992-1998, II, p. 436]. I think that beauty plays a kind of exciting emotional role in abductive reasoning, very similar to the one played by anomalies and surprise. Cf. also [Maddalena, 2005, p. 247].

ment devoted to construct external configurations that can count as representations.

An example of cognitive manipulating is the diagrammatic demonstration illustrated in Figure 1, taken from the field of elementary geometry. In this case a simple manipulation of the triangle in Figure 1(a) gives rise to an external configuration – Figure 1(b) – that carries relevant semiotic information about the internal angles of a triangle "anchoring" new meanings.

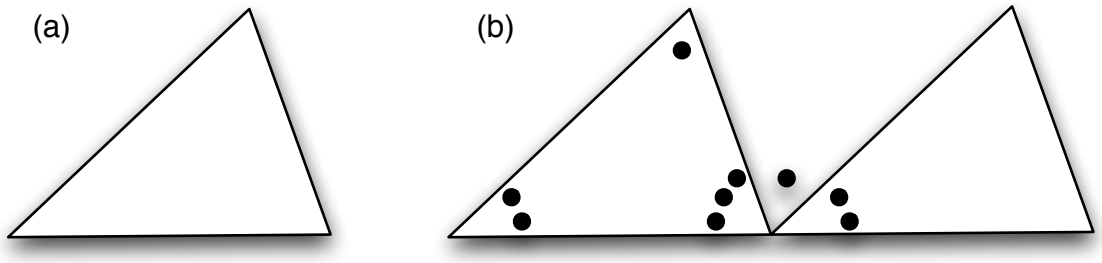

Figure 1: Diagrammatic demonstration that the sum of the internal angles of any triangle is 180°. (a) Triangle. (b) Diagrammatic manipulation/construction.

The entire process through which an agent arrives at a physical action that can count as cognitive manipulating can be understood by means of the concept of manipulative abduction. In this perspective manipulative abduction is a specific case of cognitive manipulating in which an agent, when faced with an external situation from which it is hard or impossible to extract new meaningful features of an object, selects or creates an action that structures the environment in such a way that it gives information which would be otherwise unavailable and which is used specifically to infer explanatory hypotheses.

In this way the semiotic result is achieved on external representations used in lieu of the internal ones. Here action plays an epistemic and not merely performatory role, for example relevant to abductive reasoning. The process also illustrates a synthesis between a constructive procedure of motor origin (the putting the new segment end to end parallel to one side in the externally represented given triangle), followed by a sensory procedure, "visual" (calculation of the sizes of the now clearly – externally – "seen" angles).[41]

41 "The essential step in the construction of the Euclidean space, has been the possibility of the *division* of a motor field; and here we come up against an evident physiological impossibility. Greek geometry resolved this problem of the division of a segment into equal segments by the discovery of Thales' Theorem: equidistant parallel lines cut two secants in proportional segments" [Thom, 1980, p. 134]. Furthermore, following Thom, I think this ancient Greek geometry example already represents the quintessence of the scientific approach, that is "[…] replacing a non-local operation (for example, taking the intersection of two lines in a plane) by a verbal description the formal analysis of which became the demonstration that it was virtually autonomous, that is, able to be rendered independent of the non-local intuitive approaches which described it" [Thom, 1980, p. 135]. The use of literary symbols, which are empty of sense, together with the axiomatic approach realizes the localization of the non-local intuition of the plane (and of space).

3.3 Geometrical Construction is a Kind of Manipulative Abduction

Let's quote Peirce's passage about mathematical constructions. Peirce says that mathematical and geometrical reasoning "[…] consists in constructing a diagram according to a general precept, in observing certain relations between parts of that diagram not explicitly required by the precept, showing that these relations will hold for all such diagrams, and in formulating this conclusion in general terms. All valid necessary reasoning is in fact thus diagrammatic" [Peirce, 1931-1958, 1.54]. This passage clearly refers to a situation like the one I have illustrated in the previous subsection. This kind of reasoning is also called by Peirce "theorematic" and it is a kind of "deduction" necessary to derive significant theorems (Necessary Deduction]: "[…] is one which, having represented the conditions of the conclusion in a diagram, performs an ingenious experiment upon the diagram, and by observation of the diagram, so modified, ascertains the truth of the conclusion" [Peirce, 1931-1958, 2.267]. The experiment is performed with the help of "[…] imagination upon the image of the premiss in order from the result of such experiment to make corollarial deductions to the truth of the conclusion" [Peirce, 1976, IV, p. 38]. The "corollarial" reasoning is mechanical (Peirce thinks it can be performed by a "logical machine") and not creative, "A Corollarial Deduction is one which represents the condition of the conclusion in a diagram and finds from the observation of this diagram, as it is, the truth of the conclusion" [Peirce, 1931-1958, 2.267] (cf. also [Hoffmann, 1999]).

In summary, the point of theorematic reasoning is the transformation of the problem by establishing an unnoticed point of view to get interesting – and possibly new – insights. The demonstrations of "new" theorems in mathematics are examples of theorematic deduction.

Not dissimilarly Kant says that in geometrical construction of external diagrams "[…] I must not restrict my attention to what I am actually thinking in my concept of a triangle (this is nothing more than the mere definition); I must pass beyond it to properties which are not contained in this concept, but yet belong to it" [Kant, 1929, A718-B746, p. 580].

Theorematic deduction can be easily interpreted in terms of manipulative abduction. We have seen that manipulative abduction is a kind of abduction, mainly model-based, that exploits external models endowed with delegated (and often implicit) cognitive and semiotic roles and attributes:

1. the model (diagram) is external and the strategy that organizes the manipulations is unknown a priori;
2. the result achieved is new (if we, for instance, refer to the constructions of the first creators of geometry), and adds properties not contained before in the concept (the Kantian to "pass beyond" or "advance beyond" the given concept [Kant, 1929, A154-B193/194, p. 192]). [42]

[42] Of course in the case we are using diagrams to demonstrate already known theorems (for instance in didactic settings), the strategy of manipulations is not necessary unknown and the result is not new, like in the Peircean case of corollarial deduction.

Iconicity in theorematic reasoning is central. Peirce, analogously to Kant, maintains that "[…] philosophical reasoning is reasoning with words; while theorematic reasoning, or mathematical reasoning is reasoning with specially constructed schemata" [Peirce, 1931-1958, 4.233]; moreover, he uses diagrammatic and schematic as synonyms, thus relating his considerations to the Kantian tradition where schemata mediate between intellect and phenomena.[43] The following is the famous passage in the Critique of Pure Reason ("Transcendental Doctrine of Method"):

> Suppose a philosopher be given the concept of a triangle and he be left to find out, in his own way, what relation the sum of its angles bears to a right angle. He has nothing but the concept of a figure enclosed by three straight lines, and possessing three angles. However long he meditates on this concept, he will never produce anything new. He can analyse and clarify the concept of a straight line or of an angle or of the number three, but he can never arrive at any properties not already contained in these concepts. Now let the geometrician take up these questions. He at once begins by constructing a triangle. Since he knows that the sum of two right angles is exactly equal to the sum of all the adjacent angles which can be constructed from a single point on a straight line, he prolongs one side of his triangle and obtains two adjacent angles, which together are equal to two right angles. He then divides the external angle by drawing a line parallel to the opposite side of the triangle, and observes that he has thus obtained an external adjacent angle which is equal to an internal angle – and so on.[44] In this fashion, through a chain of inferences guided throughout by intuition, he arrives at a fully evident and universally valid solution of the problem [Kant, 1929, A716-B744, pp. 578-579].

We can depict the situation of the philosopher described by Kant at the beginning of the previous passage taking advantage of some ideas coming from the catastrophe theory. As a human being who is not able to produce anything new relating to the angles of the triangle, the philosopher experiences a feeling of frustration (just like the Kölher's monkey which cannot keep the banana out of reach). The bad affective experience "deforms" the organism's regulatory structure by complicating it and the cognitive process stops altogether. The geometer instead "at once constructs the triangle", that is, he makes an external representation of a triangle and acts on it with suitable manipulations. Thom thinks that this action is triggered by a "sleeping phase" generated by possible previous frustrations which then change the cognitive status of the geometer's available and correct internal idea of triangle (like the philosopher, he "has nothing but the concept of a figure enclosed by three straight lines, and possessing three angles", but his action is triggered by a sleeping phase). Here the idea of the triangle is no longer the occasion for

[43] Schematism, a fruit of the imagination is, according to Kant, "[…] an art concealed in the depths of the human soul, whose real modes of activity nature is hardly likely ever to allow us to discover, and to have open to our gaze" [Kant, 1929, A141-B181, p. 183].

[44] It is Euclid's Proposition XXXII, Book I, cf. above Figure 1.

"meditation", "analysis" and "clarification" of the "concepts" at play, like in the case of the "philosopher". Here the inner concept of triangle – symbolized as insufficient – is amplified and transformed thanks to the sleeping phase (a kind of Kantian imagination active through schematization) in a prosthetic triangle to be put outside, in some external support. The instrument (here an external diagram) becomes the extension of an organ:

What is strictly speaking the end […] [in our case, to find the sum of the internal angles of a triangle] must be set aside in order to concentrate on the means of getting there. Thus the problem arises, a sort of vague notion altogether suggested by the state of privation. […] As a science, heuristics does not exist. There is only one possible explanation: the affective trauma of privation leads to a folding of the regulation figure. But is it is to be stabilized, there must be some exterior form to hold on to. So this anchorage problem remains whole and the above considerations provide no answer as to why the folding is stabilized in certain animals or certain human beings whilst in others (the majority of cases, needless to say!) it fails [Thom, 1988, pp. 63–64].[45]

As we have already said, for Peirce the whole mathematics consists in building diagrams that are "[…] (continuous in geometry and arrays of repeated signs/letters in algebra) according to general precepts and then [in] observing in the parts of these diagrams relations not explicitly required in the precepts" [Peirce, 1931-1958, 1.54]. Peirce contends that this diagrammatic nature is not clear if we only consider syllogistic reasoning "which may be produced by a machine" but becomes extremely clear in the case of the "logic of relatives, where any premise whatever will yield an endless series of conclusions, and attention has to be directed to the particular kind of conclusion desired" [Peirce, 1987, pp. 11–23].

In ordinary geometrical proofs auxiliary constructions are present in terms of "conveniently chosen" figures and diagrams where strategic moves are important aspects of deduction. The system of reasoning exhibits a dual character: deductive and "hypothetical". Also in other – for example logical – deductive frameworks there is room for strategic moves which play a fundamental role in the generations of proofs. These strategic moves correspond to particular forms of abductive reasoning.

We know that the kind of reasoned inference that is involved in creative abduction goes beyond the mere relationship that there is between premises and conclusions in valid deductions, where the truth of the premises guarantees the truth of the conclusions, but also beyond the relationship that there is in probabilistic reasoning, which renders the conclusion just more or less probable. On the contrary, we have to see creative abduction as formed by the application of heuristic procedures that involve all kinds of good and bad inferential actions, and not only the mechanical application of rules. It is only

[45] A full analysis of the Kölher's chimpanzee getting hold of a stick to knock a banana hanging out of reach in terms of the mathematical models of the perception and the capture catastrophes is given in [Thom, 1988, pp. 62–64]. On the role of emotions, for example frustration, in scientific discovery cf. [Thagard, 2002].

by means of these heuristic procedures that the acquisition of new truths is guaranteed. Also Peirce's mature view illustrated above on creative abduction as a kind of inference seems to stress the strategic component of reasoning.

Many researchers in the field of philosophy, logic, and cognitive science have sustained that deductive reasoning also consists in the employment of logical rules in a heuristic manner, even maintaining the truth preserving character: the application of the rules is organized in a way that is able to recommend a particular course of actions instead of another one. Moreover, very often the heuristic procedures of deductive reasoning are performed by means of model-based abductive steps where iconicity is central. We have seen that the most common example of manipulative creative abduction is the usual experience people have of solving problems in geometry in a model-based way trying to devise proofs using diagrams and illustrations: of course the attribute of creativity we give to abduction in this case does not mean that it has never been performed before by anyone or that it is original in the history of some knowledge (they actually are cases of Peircean corollarial deduction).[46]

3.4 External Diagrammatization and Iconic Brain Co-Evolution

Following our previous considerations it would seem that diagrams can be fruitfully seen from a semiotic perspective as external representations expressed through icons and symbols, aimed at simply "mimicking" various humans' internal images. However, we have seen that they can also play the role of creative representations human beings externalize and manipulate not just to mirror the internal ways of thinking of human agents but to find room for concepts and new ways of inferring which cannot – at a certain time – be found internally "in the mind".

In summary, we can say that

- diagrams as external iconic (often enriched by symbols) representations are formed by external materials that either mimic (through reification) concepts and problems already internally present in the brain or creatively express concepts and problems that do not have a semiotic "natural home" in the brain;
- subsequent internalized diagrammatic representations are internal reprojections, a kind of recapitulations (learning), in terms of neural patterns of activation in the brain ("thoughts", in Peircean sense), of external diagrammatic representations. In some simple cases complex diagrammatic transformations – can be "internally" manipulated like external objects and can further originate new internal reconstructed representations through the neural activity of transformation and integration.

[46] We have to say that model-based abductions – which for example exploit iconicity – also operate in deductive reasoning. On the role of strategies and heuristics in deductive proofs cf. [Magnani, Forthcoming, chapter seven].

I have already stressed that this process explains – from a cognitive point of view – why human agents seem to perform both computations of a connectionist type such as the ones involving representations as

- (I Level) patterns of neural activation that arise as the result of the interaction (also presemiotic) between body and environment (and suitably shaped by the evolution and the individual history): pattern completion or image recognition,

and computations that use representations as

- (II Level) derived combinatorial syntax and semantics dynamically shaped by the various artificial external representations and reasoning devices found or constructed in the semiotic environment (for example iconic representations); they are – more or less completely – neurologically represented contingently as patterns of neural activations that "sometimes" tend to become stabilized meaning structures and to fix and so to permanently belong to the I Level above.

It is in this sense we can say the "System of Diagrammatization", in Peircean words, allows for a self-controlled process of thought in the fixation of originally vague beliefs: as a system of learning, it is a process that leads from "absolutely undefined and unlimited possibility" [Peirce, 1931-1958, 6.217] to a fixation of belief and "by means of which any course of thought can be represented with exactitude" [Peirce, 1931-1958, 4.530]. Moreover, it is a system which could also improve other areas of science, beyond mathematics, like logic, it "[…] greatly facilitates the solution of problems of Logic. […] If logicians would only embrace this method, we should no longer see attempts to base their science on the fragile foundations of metaphysics or a psychology not based on logical theory" [Peirce, 1931-1958, 4.571].

As already stressed the I Level originates those sensations (they constitute a kind of "face" we think the world has), that provide room for the II Level to reflect the structure of the environment, and, most important, that can follow the computations suggested by the iconic external structures available. It is clear that in this case we can conclude that the growth of the brain and especially the synaptic and dendritic growth are profoundly determined by the environment. Consequently we can hypothesize a form of co-evolution between what we can call the iconic brain and the development of the external diagrammatic systems. Brains build iconic signs as diagrams in the external environment learning from them new meanings through interpretation (both at the spatial and sentential level) after having manipulated them.

When the fixation is reached – imagine for instance the example above, that fixes the sum of the internal angles of the triangle – the pattern of neural activation no longer needs a direct stimulus from the external spatial representation in the environment for its construction and can activate a "final logical interpretant", in Peircean terms. It can be neurologically viewed as a fixed internal record of an external structure (a fixed belief in Peircean terms) that can exist also in the absence of such external structure. The pattern of neural

activation that constitutes the I Level Representation has kept record of the experience that generated it and, thus, carries the II Level Representation associated to it, even if in a different form, the form of semiotic memory and not the form of the vivid sensorial experience for example of the triangular construction drawn externally, over there, for instance in a blackboard. Now, the human agent, via neural mechanisms, can retrieve that II Level Representation and use it as an internal representation (and can use it to construct new internal representations less complicated than the ones previously available and stored in memory).

At this point we can easily understand the particular mimetic and creative role played by external diagrammatic representations in mathematics:

1. some concepts, meanings, and "ways of [geometrical] inferring" performed by the biological human agents appear hidden and more or less tacit and can be rendered explicit by building external diagrammatic mimetic models and structures; later on the agent will be able to pick up and use what was suggested by the constraints and features intrinsic and immanent to their external semiotic materiality and the relative established conventionality: artificial languages, proofs, new figures, examples, etc.;
2. some concepts, meanings, and "new ways of inferring" can be discovered only through a problem solving process occurring in a distributed interplay between brains and external representations. I have called this process externalization (or disembodiment) of the mind: the representations are mediators of results obtained and allow human beings

    - to re-represent in their brains new concepts, meanings, and reasoning devices picked up outside, externally, previously absent at the internal level and thus impossible: first, a kind of alienation is performed, second, a recapitulation is accomplished at the neuronal level by re-representing internally that which has been "discovered" outside. We perform cognitive geometric operations on the structure of data that synaptic patterns have "picked up" in an analogical way from the explicit diagrammatic representations in the environment;
    - to re-represent in their brains portions of concepts, meanings, and reasoning devices which, insofar as explicit, can facilitate inferences that previously involved a very great effort because of human brain's limited capacity. In this case the thinking performance is not completely processed internally but in a hybrid interplay between internal (both tacit and explicit) and external iconic representations. In some cases this interaction is between the internal level and a computational tool which in turn can exploit iconic/geometrical representations to perform inferences.

An evolved mind is unlikely to have a natural home for complicated concepts like the ones geometry introduced, as such concepts do not exist in a definite way in the natural (not artificially manipulated) world: so whereas evolved minds could construct spatial frameworks and perform some trivial spatial inferences in a more or less tacit way by exploiting modules shaped by natural

selection, how could one think exploiting explicit complicated geometrical concepts without having picked them up outside, after having produced them?

Let me repeat that a mind consisting of different separated implicit templates of thinking and modes of inferences exemplified in various exemplars expressed through natural language cannot come up with certain mathematical and geometrical entities without the help of the external representations. The only way is to extend the mind into the material world, exploiting paper, blackboards, symbols, artificial languages, and other various semiotic tools, to provide semiotic anchors[47] for finding ways of inferring that have no natural home within the mind, that is for finding ways of inferring and concepts that take us beyond those that natural selection and previous cultural training could enable us to possess at a certain moment.

Hence, we can hypothesize – for example – that many valid spatial reasoning habits which in human agents are performed internally have a deep origin in the past experience lived in the interplay with iconic systems at first represented in the environment. As I have just illustrated other recorded thinking habits only partially occur internally because they are hybridized with the exploitation of already available or suitably constructed external diagrammatic artifacts.

4 Mimetic Minds as Semiotic Minds

I contend that there are external representations that are representations of other external representations. In some cases they carry new scientific knowledge. To make an example, Hilbert's Grundlagen der Geometrie is a "formal" representation of the geometrical problem solving through diagrams: in Hilbertian systems solutions of problems become proofs of theorems in terms of an axiomatic model. In turn a calculator is able to re-represent (through an artifact) (and to perform) those geometrical proofs with diagrams already performed by human beings with pencil and paper. In this case we have representations that mimic particular cognitive performances that we usually attribute to our minds (cf. the first sections of this article).

We have seen that our brains delegate cognitive (and epistemic) roles to externalities and then tend to "adopt" and recapitulate what they have checked occurring outside, over there, after having manipulated – often with creative results – the external invented structured model. A simple example: it is relatively neurologically easy to perform an addition of numbers by depicting in our mind – thanks to that brain device that is called visual buffer – the images of that addition thought as it occurs concretely, with paper and pencil, taking advantage of external materials. We have said that mind representations are also over there, in the environment, where mind has objectified itself in various semiotic structures that mimic and enhance its internal representations.

[47] [Enfield, 2005, Callagher, 2005] point out the role of the body itself as and "anchoring" of cognitive processes, for instance in the case of human gestures linked to the expression of meanings.

Turing adds a new structure to this list of external objectified devices: an abstract tool, the (Universal) Logical Computing Machine (LCM), endowed with powerful mimetic properties. We have concluded the subsection 1.1 remarking that the creative "mind" is in itself extended and, so to say, both internal and external: the mind is semiotic because transcends the boundary of the individual and includes parts of that individual's environment, and thus constitutively artificial. Turing's LCM, which is an externalized device, is able to mimic human cognitive operations that occur in that interplay between the internal mind and the external one. Indeed Turing already in 1950 maintains that, taking advantage of the existence of the LCM, "Digital computers […] can be constructed, and indeed have been constructed, and […] they can in fact mimic the actions of a human computer very closely" [Turing, 1950, p. 435].

In the light of my perspective both (Universal) Logical Computing Machine (LCM) (the theoretical artifact) and (Universal) Practical Computing Machine (PCM) (the practical artifact) are mimetic minds because they are able to mimic the mind in a kind of universal way (wonderfully continuing the activity of disembodiment of minds and of semiotic delegations to the external materiality our ancestors rudimentary started). LCM and PCM are able to re-represent and perform in a very powerful way plenty of cognitive skills of human beings. Universal Turing Machines are discrete-state machines, DMS, "with a Laplacian behavior" [Longo, 2002, Lassègue, 1998, Lassègue, 1999]: "[…] it is always possible to predict all future states") and they are equivalent to all formalisms for computability (what is thinkable is calculable and mechanizable), and because universal they are able to simulate – that is to mimic – any human cognitive function, that is what is usually called mind. A natural consequence of this perspective is that Universal Turing machines do not represent (against classical AI and modern cognitivist computationalism) a "knowledge" of the mind and of human intelligence. Turing is perfectly aware of the fact that brain is not a DSM, but as he says, a "continuous" system, where instead a mathematical modeling can guarantee a satisfactory scientific intelligibility (cf. his studies on non-Laplacian mathematical models of morphogenesis).

We have seen that our brains delegate meaningful semiotic (and of course cognitive and epistemic) roles to externalities and then tend to "adopt" what they have checked occurring outside, over there, in the external invented structured and model. And a large part of meaning formation takes advantage of the exploitation of external representations and mediators. Our view about the disembodiment of mind certainly involves that the Mind/Body dualist view is less credible as well as Cartesian computationalism. Also the view that mind is computational independently of the physical (functionalism) is jeopardized. In my perspective on human cognition in terms of mimetic minds we no longer need Descartes dualism: we only have semiotic brains that make up large, integrated, material cognitive systems like for example LCMs and PCMs. These are new independent semiotic agencies that constitute real artificial minds aiming at "universally" imitating human cognition. In this perspective what we usually call mind simply consists in the union of both the chang-

ing neural configurations of brains together with those large, integrated, and material cognitive systems the brains themselves are continuously building in an infinite semiotic process.

Minds are material like brains, in so far as they take advantage of intertwined internal and external semiotic processes. It seems to me at this point we can better and more deeply understand Peirce's semiotic motto "man is an external sign" in the passage we have completely quoted above in section 2.1: "[…] as the fact that every thought is a sign, taken in conjunction with the fact that life is a train of thoughts, proves that man is a sign; so, that every thought is an external sign, proves that man is an external sign" [Peirce, 1931-1958, 5.324]. The only problem seems "how meat knows": we can reverse the Cartesian motto and say "sum ergo cogito".

We have seen that our brains delegate meaningful cognitive (and epistemic) roles to externalities and then tend to "adopt" what they have checked occurring outside, over there, in the external invented structures and models. And a large part of meaning formation takes advantage of the exploitation of external representations and mediators. We have said that PCMs can be considered mimetic minds (they are ideal "practical" – in Turing's sense – agents): what is in turn the cognitive status of "logical agents" from the point of view of their demonstrative aspect?

5 Conclusion

The main thesis of this article is that the externalization/disembodiment of mind is a significant cognitive perspective able to unveil some basic features of creative abductive thinking and its cognitive and computational problems. Its fruitfulness in explaining the semiotic interplay between internal and external levels of cognition is evident. I maintained that various aspects of creative meaning formation could take advantage of the research on this interplay: for instance study of external mediators can provide a better understanding of the processes of explanation and discovery in science and in some areas of artificial intelligence related to mechanizing discovery processes.

We have seen how the cognitive referral to the central role of the relation between meaningful behavior and dynamical interactions with the environment becomes critical to the problem of modeling up-to-date artificial systems devoted to performing creative and explanatory tasks: I contend that the epistemological role of those artifacts, such as computers, which I called "mimetic minds", can be further studied, taking advantage of research on hypercomputation. The imminent construction of new types of universal "abstract" and "practical" machines will constitute important and interesting new "mimetic minds" externalized and available over there, in the environment, as sources of the mechanisms underlying the emergence of new meaning processes. They will provide new tools for creating meaning in classical areas like analogical, visual, and spatial inferences, both in science and everyday situations, thereby extending epistemological and psychological theory.

Finally, the externalization/disembodiment of mind is a significant cognitive perspective able to unveil some aspects of creative meaning formation central to psychoanalytic research and therapy. I have highlighted some Junghian analysis regarding the role of certain external artifacts where the mobility and disposability of psychic energy are seen as the secret of cultural development both at the collective and individual level. I have contended that symbols, in a psychoanalytic sense, are artifacts/tools that maximize abducibility, because they maximize the recoverability of something hidden, not yet grasped by consciousness.

References

[Ambrosio, 2007] C. Ambrosio. Iconicity and Homomorphism in Picasso's Guernica: a Study of Creativity Across the Boundaries. PhD thesis, University of London, 2007.

[Barwise and Etchemendy, 1990] J. Barwise and J. Etchemendy. Visual information and valid reasoning visualization in mathematics. In W. Zinnermann, editor, Mathematical Association of America, Washington, DC, 1990.

[Bermúdez, 2003] J. L. Bermúdez. Thinking without Words. Oxford University Press, Oxford, 2003.

[Brent, 2000] J. Brent. A brief introduction to the life and thought of Charles Sanders Peirce. In J. Muller and J. Brent, editors, Peirce, Semiosis, and Psychoanalysis, pages 1–14. John Hopkins, Baltimore and London, 2000.

[Brooks, 1991] R. A. Brooks. Intelligence without representation. Artificial Intelligence, 47:139–159, 1991.

[Callagher, 2005] S. Callagher, editor. How the Body Shapes the Mind. Oxford University Press, Oxford, 2005.

[Cangelosi, 2007] A. Cangelosi. Adaptive agent modeling of distributed language: investigations on the effects of cultural variation and internal action representations. Language Sciences, 29:633–649, 2007.

[Carruthers, 2002] P. Carruthers. The cognitive function of language. Behavioral and Brain Sciences, 25(6):657–674, 2002.

[Chomsky, 1986] N. Chomsky. Knowledge of Language. Its Nature, Origins, and Use. Praeger, New York, 1986.

[Clark, 1997] A. Clark. Being There: Putting Brain, Body, and World Together Again. The MIT Press, Cambridge, MA, 1997.

[Clark, 2003] A. Clark. Natural-Born Cyborgs. Minds, Technologies, and the Future of Human Intelligence. Oxford University Press, Oxford, 2003.

[Clark, 2006] A. Clark. Language, embodiment, and the cognitive niche. Trends in Cognitive Science, 10(8):370–374, 2006.

[Clowes and Morse, 2005] R. W. Clowes and A. Morse. Scaffolding cognition with words. In L. Berthouze, F. Kaplan, H. Kozima, H Yano, J. Konczak, G. Metta, J. Nadel, G. Sandini, G. Stojianov, and C. Balkenius, editors, Proceedings of the Fifth International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems, pages 101–105, Nara, 2005.

[Colapietro, 2000] V. Colapietro. Futher consequences of a singular capacity. In J. Muller and J. Brent, editors, Peirce, Semiosis, and Psychoanalysis, pages 136–58. John Hopkins, Baltimore and London, 2000.

[Damasio, 1994] A. R. Damasio. Descartes' Error. Putnam, New York, 1994.

[Dennett, 1991] D. Dennett. Consciousness Explained. Little, Brown, and Company, New York, 1991.

[Donald, 2001] M. Donald. A Mind So Rare. The Evolution of Human Consciousness. Norton, London, 2001.

[Enfield, 2005] N. Enfield. The body as a cognitive artifact in kinship representations: hand gestures diagrams by speakers of lao. Current Anthropology, 46:51–81, 2005.

[Gasser, 2004] M. Gasser. The origins of arbitrariness in language. In T. Regier K. D. Forbus, D. Gentner, editor, CogSci 2004, XXVI Annual Conference of the Cognitive Science Society, Chicago, IL, 2004. CD-Rom.

[Gatti and Magnani, 2005] A. Gatti and L. Magnani. On the representational role of the environment and on the cognitive nature of manipulations. In L. Magnani and R. Dossena, editors, Computing, Philosophy and Cognition, pages 227–242. College Pubblications, London, 2005.

[Gibson, 1979] J. J. Gibson. The Ecological Approach to Visual Perception. Houghton Mifflin, Boston, MA, 1979.

[Giere, 2006] R. N. Giere. The role of agency in distribited cognitive systems. Philosophy of Science, 73:710–719, 2006.

[Girard, 1977] R. Girard. Violence and the Sacred [1972]. Johns Hopkins University Press, Baltimore, MD, 1977.

[Girard, 1986] R. Girard. The Scapegoat [1982]. Johns Hopkins University Press, Baltimore. MD, 1986.

[Glenberg and Kaschak, 2003] A. M. Glenberg and M. P. Kaschak. The body's contribution to language. In B. H. Ross, editor, The Psychology of Learning and Motivation, Vol. 43., pages 93–126. Academic Press, San Diego, 2003.

[Gomes et al., 200] A. Gomes, C. N. El-Hani, R. Gudwin, and J. Queiroz. Toward emergence of meaning processes in computers from Peircian semiotics. Mind and Society , 6:173–187, 200.

[Gonzalez and Haselager, 2005] M. E. Q. Gonzalez and W. F. G. Haselager. Creativity: surprise and abductive reasoning. The British Journal for the Philosophy of Science, 153(1):325–341, 2005.

[Gooding, 1990] D. Gooding. Experiment and the Making of Meaning. Kluwer, Dordrecht, 1990.

[Harris, 1989] R. Harris. How does writing restructure thought? Language and Communication, 9(2/3):99–106, 1989.

[Heeffer, 2007] A. Heeffer. Abduction as a strategy for concept formation in mathematics: Cardano postulating a negative. In O. Pombo, editor, International Meeting Abduction and the Process of Scientific Discovery, pages 179–194. Centro de Filosofia das Ciências da Universidade de Lisboa, Lisbon, 2007.

[Hoffmann, 1999] M. H. G. Hoffmann. Problems with peirce's concept of abduction. Foundations of Science, 4(3):271–305, 1999.

[Kant, 1929] I. Kant. Critique of Pure Reason. MacMillan, London, 1929. Translated by N. Kemp Smith, originally published 1787, reprint 1998.

[Karmiloff-Smith, 1992] A. Karmiloff-Smith. Beyond Modularity: A Developmental Perspective on Cognitive Science. The MIT Press, Cambridge, MA, 1992.

[Kirsh and Maglio, 1994] D. Kirsh and P. Maglio. On distinguishing epistemic from pragmatic action. Cognitive Science, 18:513–549, 1994.

[Kruijff, 2005] G.-J.-M. Kruijff. Peirce's late theory of abduction: a comprehensive account. Semiotica, 153(1/4):431–454, 2005.

[Lassègue, 1998] J. Lassègue. Turing. Les Belles Lettres, Paris, 1998.

[Lassègue, 1999] J. Lassègue. Turing entre formel et forme; remarque sur la convergence des perspectives morphologiques. Intellectica, 35(2):185–198, 1999.

[Logan, 2006] R. K. Logan. The extended mind model of the origin of language and culture. In N. Gontier, J. P. Van Bendegem, and D. Aerts, editors, Evolutionary epistemology, Language and Culture, pages 149–167. Berlin/New York, Springer, 2006.

[Longo, 2002] G. Longo. Laplace, Turing, et la géométrie impossible du "jeu de l'imitation": aléas, determinisme et programmes dans le test de Turing. Intellectica, 35(2):131–61, 2002.

[Loula et al., Forthcoming] A. Loula, R. Gudwin, C. N. El-HaniD., and J. Queiroz. Emergence of self-organized symbol-based communication in artificial creatures. Cognitive Systems Research, Forthcoming.

[Love, 2004] N. Love. Cognition and the language myth. Language Sciences, 26:525–544, 2004.

[Maddalena, 2005] G. Maddalena. Abduction and metaphysical realism. Semiotica, 153(1/4):243–259, 2005.

[Magnani and Belli, 2006] L. Magnani and E. Belli. Agent-based abduction: being rational through fallacies. In L. Magnani, editor, Model-Based Reasoning in Science and Engineering. Cognitive Science, Epistemology, Logic, pages 415–439, London, 2006. College Publications.

[Magnani, 2001a] L. Magnani. Abduction, Reason, and Science. Processes of Discovery and Explanation. Kluwer Academic/Plenum Publishers, New York, 2001.

[Magnani, 2001b] L. Magnani. Philosophy and Geometry. Theoretical and Historical Issues. Kluwer Academic Publisher, Dordrecht, 2001.

[Magnani, 2007] L. Magnani. Morality in a Technological World. Knowledge as Duty. Cambridge University Press, Cambridge, 2007.

[Magnani, Forthcoming] L. Magnani. Abductive Cognition. The Eco-Cognitive Dimension of Hypothetical Reasoning. Forthcoming.

[Menary, 2007] R. Menary. Writing as thinking. Language Sciences, 29:621–632, 2007.

[Mithen, 1999] S. Mithen. Handaxes and ice age carvings: hard evidence for the evolution of consciousness. In A. R. Hameroff, A. W. Kaszniak, and D. J. Chalmers, editors, Toward a Science of Consciousness III. The Third Tucson Discussions and Debates, pages 281–296, Cambridge, 1999. MIT Press.

[Monekosso et al., 2004] N. Monekosso, P. Remagnino, and F. J. Ferri. Learning machines for chance discovery. In A. Abe and R. Oehlmann, editors, The 1st European Workshop on Chance Discovery , pages 84–93, Valencia, Spain, 2004.

[Moorjani, 2000] A. Moorjani. Peirce and psychopragmatics. In J. Muller and J. Brent, editors, Peirce, Semiosis, and Psychoanalysis, pages 102–121. John Hopkins, Baltimore and London, 2000.

[Peirce, 1931-1958] C. S. Peirce. Collected Papers of Charles Sanders Peirce. Harvard University Press, Cambridge, MA, 1931-1958. vols. 1-6, Hartshorne, C. and Weiss, P., eds.; vols. 7-8, Burks, A. W., ed.

[Peirce, 1976] C. S. Peirce. The New Elements of Mathematics by Charles Sanders Peirce. Mouton/Humanities Press, The Hague-Paris/Atlantic Higlands, NJ, 1976. vols I-IV, edited by C. Eisele.

[Peirce, 1987] C. S. Peirce. Historical Perspectives on Peirce's Logic of Science: A History of Science. Mouton, Berlin, 1987. vols. I-II, edited by C. Eisele.

[Peirce, 1992-1998] C. S. Peirce. The Essential Peirce. Selected Philosophical Writings. Indiana University Press, Bloomington and Indianapolis, 1992-1998. Vol. 1 (1867-1893), ed. by N. Houser & C. Kloesel; vol. 2 (1893-1913) ed. by the Peirce Edition Project.

[Pinker, 2003] S. Pinker. Language as an adaptation to the cognitive niche. In M. H. Christiansen and S. Kirby, editors, Language Evolution, pages 16–37. Oxford University Press, Oxford, 2003.

[Prigogine and Stengers, 1984] I. Prigogine and I. Stengers, editors. Order out of Chaos. Man's New Dialogue with Nature. Bantam, London, 1984.

[Queiroz and Merrell, 2005] J. Queiroz and F. Merrell, editors. Abduction: between subjectivity and objectivity, volume 153. De Gruyter, Berlin–New York, 2005. Special Issue of the Journal Semiotica.

[Stenning, 2000] K. Stenning. Distinctions with differences: comparing criteria for distinguishing diagrammatic from sentential systems. In M. Anderson, P. Cheng, and V. Haarslev, editors, Theory and Application of Diagrams, pages 132–148. Springer, Berlin, 2000.

[Sterelny, 2004] K. Sterelny. Externalism, epistemic artefacts and the extended mind. In R. Schantz, editor, The Externalist Challenge, pages 239–254. De Gruyter, Berlin–New York, 2004.

[Svensson and Ziemke, 2004] H. Svensson and T. Ziemke. Making sense of embodiment: simulation theories and the sharing of neural circuitry between sensorimotor and cognitive processes. In K. D. Forbus, D. Gentner, and T. Regier, editors, CogSci 2004, XXVI Annual Conference of the Cognitive Science Society, Chicago, IL, 2004. CD-Rom.

[Thagard, 2002] P. Thagard. The passionate scientist: emotion in scientific cognition. In P. Carruthers, S. Stich, and M. Siegal, editors, The Cognitive Basis of Science, pages 235–250. Cambridge University Press, Cambridge, 2002.

[Thom, 1980] R. Thom. Modèles mathématiques de la morphogénèse. Christian Bourgois, Paris, 1980. Translated by W. M. Brookes and D. Rand, Mathematical Models of Morphogenesis, Ellis Horwood, Chichester, 1983.

[Thom, 1988] R. Thom. Esquisse d'une sémiophysique. InterEditions, Paris, 1988. Translated by V. Meyer, Semio Physics: a Sketch, Addison Wesley, Redwood City, CA, 1990.

[Thomas, 1999] H. J. Thomas. Are theories of imagery theories of imagination? An active perception approach to conscious mental content. Cognitive Science, 23(2):207–245, 1999.

[Turing, 1950] A. M. Turing. Computing machinery and intelligence. Mind, 49:433–460, 1950.

[Wheeler, 2004] M. Wheeler. Is language and ultimate artifact? Language Sciences, 26:693–715, 2004.

[Wilson, 1993] J. Q. Wilson. The Moral Sense. Free Press, New York, 1993.

[Woods, 2004] J. Woods. The Death of Argument. Kluwer Academic Publishers, Dordrecht, 2004.

[Zhang, 1997] J. Zhang. The nature of external representations in problem solving. Cognitive Science, 21(2):179–217, 1997.

[Zlatev, 2007] J. Zlatev. Embodiment, language, and mimesis. In T. Ziemke, J. Zlatev, and R. Frank, editors, Body, Language and Mind I: Embodiment, pages 297–338. Berlin, De Gruyter, 2007.

[Zwaan, 2004] R. A. Zwaan. The immersed experiencer: toward an embodied theory of language comprehension. In B. H. Ross, editor, The Psychology of Learning and Motivation, Vol. 44., pages 35–62. Academic Press, New York, 2004.

Science is Culture:

Neuroeconomics and Neuromarketing. Practical Applications and Ethical Concerns

By Sarah Rebecca Anne Belden

1 Introduction

Neuroeconomics is a relatively new transdisciplinary field, which developed out of Neuroscience. This burgeoning discipline analyses our brain activity when we calculate risks and evaluate rewards, and utilizes brain-scanning technology to study how people make decisions, evaluate personal choices and even decide which products to buy. Since the late 1990's a group of interdisciplinary scholars have begun to combine social and natural scientific approaches in this emerging discipline, combining both theoretical and empirical tools from neuroscience, psychology and economics into a single approach. The resulting synthesis has provided insights valuable to all three parent-disciplines, which recently conducted studies, seem to support. Often utilizing a variety of neuroimaging techniques and interventions such as fMRI, PET, MEG and EEG, ERP and SSPT, scientific researchers in this new field have sought to better understand the decision-making processes of individuals in order to build more precise economic behavioral models. These combined theories have already begun to restructure our neurobiological understanding of the decision-making process, and concurrently a number of recent neurobiological findings have provided great insight into some of the already existing theories in the psychological and economic branches of this discipline.

Since the 1990's however, a newer and more radical branch of Neuroeconomics has been born called Neuromarketing, which is aimed at revealing consumer preferences using these same brain-imaging techniques. Rather than simply trying to use science to better understand the decision-making processes of individuals, these neuromarketing studies test subjects' reactions to certain stimuli, which are then recorded with the aim of revealing consumer preferences. The results of these experiments are aimed at building targeted

advertising campaigns, designing new consumer products and shopping environments and even determining the reasoning behind subjects' preferences to certain brands such as Coke or Pepsi.

While this may be good news for Madison Avenue and the billion dollar advertising and marketing industries, as well as those corporations who employ these companies to help sell their products, the very idea of using brain scans to determine our private and personal predilections for the purpose of selling us more products seems rather invasive, if not Orwellian to say the least. Not only are there ethical concerns surrounding this new area of study, regarding the practical applications and their likely implications for individuals and society, but even more alarming, is the gusto with which the press, marketing firms, and Big Business have embraced the idea of "peering inside peoples heads" in order to better pin point their needs, desires and preferences as consumers. The idea of a "hard science," which can be utilized to uncover the holy grail of marketing or a magic "buy button" in our brain, is just too good for these industries to pass up, not to mention the scientists who have much to gain from peddling, what some call, pseudoscience for profit. At this stage neuromarketing is far from a "hard science" and the public should maintain a healthy dose of skepticism with regard to the practical applications of these neuroimaging techniques, which require many layers of signal processing, statistical analysis and a complex set of assumptions in order to interpret the psychological significance of these brain scans. But the public should also be aware of the ethical implications of this new type of neuroscience: how it is utilized; what its applications are; whether or not these new techniques are invasive and what the possible implications for society may be.

Hailed by some leading market researchers as the most important advance in their industry for a century, Neuromarketing has already been dismissed by skeptical neuroscientists as verging on a pseudo-scientific scam. A recent editorial in *Nature Neuroscience*, for example, suggested that many cognitive scientists who had watched colleagues in molecular science grow rich were now 'jumping on the commercial bandwagon,' adding that, "According to this view, neuromarketing is little more than a new fad, exploited by scientists and marketing consultants to blind corporate clients with science." Despite this, interest in Neuromarketing is growing rapidly. In 1998 less than 20 papers a year were published that examined the brain and decision-making, however, by 2008 nearly 200 articles relating to this particular area of study had been published. As reported in *Advances in Clinical Neuroscience and Rehabilitation* magazine there has been a marked increase in the number of articles in scientific journals and congresses organized around this new topic; entire issues have been devoted to neuromarketing in advertising and marketing trade publications; and it has even been reported that several new fMRI facilities, intended for Neuromarketing rather than medical purposes were opened in 2005 alone, in the United States. This is evidence enough to sound the alarm, however, while the public seems well aware of the ethical issues surrounding new scientific advances in molecular genetics, there has been little public awareness with regard to the ethical implications of neuroscience and neuromarketing.

2 Short History of Neuroeconomics

The first paper to explicitly combine neuroscientific data and a rigorous mathematical theory from the social sciences was Peter Shizgal and Kent Conover's 1996 review in Current Directions in Psychological Science: "On the neural computation of Utility." This paper sought to describe the neurobiological substrate for choice in rats using a normative economic theory. In 1999 this was followed by Michael Platt and Paul Glimcher's publication of "Neural correlates of decision variables in parietal cortex" which described a formal economic-mathematical approach for the physiological study of the sensory-motor process, or decision-making. This paper demonstrated that the activity of individual neurons in the posterior parietal cortex encoded both the probability and magnitude of reward, as would be predicted by most economic theories if these neurons participated in decision-making. This was rapidly followed by a multitude of papers uniting both economic and psychological theories with measurements in human and animal brains.

The first of these neuroeconomic studies in humans were a pair of papers published in 2001, which reflected collaboration between the fMRI pioneers Hans Breiter, Shizgal, and the Princeton psychologist/economist Daniel Kahneman (who would win the Nobel prize for his contribution to behavioral economics the following year). That paper employed the psychological Prospect theory of choice developed by Kahneman. The second of these papers reflected collaboration between the economists Kevin McCabe, his colleague Vernon Smith and a team that included economists, a psychologist and a biomedical engineer (McCabe et al., 2001). This study represented the first use of game theory in a human neurobiological experiment. In that paper, subjects played a trust game either against anonymous human opponents or against a computer. The neurobiological data revealed that in some subjects the medial prefrontal cortex is more active when subjects play a cooperative strategy than when they show a lack of trust in their game theoretic opponent.

Since the publication of these studies, perhaps the most critical insight has been evidence that the decision-making systems of the brain can be viewed as a two-part system. The first of these two parts are made up of the frontal cortex and the basal ganglia, the areas that learn and compute the values of available actions. The outputs of these structures are subsequently passed to the second part of the system; the fronto-parietal circuits, which then decide between the options and pass this information along to the motor system for execution. These are the areas that ultimately contribute to our decision making process.

With this plethora of research, Neuroeconomics has seen a steady growth. Today, a number of Centers for the study of Neuroeconomics have emerged at Universities throughout the world. In addition to these research centers, The Society for Neuroeconomics established itself as the main center for this emerging discipline in 2005. In 2009 the Society published, in collaboration

with Academic Press, "Neuroeconomics: Decision-Making and the Brain," which now serves both as a textbook for many graduate courses in Neuroeconomics, as well as a Handbook of Neuroeconomics for researchers in the field.

3 Short History of Neuromarketing

Neuroeconomics is a purely academic discipline concerned with the basic mechanisms of decision-making. In contrast, Neuromarketing is a more applied field concerned with the application of brain scanning technology to the traditional goals and questions of interest of the marketing industry. While the notion of Neuromarketing has been around for some 30 odd years, Professor Ale Smidts from Erasmus University is said to have first coined the term in 2002, and the first marketer to use fMRI was Gerry Zaltman at Harvard University beginning in 1999. The first marketing conference, which focused on the burgeoning field of Neuromarketing in 2004, was held at Baylor College of Medicine in Houston. While the most utilized and well recognized brain-imaging techniques are fMRI (functional magnetic resonance imaging), QEEG (Quantitative electroencephalography) and MEG (magneto encephalography), earlier forms of these techniques were being utilized as early as the late 1960's.

Before the development of these more sophisticated technologies researchers used pupilometers – devices that measure spontaneous pupil dilation as indicators of peoples' interest while they were looking at advertising or print advertisements. During this time, researchers also employed the use of GSR (Galvanic Skin Response) as a possible indicator of people's emotional response to advertisements. Later, new technology for eye tracking was developed which revealed exactly where on the page (or TV screen) people's eyes were focused. And finally, in the 1970's Herbert Krugman and Flemming Hansen began to explore processes that occur in the right/left brain hemispheres using electroencephalograph (EEG) brain wave technology. Each of these technologies was heralded at the time as groundbreaking, however none of these found widespread use for the purpose of marketing.

In 1981 SST (Steady State Topography) was utilized by Professor Richard Silberstein at Swinburne University, where he used this technology in clinical applications for possible use in marketing. The latest, and perhaps most widely known technologies are fMRI (functional magnetic resonance) and MEG (magneto-encephalography) which are both utilized as brain scanning devices. Both technologies show which areas of the brain "light up" when stimulated, producing a snapshot of the subjects brain. While there has been a great deal of hype surrounding these technologies and their potential applications for marketing, very few studies in peer reviewed journals have actually been published, deploying them for the use of marketing. One of the earliest studies conducted, utilizing these newer technologies was one performed by Professor Ambler and his colleagues at the London Business School. This study asked subjects who were placed in a MEG scanner, which of 3 brands they would purchase when given a choice. The results indicated that familiar

brands stimulate the right parietal cortex in the brain. The authors thus, theorized that this area of the brain was a possible "location of brand equity."

In 2000, Rossiter et al used SST to monitor brain waves while people watched TV ads. They were able to predict what scenes people would recognize a week later. They found they could predict this from activity in the left-brain at the time of exposure in the posterior region of the frontal cortex. Prior to this, it was thought that the crucial processing for pictures would be in the right hemisphere of the brain. Since 2000, many other similar studies have been conducted, which have resulted in relatively minor findings, most likely, due to the subjective and highly interpretive nature of this type of research. While each of these techniques has its strengths and weaknesses, there is also a great deal of detailed interpretation which goes into understanding the meaning of increased brain activation and in specifying what mental process is signified by an activation.

Most imaging studies report activations arising from the difference between two tasks. For each brain area, the signal during the task is compared to the signal at rest; those areas of the brain with stronger signals during the task are presumed to be processing the information. A very recent breakthrough however, may be able to detect the activity of an individual neuron in the future. At this stage however, the smallest brain area that can be represented - a voxel, is the size of a grain of rice and contains tens of thousands of neurons. It is interesting to note that there are about 100 billion neurons in the typical brain, but current fMRI resolution is only about 150,000 voxels. The changes in blood flow in a voxel thus, indicate increased activity of not a single neuron but a huge pod of tens of thousands of neurons.

4 Practical Applications: A Dubious Aim

In addition to some of the earlier Neuromarketing studies and applications already described herein, there are several other case studies that are of interest. These studies offer us a glimpse into exactly what these new technologies are being adapted for and how they are being applied, which is more often than not, for the sole purpose of marketing products to consumers. One such example is a study employed by Daimler Chrysler utilizing fMRI technology to see how consumers perceive their cars. These scans concluded that many sports cars activated the ventromedial prefrontal cortex, or what is called the "reward" centre of the brain, which is also reportedly activated by alcohol, drugs and sex. When shown a frontal view of these cars, the area of the brain that processes human faces was also shown to "light up." Boston based Ad agency Arnold Worldwide, hired by Jack Daniels employed similar brain imaging studies recently carried out at Harvard's McLean Hospital. These studies use fMRI scans to measure subjects' emotional responses to images associated with the activity of drinking in 25-34 year olds. The scans "help give us empirical evidence of the emotion of decision-making," says Baysie Wightman, head of Arnold's new science-focused Human Nature Department. These results apparently helped shape Jack Daniel's 2007 ad campaigns geared towards this particular demographic.

According to an article in the *Journal of Advertising Research* in 2001, another Australian study of TV commercials using brain wave technology (Steady-state Probe Topography) indicated that the left-brain was crucially involved in long-term memory for pictures. This was contrary to expectation, as it was previously thought that crucial processing of pictures was located in the right brain. Using the newer brain scanning technologies, the first studies of brands started to appear in 2002. One study performed in 2002 at the Psychology Department at the University of Los Angeles looked at exactly where brand names are processed in the brain and found more activity in the right brain than the left. Another study performed that same year at the London Business School examined people making a choice between brands and brand familiarity. Indicators showed up mostly in the right brain, in a place called the parietal cortex. Researchers apparently have their fingers crossed that this will turn out to be where brand equity resides, which no doubt will fuel a slew of additional studies in this specific area.

While much of the research is still mostly academic, many experts anticipate that that it's just a matter of time before these findings become a routine part of every competitive corporation's marketing plans. Some findings, such as the aforementioned discovery, which focuses on how the brain interprets brand names, are already enticing advertisers. Take, for example, the classic taste test. P. Read Montague of Baylor College of Medicine, who performed his version of the Pepsi Challenge with the use of an fMRI machine in 2004. In this study researchers repeated the famous Pepsi/Coca-Cola blind taste test challenge while scanning the brains of volunteers. When ignorant of which beverage they were sampling, the subjects favored Pepsi with their scans revealing activation of the ventromedial prefrontal cortex (a reward centre). However, when Montague repeated the test and told them what they were drinking, three out of four people said they preferred Coke. When aware of which brand they tasted, the scans revealed activity in the hippocampus, midbrain and dorsolateral prefrontal cortex – areas associated with memory, emotions and emotional information processing. This led the researchers to conclude that a preference for Coke is more influenced by the brand image than by the taste itself. Montague states that, "This showed that the brand alone has value in the brain above and beyond the desire for the content of the can."

Various studies have used verbal reports (e.g. scene recognition, brand preference); behavior (e.g. purchase vs. non-purchase); and even different segment reactions (e.g. Democrats vs. Republican brains are said to react differently to advertisements) to evaluate video clips and TV advertisements, study decision making among shoppers and even to investigate the likely impact of political advertising during the recent presidential elections. A study at the University of California, Los Angeles, for example reported differences in the neural responses of Democrats and Republicans to commercials depicting the 9/11 terrorist attacks. For the most part however, studies have been focused thus far, on the so-called 'known centers' such as: the rewards center, self-referencing center; and face recognition center. This has resulted in numerous

neuromarketing studies, which increasingly focus on the various 'known centers' in the brain, however the actual scientific data about these 'known centres' is very limited. A number of findings converge on the prefrontal cortex located in the lower forehead but no-one is clear yet as to precisely what all this means, thus, this should be considered more speculative at this point than anything else.

While the implications for marketing are problematic and mostly in the realm of speculation for the moment - we can, no doubt, expect a continuing accumulation of these studies in the near future. In any new scientific field, there is often a period where there is more speculation than proven research. This, coupled with the increasingly commercial nature of science, has resulted in a proliferation of pseudo experts in marketing, whose exaggerated claims and "powerful new marketing services," may do injustice to the real scientific research being conducted within this new discipline.

5 Critiques & Potential Ethical Concerns

Within the realm of Neuroeconomics and Neuromarketing there are a number of causes for concern. These are not only ethical, but also practical in nature. Concerning the applications of neurotechnology, there are a host of implications for individuals and society which should be considered carefully before these are put into wide spread use. Other potential implications may be considered more philosophical in nature, concerning the way we think about ourselves as persons, moral agents and even spiritual beings. In fact, there has already been a campaign organized against one such research project at Emory University. A national watchdog group headed by Ralph Nader called Commercial Alert has objected to Emory allowing Brighthouse, an Atlanta marketing consultancy, to use the university's neuroscience facilities for neuromarketing research. Commercial Alert has asked the Office for Human Research Protections, a division of the U.S. Department of Health and Human Services, to investigate whether the project violates federal guidelines for medical research.

Commercial Alert contends that it is wrong to use medical research for marketing instead of for the improvement and well being of humankind. The University has reviewed and approved the research, and states that the studies are making important contributions to Science, which will soon be published in scientific journals. However, it has been recently revealed that the university now no longer conducts this neuromarketing research on campus. Instead, Joey Reiman, who is an adjunct professor at Emory's business school and the proprietor of Brighthouse marketing consultancy, says that the university studies how the brain reacts to preferences, and then passes this information over to his consulting company, which is then hired by corporate clients. This raises many ethical questions about how this research is being used and such conflicts of interest are clearly a cause for concern. This type of research in the name of scientific knowledge is common, however selling this information to corporations whose job it is to manipulate people for profit is a dubious enterprise at best.

Despite how this information is or is not used, a much more philosophical question might be, how such invasive neuroimaging techniques are breaching the privacy of the human mind. This technological progress is making it possible to monitor and manipulate the human mind with increasing precision and with these techniques it may be possible to not only infringe upon the privacy of the human mind, but to judge people based not only by their actions, but also by their thoughts and predilections.

Brief Description of Technologies

Positron Emission Tomography or **PET** scans, were developed in the mid- 1970s, PET was the first scanning method to give functional information about the brain. Both PET and FMRI provide information about neural activity in different brain regions as indicated by the level of cerebral blood flow. With FMRI, the magnetic consequences of blood oxygenation are measured, whereas PET measures blood flow by first injecting people with a liquid radioactive tracer and measuring changes in radiation.

FMRI or **Functional Magnetic Resonance Imaging** and **MRI** or **Magnetic Resonance** Imaging require no radioactive materials and produce images at a higher resolution than PET. Originally used to take snapshots of what various brain injuries looked like, researchers realized that they could also use MRI machines to see which parts of the brain were being utilized in specific tasks, such as perception, language and memory – hence the term 'functional' MRI. This method involves very rapid scanning of the brain to see which areas of the brain are activated. When neural activity increases and the blood oxygenation in a region increases, this changes its magnetic properties. Increased neural action draws a bigger blood supply to support its work, which shows up—millisecond by millisecond —on an fMRI scan as magnetic changes. So, what fMRI detects is not neural activity directly, but magnetic changes that are blood-oxygen level dependent. The method is non invasive so multiple scans can be done on the same subject.

Magneto encephalography, or **MEG** is a very different brain scanning technique but used for similar purposes. The big advantage of MEG scans is that they are able to measure activity in the brain extremely quickly - every 1/1000 of a second, which is similar to the rate at which the brain works - essentially 'the speed of thought'. This method is closely related to electroencephalography or EEG, since they both try to measure the same neuronal currents. Electrical currents in the brain's neuronal circuitry give rise to very weak magnetic fields that can be picked up by superconducting detectors arranged around the outside of the head. The main disadvantages of MEG are that it is more expensive and not as good as fMRI at localizing, where, precisely in the brain, activity is taking place.

ERP – **Event Related Potentials,** also called **Evoked Response Potentials**
uses electrodes on the scalp to measure voltage fluctuations resulting from electrical activity in the brain. The "baseline" activity is then averaged out, leaving just the electrical responses evoked by each stimulus presentation. The location of where the activity is generated inside the brain has to be imputed mathematically. In animal studies and patients undergoing brain surgery, another way to localize ERP sources is to place electrodes directly on the brain.

SSPT or **Steady State Probe Topography** is used for monitoring activity during dynamic stimulus sequences, such as TV commercials. SSPT measures steady-state visually evoked potentials (SSVEP) and records at the rate of 13 times per second from 64 electrodes in a lightweight skullcap.

While important strides are being made in understanding the relation between the mind and the brain, our understanding of why people behave the way they do is closely bound up with the content of our laws, morals, social mores and religious beliefs. This is thus, a topic, which holds great philosophical weight for mankind and society as a whole.

We may also want to consider the physical invasiveness of some of these techniques, such as the PET scan, which utilizes radioactive tracers to detect brain activity in subjects, or even more invasive procedures carried out on patients in brain surgery, where electrodes are placed directly on the brain. We might also want to ask questions about the way in which many of these studies are conducted. Often subjects are lead to believe they are being tested for specific information, when in fact the tests being administered are employed for the purpose of obtaining other personal information surreptitiously, in studies designed for a completely different purpose. Perhaps it is not in an individual's best interest to have such personal information available to others, especially when considering that it will most likely be utilized by corporations and marketing firms who wish to use it to sell more of their products and make higher profits.

Another practical problem here is that the media, the public, the corporations and marketing firms interested in this new technology seem to think that it is completely full proof. For example, the general conception seems to be that brain scans "do not lie." This has created a great deal of misinformation and media reporting, which has outstripped any current scientific substance. This promotional hype has in turn, led some scientists, researchers and even universities to jump on the bandwagon in order to take advantage of the corporate dollars being spent by these dubious enterprises. Bearing these questions in mind, perhaps it is time we weigh the potential effects and possible ramifications of such research and how this may be used going forward in society at large. Will the research generated by this new discipline further our quest to better understand the mind and brain and add to the betterment of society as a whole? Or will it simply be usurped and corrupted by the all-powerful corporations who are already dictating so much of what is being funded in science now? Is it wise to allow precious funding dollars and University facilities to be used for the purpose of bolstering already ubiquitous and rampant consumerism? Wouldn't this funding be better used for the health and betterment of society rather than for capitalistic purposes? And will there be proper regulation for this type of research imposed, as in the case with biotechnology or stem cell research? These are the hard questions we must ask, not only for the preservation of the scientific community, but also for society at large.

References

Addison T. *More science: more sense or nonsense?* Ad-Map, Issue; 461:24 May 2005.

Tim Ambler, John Stins, Sven Braeutigam, Steven Rose & Stephen Swithenby. *Salience and Choice: Neural correlates of shopping decisions.* London Business School. Centre for Marketing Working Paper No. 01-902. April 2002.

Brammer, Michael. *Brain Scam?* Nature Neuroscience Vol.7. No. 7 683 July 2004.

Farah J. Martha, *Neuroethics: the practical and the philosophical.* Trends in Cognitive Science, Vol.9 No.1, January 2005.

Gabrieli J. D. E. *Cognitive Neuroscience of Human Memory*, Annual Review of Psychology, Vol. 49, 1998.

Glimcher Paul W. *Neuroeconomic,s* Scholarpedia, 3(10):1759, 2008.

Krugman HE. *Brain wave measures of media involvement.* Journal of Advertising Research,11:3-10, 1971.

Lovel Jim. *Nader Group Slams Emory for Brain Research*. Atlanta Business Chronicle, December 5, 2003.

McClure SM et al. *Neural Correlates of Behavioral Preference for Culturally Familiar Drinks*. Neuron 44 (2):379–87, 2004.

Park, Alice. *Marketing to your mind.* Time Magazine, January 19, 2007.

Possidonia Gontijo, Janice Rayman, Shi Zhang & Eran Zaidel. *How Brand Names are Special: Brands, Words and Hemispheres*. Psychology Department, Anderson School of Management University of Los Angeles, California. 2002

John Rossiter, Richard Silberstein, P. Harris, G. Nield,. *Brain-imaging detection of visual scene encoding in long-term memory for TV commercials*. Journal of Advertising Research, 41, 13-21, 2001.

Sutherland Max. *Neuromarketing: What's it all about?* www.sutherlandsurvey.com, March, 2007

David Lewis, Darren Bridger, Nueromarketing, Advances in Clinical Neuroscience and Rehabilitation, July / August 2005, 5 (3)

Symposia Call for Papers
BICS 2010
Brain-inspired Cognitive Systems

Madrid, Spain, July 14-16, 2010

Sixth International ICSC Symposium on Neural Computation (NC 2010)
Fifth International ICSC Symposium on Biologically Inspired Systems (BIS 2010)
Fourth International ICSC Symposium on Cognitive Neuroscience (CNS 2010)
Third International ICSC Symposium on Models of Consciousness (MoC 2010)

www.bicsconference.org

Motivation

Brain Inspired Cognitive Systems - BICS 2010 aims to bring together leading scientists and engineers who use analytic and synthetic methods both to understand the astonishing processing properties of biological systems and, specifically those of the living brain, and to exploit such knowledge to advance engineering methods for building artificial systems with higher levels of cognitive competence.

BICS 2010 is a meeting point of cognitive systems engineers and brain scientists where cross-domain ideas are fostered in the hope of getting new emerging insights on the nature, operation and extractable capabilities of brains. This multiple approach is necessary because the progressively more accurate data about brains is producing a growing need of both a quantitative and theoretical understanding and an associated capacity to manipulate this data and translate it into engineering applications rooted in sound theories.

BICS 2010 is intended for both researchers that aim to build brain inspired systems with higher cognitive competences, and as well to life scientists who use and develop mathematical and engineering approaches for a better understanding of complex biological systems like the brain.

BICS 2010 is organized around four major interlaced focal symposia that are organized into patterns that encourage cross-fertilization across the symposia topics. This emphasizes the role of BICS as a major meeting point for researchers and practitioners in the areas of biological and artificial cognitive systems. Debates across disciplines will enrich researchers with complementary perspectives from diverse scientific fields.

Dates

Submission of contributions: November 30, 2009
Notification of acceptance: February 28, 2010
Final contributions due: April 30, 2010
Conference: July 14-16, 2010

Program Committee

Jaime Gómez (*Technical University of Madrid*, Spain)
Chair of the PC

Amir Hussain (*University of Stirling*, UK)
NC Chair

Leslie Smith (*University of Stirling*, UK)
BIS Chair

Igor Aleksander (*Imperial College*, UK)
CNS Chair

Antonio Chella (*University of Palermo*, Italy)
MoC Chair

David Gamez (*Imperial College*, London, UK)
Hugo Gravato Marques (*University of Essex*, UK)
Alexei Samsonovich (*George Mason University*, VA, USA)
Raul Arrabales (*Universidad Carlos III*, Madrid, Spain)
Pentti Haikonen (*University of Illinois*, Springfield, IL, USA)
Tom Ziemke (*University of Skövde*, Sweden)
David Balduzzi (*University of Wisconsin*, WI, USA)
Riccardo Manzotti (*IULM*, Milan, Italy)
James Albus (*George Mason University*, VA, USA)
James Austin (*Cybula Ltd*, UK)
Giacomo Indiveri (*University of Zurich*, Switzerland)
Alister Hamilton (*University of Edinburgh*, UK)
F. Claire Rind (*Newcastle University*, UK)
Sue Denham (*University of Plymouth*, UK)
Philip Hafliger (*University of Oslo*, Norway)
David Windridge (*University of Surrey*, UK)
Luis Rocha (*Indiana University*, Bloomington, USA)
Shun-ichi Amari (*RIKEN Brain Science Institute*, Japan)
Jose C. Principe (*University of Florida*, USA)
Professor Ron Sun (*Rensselaer Polytechnic Institute*, USA)
Anil K Seth (*University of Sussex*, UK)
Bernard Widrow (*Stanford University*, USA)
Stephen Grossberg (*Boston University*, USA)
Umamaheshwari Ramamurthy (*University of Memphis*, TN, USA)
Hans-Heinrich Bothe (*Technical University of Denmark*, Denmark)
Marcilio Souto (*Federal University of Rio Grande do Norte*, Brazil)
Irene Macaluso (*Trinity College*, Dublin, Ireland)
Will Browne (*University of Reading*, UK)
Petros A. M. Gelepithis (*National University of Athens*, Greece)

Conference Scope

Neural Computation (NC)

NeuroComputational (NC) Systems · NC Hybrid Systems · NC Learning · NC Control Systems · NC Signal Processing · NC Architectures · NC Devices · NC Perception and Pattern Classifiers · Support Vector Machines · Fuzzy or Neuro-Fuzzy Systems · Evolutionary Neural Networks · Biological Neural Network Models · NC Applications

Biologically Inspired Systems (BIS)

Brain Inspired (BI) Systems · BI Vision · BI Audition and sound processing · BI Other sensory modalities · BI Motion processing · BI Robotics · BI Adaptive and Control systems · BI Evolutionary systems · BI Oscillatory systems · BI Signal processing · BI Learning · Neuromorphic systems

Cognitive Neuroscience (CNS)

CN of vision · CN of non-vision sensory modalities · CN of volition · Systems Neuroscience · Attentional Mechanisms · Affective Systems · Language · Cortical Models · Sub-Cortical Models · Cerebellar Models · Neural correlates

Models of consciousness (MoC)

World awareness · Self-awareness · Imagination· Qualia models · Virtual Machine Approaches · Formal Models of Consciousness · Control Theoretical Models · Developmental/Infant Models · Will and Volition · Emotion and Affect · Philosophical implications · Neurophysiological Grounding · Enactive approaches · Heterophenomenology · Analytic/Synthetic phenomenology

Organizing Committee

Ricardo Sanz
General Chair

Ramon Galán
Chair of the Organizing Committee

Carlos Hernández (Publications)
Iñaki Navarro (Media)
Manuel Rodríguez (Finance)

E-Mail and Symposia Website

info@bicsconference.org
oc@bicsconference.org
pc@bicsconference.org

www.bicsconference.org

ASLab
UPM Autonomous Systems Laboratory

*The **Autonomous Systems Laboratory** (ASLab) is a research group of the **Technical University of Madrid** (www.upm.es) focused on the development of technology for robust autonomy.*

If you've read, thought or done anything at all about Autonomous Systems (ASys), you'll probably know at least three things: ASys are the most exciting target for technical research; ASys can be really, we mean really, complicated; and lastly ASys are absolutely, ourageously, often unaffordably expensive in effort to build.

While ASLab has been created to change all that, it is not so different from the normal models for academic research. But, in a sense, we do all our research activities in cognitive science from an industrial-biased stance. We want to develop technology for autonomous systems to be deployed into the real world, so they will free humans from supervising them once they're up and running. The ASys shall self-manage.

ASLab research topics:

Cognitive control architectures
Integrated controllers
Model-based control systems
Ontologies for autonomous systems
Development processes for complex controllers
Reusable control components
Real-time middleware and platforms for distributed control
Retargetability of embedded control components
Technology of systems self-awareness
Philosophical implications of the technology of self-aware machines

Recent/ongoing research projects:

C3: Conscious Cognitive Control
HUMANOBS: Humanoids that Learn Socio-communicative Skills by Imitation
COMPARE: A Component Approach for Real-time and Embedded
AMS: Autonomous Modular Systems
MERCED: A Market Enabler for Re-targetable COTS
ICEA: Integrating Cognition, Emotion and Autonomy
HRTC: Hard real-time CORBA
GENESYS: Generic Embedded Systems Platform

Please visit our website: www.aslab.org

Journal of Mind Theory

Journal of
Mind Theory
Rigor in cognitive science vol. 0, n° 2, 2008

ISBN 9788461330218-1