Journal of Mind Theory

Rigor in cognitive science

volume 0, number 1, 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

Vindication of a Rigorous Cognitive Science *v*

Toward a Computational Theory of Mind *1*

The Mind as an Evolving Anticipative Capability *39*

The Challenges for Implementable Theories of Mind *99*

Interview:

Questions for a Journal of Mind Theory *111*

JMT Vol. 0 No. 2

Vindication of a Rigorous Cognitive Science *v*

MENS, a mathematical model for cognitive systems *129*

The Unbearable Heaviness of Being in Phenomenologist AI *181*

Pragmatics and Its Implications for Multiagent Systems *193*

Mimetic Minds as Semiotic Minds.
How Hybrid Humans Make Up Distributed Cognitive Systems *217*

Science is Culture:

Neuroeconomics and Neuromarketing.
Practical Applications and Ethical Concerns *249*

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 Laboratory of Autonomous Systems. Prior to this, he received his degree in Computer Sciences and worked for several years as a consultant in France, and as team leader in Spain for major technology companies. Since 2004, he has returned to Academia, where he currently continues his role 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. Professor 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, ren-

dered 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, they have together developed the model Memory Evolutive System 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

Toward a Computational Theory of Mind

James Albus

Krasnow Institute for Advanced Studies

Abstract
Scientific knowledge of the brain and technology of intelligent systems has developed to a point where a computational theory of mind is feasible. This paper briefly describes the RCS (Real-time Control System) reference model architecture that has been used successfully over the past 30 years for designing intelligent systems for a wide variety of applications. It then suggests how RCS can be mapped onto the neuronal structure of the brain, and vice versa. Both RCS and the brain are goal-directed and sensory-interactive intelligent control systems. Both are hierarchical in structure and partitioned into behavior generating and sensory processing hierarchies. Both rely heavily on an internal model of the external world for perception and behavior. The world model is used in perception for focusing attention, segmentation, grouping, prediction, and classification. It is used in behavior for decision-making, planning, and control. Both RCS and the brain have value judgment processes that assign worth to perceived objects, events, situations, and scenarios; and estimate the cost, risk, and benefit of plans for future behavior. The formal structure of RCS provides a framework for a computational theory of mind that is both quantitative and experimentally testable.

Keywords
Artificial intelligence, brain, cognitive science, mind, neuroscience, robotics.

1 Introduction

> *The task of neural science is to provide explanations of behavior in terms of activities of the brain.*
>
> -- Eric Kandel [26]

We are at a historical tipping point. The fundamental neuroscience and the technology of intelligent systems have matured to a level where it is possible to hypothesize quantitative theories of mind. The computational power and software engineering tools now exist to build experimental models of the brain that test theoretical predictions against observed performance in real world environments. There are good reasons to believe that a scientific expla-

nation of mind in terms of neuronal activities in the brain is feasible within the foreseeable future.

Much is known in the neuroscience and brain modeling communities regarding how the brain functions [26], [19], [18], [24], [27], [13], [14], [23]. Much is known in the computer science and intelligent systems engineering 1community about how to embed knowledge in computer systems [38], [46], [30], [8]. Researchers in robotics, automation, and control systems have learned how to build intelligent systems capable of performing complex operations in real-world, uncertain, and sometimes hostile, environments [20], [28], [31]. Computer hardware is approaching the estimated speed and memory capacity of the human brain and is increasing by an order of magnitude every five years [37], [29]. Reference model architectures and software development methodologies have evolved over the past three decades that provide a systematic approach to engineering intelligent systems [2].

This paper is an attempt to integrate knowledge from all of these disciplines into a framework for a computational theory of mind.

2 A Computational Theory of Mind

A computational theory of mind can be defined as a theory that models the mind as a set of processes, each of which has a computational equivalent [43]. There are many phenomena that are commonly attributed to the mind. These include imagination, thought, reason, emotion, perception, cognition, knowledge, communication, planning, wisdom, intention, motives, memory, feelings, behavior, creativity, consciousness, intelligence, intuition, and self. The fundamental hypothesis of a computational theory of mind is that each of these phenomena is a manifestation of an underlying process that has a computational equivalent.

One version of this hypothesis [2] suggests that *imagination* is a process of modeling, simulation, and visualization, i.e., generating and visualizing scenarios from assumptions about state, attributes, and relationships of objects, events, situations, and classes. *Thinking* is a process of imagining what might occur under various circumstances and analyzing the results. *Reasoning* is a process by which rules of logic are applied to representations of knowledge during the process of thinking. *Emotions* are mental states or feelings that result from a value judgment process evaluating what is good or bad, attractive or repulsive, important or trivial, loved or hated, hoped for or feared. *Feelings* are patterns of activity on particular sets of neurons that are perceived as pain, pleasure, joy, grief, hope, fear, love, hate, anxiety, or contentment. *Perception* is a process by which patterns of neural activity are interpreted as knowledge about the world, including self knowledge. *Knowledge* is information that is structured so as to be useful for thinking and reasoning. *Cognition* is a collection of processes by which knowledge is acquired and evaluated, awareness is achieved, reasoning is carried out, and judgment is exercised. *Meaning* is the set of semantic relationships that exist between the internal knowledge database and the external world. Meaning establishes what is intended or meant

by behavioral actions, and defines what entities, events, and situations in the knowledge database refer to in the world.

Understanding is what occurs when the system's internal representation of external reality is adequate for generating intelligent behavior. *Planning* is a process whereby a system imagines potential futures and selects the best course of action to achieve a goal state. *Wisdom* is the ability to make decisions that are most likely to achieve high-level long-range goals. *Introspection* is a process by which a system examines its own internal state and capabilities and reasons about its own strengths and weaknesses. *Reflection* is a process by which a system rehearses, analyzes, or thinks about the meaning of situations and events. *Reflexion* is a process whereby a system considers what others think about what it is thinking. *Attention* is a process by which an intelligent system directs sensors and focuses computational resources on what is important to its current goals – and ignores what is unimportant. *Awareness* is a condition wherein a system has knowledge of the structure, dynamics, and meaning of the environment in which it exists. *Consciousness* is a state or condition in which an intelligent system is aware of itself, its surroundings, its situation, its intentions, and its feelings [2].

According to this hypothesis, the duality of mind and brain can be explained in terms of the duality between a computing machine and the computational processes that occur in it. A machine is material, has mass, and occupies space. A computational process is immaterial, has no mass, and occupies no space. Yet, the immaterial process determines the behavior of the material machine, and the material machine is required for the process to exist. There is no need to appeal to spiritual essences that are beyond scientific investigation or to quantum effects that are inherently indeterminate.

The above hypothesis suggests that many of the phenomena of mind, once fully understood in terms of computational processes, will turn out to be less mysterious and inscrutable than many philosophers insist. By this hypothesis, the great mystery of dualism is reduced to simply the distinction between a machine and a process. This, of course, is controversial. Many in the philosophical arena would disagree. On the other hand, many in the computational sciences and neurosciences would feel it is self-evident. A full review of the multitude of viewpoints regarding the nature of the mind is far beyond the scope of this paper. A good sampling of support for the computational hypothesis can be found in [12], [13], [27], and [48].

The reader should understand that what is presented in this paper is by no means a fully formed computational theory of mind, but rather a conceptual framework that might lead eventually to such a theory. This framework is based on a set of assumptions that are stated here explicitly. They are:

1. The mind is a process, or more precisely a set of processes, and the brain is a machine in which the processes of mind occur. In short, the mind is what the brain does.

2. The phenomenon of mind results from activity within and between four elemental processes (behavior generation, sensory processing, world modeling, and value judgment) at many hierarchical levels.
3. These elemental processes of mind are supported by:
    a. *knowledge* represented in data structures that encode iconic, symbolic, declarative, procedural, and episodic knowledge; including what is perceived, known, or believed about the external environment, what is measured of the internal state of the mind and body, and what is considered to be good or bad, important or trivial.
    b. *communication* mechanisms that transport knowledge between functional modules within the brain.
4. All functional processes in the brain have computational equivalents.

These assumptions are essentially axioms that are accepted as true without proof. The purpose of axioms is to provide a starting point for a theoretical framework. The purpose of the above explicit assumptions is to provide the reader with a starting point for the particular framework described here.

The approach will be to:

1. Briefly describe a reference model architecture that has been successfully used for designing intelligent systems for a wide variety of applications.
2. Suggest how that reference architecture can be mapped onto the architecture of the brain, and vice versa.
3. Suggest how this mapping provides a framework for formulating a computational theory of mind.

3 A Reference Model Architecture

An architecture consists of functional modules, interfaces, communications, and data structures. A reference model architecture defines how the functional modules and data structures are integrated into subsystems and systems. It specifies functionality, modularity, connectivity, latency, bandwidth, reliability, semantics, and system performance. A reference model architecture is important because it enables a systematic methodology for engineering complex systems from a multiplicity of heterogeneous functional modules.

There are a number of architectures that have been developed for building intelligent systems. Some of these such as SOAR [30], ACT-R [8], Pilot's Associate [42], and various Black-Board and Expert Systems architectures [25] are designed to model high-level cognitive elements of human reasoning. However, they do not address the low-level details of perception and real-time behavior in the natural environment. Others such as Subsumption [10] and its many derivatives [11] have been designed to model low-level reactive behaviors. However, these do not address the high-level elements of cognition, knowledge representation, reasoning, and planning. Still others such as AuRA [9], CLARAty [45], and RCS [7] are hybrid architectures designed to combine high-level planning with low-level behaviors.

The cognitive architecture chosen for this paper is RCS (Real-time Control System). There are several reasons for this choice. One is that RCS addresses the full range of complexity inherent in embodied cognitive systems, from sensing through perception to cognition, decision-making, planning, and control of intelligent behavior in real-world environments.

A second reason is that RCS embodies a computational infrastructure that is plausible from a neuroscience viewpoint. RCS was originally inspired by the Marr-Albus model of the cerebellum [33], [6] and the CMAC (Cerebellar Model Articulation Controller) neural network. ([4], [5], [36]) RCS consists of a network of computational modules that accept inputs and produce outputs in a manner designed to emulate neural computation in the brain. It mimics the hierarchical structure that is observed throughout the brain. A Neutral Messaging Language (NML) that provides communications between computational modules mimics the neural pathways in the brain [39].

A third reason for selecting RCS is that it provides a mature engineering methodology that has been used by many teams of engineers over the past 30 years for building real-time controllers for a wide variety of robots and intelligent systems focused on real applications in real world environments. These include controllers for laboratory robots, machine tools, inspection machines, intelligent manufacturing systems, industrial robots, automated general mail facilities, automated stamp distribution systems, automated mining equipment, unmanned underwater vehicles, autonomous operations for nuclear submarines, and unmanned ground vehicles [2], [1], [32] RCS also provides software development tools and a simulation environment that are well documented and publicly available over the internet [39], [21].

3.1 Structure of the RCS Model

The basic form of the RCS reference model is shown in Figure 1. An internal model of the external world lies at the center. The World Model includes both a knowledge database and a set of World Modeling processes that provide three basic functions:

1. To build and maintain the knowledge database,
2. To service requests for information from Sensory Processing and Behavior Generation,

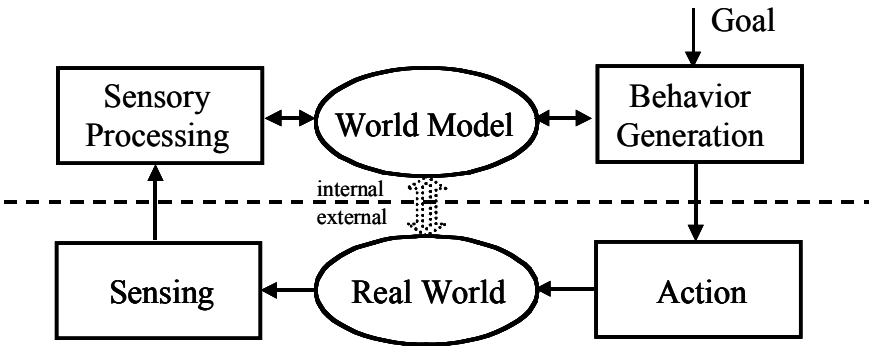

Figure 1: The fundamental structure of the RCS reference model architecture.

3. To generate predictions based on knowledge stored in the knowledge database.

The World Modeling processes generate short-term predictions for Sensory Processing to support focus of attention, segmentation, and recursive estimation. The World Modeling processes also generate long-term predictions for behavior to support decision-making, planning, and control.

Sensors and Sensory Processing processes extract information from the sensory data stream to keep the World Model current and accurate. Working together, Sensory Processing and World Modeling processes establish and maintain correspondence between the internal World Model and the external world environment.

Behavior Generation uses current knowledge in the World Model to generate actions that produce results in the external world. Behavior Generation uses the predictive capabilities of the World Model for decision making and planning to achieve or maintain goals.

The flow of information between the World Model and Sensory Processing is bi-directional. While Sensory Processing keeps the World Model updated, the World Model provides context and predictions to assist Sensory Processing in the interpretation of sensory data. The World Model provides Sensory Processing with knowledge of what is important. This is used for focusing attention. The World Model provides Sensory Processing with predictions of what kinds of objects and events to expect, where and when to expect them to occur, and what attributes and behaviors to expect them to exhibit. This information is used by Sensory Processing for segmentation, grouping, tracking, and classification of targets; and for processing of temporal sequences.

The flow of information between the World Model and behavior is also two-way. While the World Model provides Behavior Generation with information regarding the state of the external world, Behavior Generation provides the World Wodel with information about the state of the task. This enables the World Model to know what task is in progress, and what commands are currently being sent to actuators. This information enables the World Modeling processes to better predict what will happen in the future. Behavior Generation informs the World Model about plans for possible future actions. The World Modeling processes can then simulate the probable results of these possible future actions, and return an estimate of cost, benefit, and risk. This enables Behavior Generation to choose among alternative future courses of action. This two-way conversation between Behavior Generation and World Model is a planning loop.

3.2 First Level of Detail

A first level of detail in the RCS reference model is shown in Figure 2. Behavior is generated by Behavior Generation (BG) processes (Planners and Execu-

tors) supported by Task Knowledge. Task knowledge is procedural knowledge, i.e., skills, abilities, and knowledge of how to act to achieve task goals under various conditions. Task knowledge includes requirements for tools and resources, and lists of objects that are important to task performance. Task knowledge can be represented in the form of task frames, recipes, schema, state graphs, procedures, programs, rules, or flow charts. Task knowledge is used by Planners and Executors to make decisions, generate plans, and control behavior [1], [2]. Some task knowledge is acquired by learning how to do things, either from a teacher or from experience [47]. Some task knowledge is embedded in BG software in the form of algorithms and data structures containing *a priori* information about the world and the intelligent system.

Perception is enabled by Sensory Processing (SP) processes that operate on sensory input to window (i.e., focus attention), segment and group entities and events, compute attributes, do recursive estimation, and perform classification operations. The World Model is composed of a Knowledge Database (KD), a set of World Modeling (WM) processes, and a set of Value Judgment (VJ) processes.

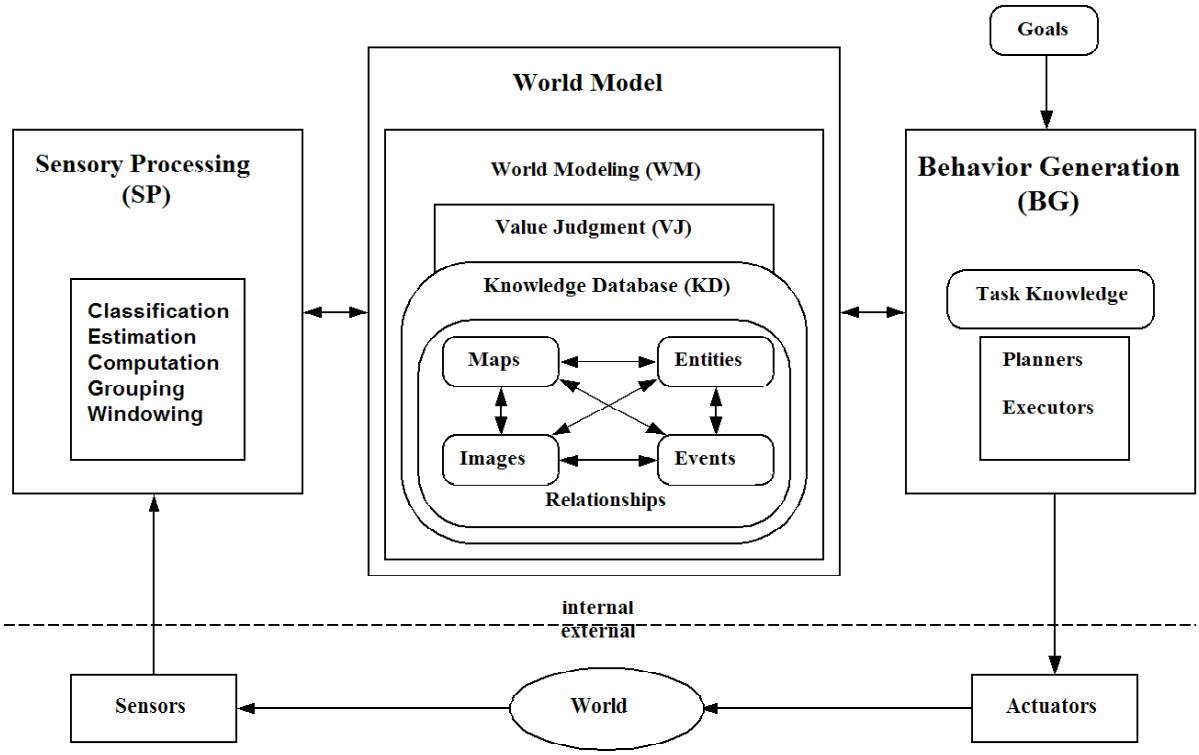

Figure 2: A first level of detail in the basic structure of intelligent systems. Processes are represented by boxes with square corners. Data structures are represented by boxes with rounded corners. Data flow between processes and pointers within KD are represented by arrows.

The Knowledge Database (KD) contains declarative and episodic knowledge. Knowledge in the KD is learned, either from a teacher or from experience. Declarative and episodic knowledge are represented in both iconic and sym-

bolic forms. Iconic forms include images and maps. These are two-dimensional projections of the three-dimensional structure of the external world. Images are in the coordinate frame of the sensors. Maps are in a coordinate frame that is ideal for planning behavior, such as an overhead view of the ground. Maps may be of different scales, and are often overlaid with icons and features such as terrain contour lines, roads, bridges, and streams. In the RCS model, images and maps are represented by two-dimensional arrays of pixels (i.e., picture elements consisting of scalars or vectors, each element of which represents a signal or attribute value.) In the brain, images exist in the retina, the visual cortex, and the somatosensory cortex. Maps exist in the posterior parietal cortex, the hippocampus, and possibly other regions.

Symbolic forms include entities, events, and relationships. Entities represent segmented regions of space. Events represent segmented intervals of time. Relationships are linkages that exist between and among entities and/or events. In RCS, entities and events can be represented by abstract data structures such as LISP frames, C structs, or C++ objects and classes. Relationships are represented by pointers. Entities and events in the KD can be linked by pointers to represent places, situations, and episodes. For example, parent-child relationships and class membership relationships can be represented by "belongs-to" or "has-part" pointers. Situations and places can be represented by graph structures and semantic nets. Episodes are strings of situations that occur over extended periods of time. Episodes can be described by linguistic structures such as words, phrases, sentences, and stories. Spatial, temporal, mathematical, social, and causal relationships can be represented by abstract data structures and pointers that link them together in networks that provide context and meaning.

Direct contact with the external world is provided by patterns of light on the retina, patterns of tactile stimulation on the skin, patterns of pressure variations in the cochlea, patterns of acceleration in the vestibular system, patterns of excitation of the organs of taste and smell, and patterns of input from internal and proprioceptive sensors in the body. Signals from sensors enter the sensory processing system in iconic form.

Patterns of signals are transformed into symbolic form through the operations of segmentation and grouping. Pixels in images are segmented and grouped into patterns, or entities, e.g., edges, surfaces, objects, groups, situations, and places. Strings of acoustic signals are segmented and grouped into events, e.g., sounds, phonemes, words, sentences, stories, and episodes. Patterns of smell and taste are grouped into classes. Patterns of proprioception and vestibular signals are grouped into body posture and gait.

These grouping and segmentation operations generate pointers that link iconic to symbolic representations, and vice versa. Pointers link pixels in images and maps to symbolic frames representing entities, events, and classes. Forward pointers linking iconic to symbolic representations provide the basis for symbolic reasoning. This enables an intelligent system to perceive the world not as a collection of pixels and signals, but as a montage of objects,

events, situations, and episodes. Meaning occurs when signals from sensors are grouped into entities and events with behavioral significance. Back-pointers link symbolic frames back to images and maps, which can be projected back onto the sensory data stream. Back-pointers provide the basis for symbol grounding. This enables the intelligent system to project context and meaning onto sensory experiences, and hence onto the external environment. Two-way links between iconic and symbolic forms provide the basis for scene and speech understanding, abductive inferencing, and symbol grounding.

The KD includes knowledge stored in long-term memory, short-term memory, and immediate experience. The RCS model assumes that long-term memory is symbolic and stored in non-volatile media. Short-term memory is symbolic and stored in volatile media such as finite state automata or recirculating delay lines. Immediate experience is iconic and stored in dynamic registers in active processes such as recursive estimation, adaptive resonance circuits, and control loops.

3.3 World Model (WM) and Value Judgment (VJ) Processes

A block diagram of WM and VJ processes is shown in Figure 3. The WM processes provide the database management and learning functions required to keep the KD current, and relevant to what BG processes require for achieving task goals. WM processes service queries for BG executors, and provide simulations for BG planners. WM processes provide model-based predictions and visualizations in support of recursive estimation, focus of attention, and segmentation processes in SP.

The VJ processes provide a variety of evaluation functions required for in-

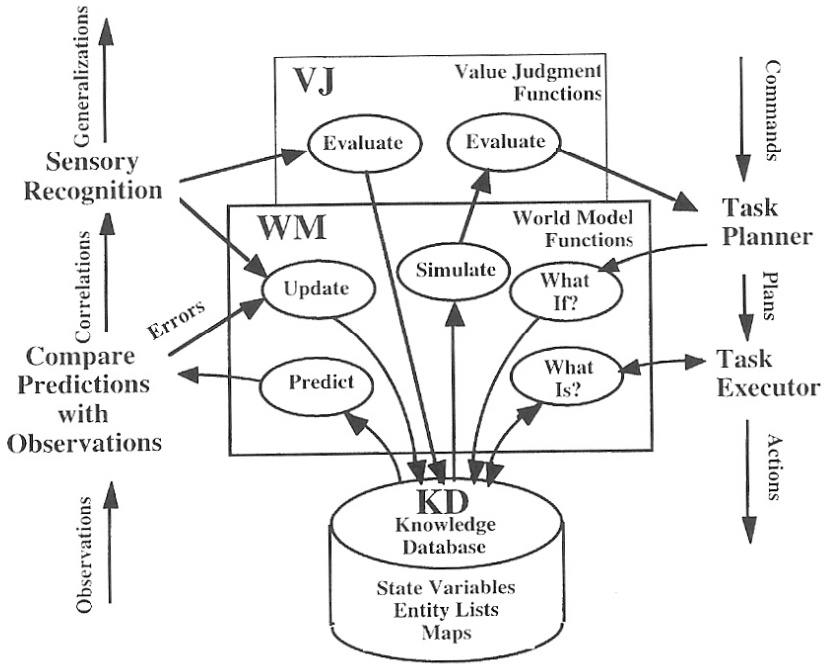

Figure 3: A block diagram of WM and VJ processes (From [2])

telligent decision-making, planning, focus of attention, and control. Evaluations from VJ processes provide criteria for decision-making by BG in the selection of goals and assignment of priorities to tasks. VJ processes provide the criteria for selecting modes of behavior such as aggressive vs. cautious, or fight vs. flee. VJ processes evaluate the cost, benefit, and risk of expected results of hypothesized plans.

VJ processes compute levels of confidence for perceptual hypotheses and assign worth to entities, events, situations, and episodes entered in the KD by WM. VJ processes evaluate the behavioral significance of objects, events, and situations, and provide the basis for deciding whether something is worth storing in long-term memory. VJ processes provide the basis for assessment of what is good or bad, attractive or repulsive, beautiful or ugly. For intelligent systems, VJ processes determine what is to be feared or hoped for, loved or hated. In highly intelligent systems, VJ processes may generate feelings and beliefs that provide a sense of duty, justice, and morality.

3.4 The RCS Hierarchy

A fundamental feature of the RCS reference model architecture is its hierarchical structure. The BG hierarchy consists of echelons of nodes containing intelligent agents. An example of a RCS organizational hierarchy is shown in Figure 4. This is the version of RCS (4D/RCS) that was developed for the Army Research Laboratory Experimental Unmanned Vehicle (XUV) program,

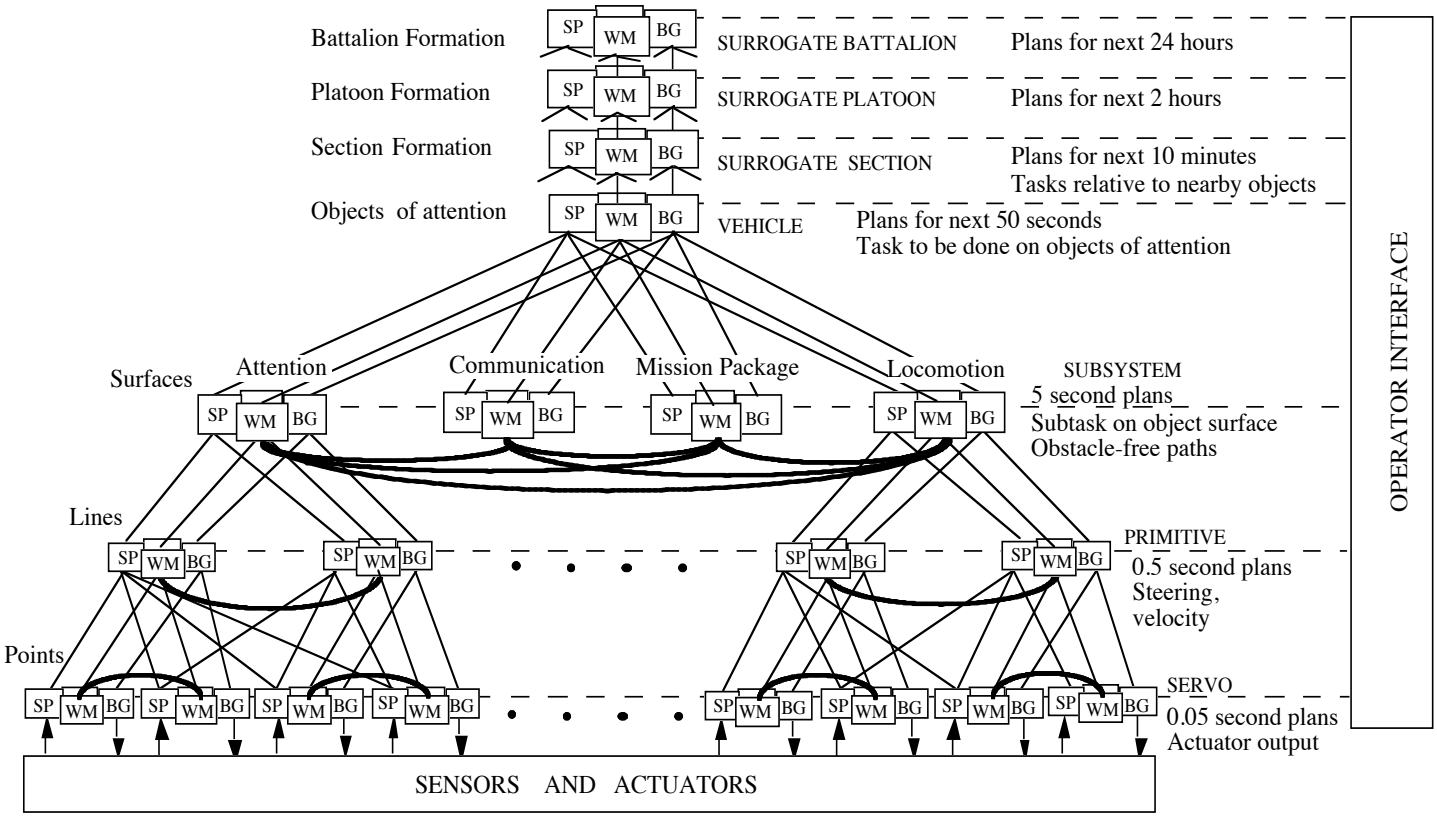

Figure 4: The 4D/RCS reference model architecture developed for the Army Research Laboratory Demo III experimental unmanned ground vehicle program. This example is for an autonomous vehicle in a scout platoon (from [1]).

[41] and since adopted for the Autonomous Navigation System being developed for the U.S. Army Future combat system [1], [32], [34].

At each echelon, BG processes in RCS nodes decompose task commands from a higher echelon node into subtasks for one or more subordinate nodes in a lower echelon. At each level, SP processes focus attention, segment and group patterns of sensory data from lower levels, and compute attributes and classifications of those patterns for higher echelons. At each level and each echelon, WM processes use SP results to maintain KD data with range and resolution needed to support BG planning and control functions in each node. VJ processes provide evaluations to support decision-making, planning, control, memory, and focus of attention in BG, WM, and SP processes in each node.

Each 4D/RCS node has a well-defined role, set of responsibilities, span of control, range of interest, and resolution of detail in space and time. These nodes can be configured to model any management style, and can be reconfigured at any time in response to changing task priorities and resource availabilities.

The example in Figure 4 is a reference model for a controller for a single scout vehicle in a section of a platoon attached to a battalion. A similar reference model might be developed for a single human being embedded in a social structure consisting of an immediate family, an extended family, and a tribe. Processing nodes are organized such that the BG processes form a chain of command. There are horizontal communication pathways within nodes, and information in the knowledge database is shared between WM processes in nodes above, below, and at the same level within the same subtree. On the right in Figure 4, are examples of the functional characteristics of the BG processes at each echelon. On the left, are examples of the scale of maps generated by SP-WM processes and populated by the WM in the KD at each level. VJ processes are hidden behind WM processes in the diagram. A control loop may be closed at every node. An operator interface provides input to, and obtains output from, processes in every node. Numerical values in the figure are representative examples only. Actual numbers depend on parameters of specific vehicle dynamics. It should be noted that there are not necessarily the same number of SP levels as BG echelons. This is discussed later in more detail.

The BG process in each node has a well-defined and limited set of task skills. Each echelon in the BG hierarchy relies on the echelon above to define goals and priorities, and provide long-range plans. Each node relies on the echelon below to carry out the details of assigned tasks. Within each node, the KD provides a model of the external world at a range and resolution that is appropriate for the behavioral decision-making and planning activities that are the responsibility of that node. This hierarchical distribution of roles and responsibilities provides a way to manage computational complexity as systems scale up to real-world tasks and environments.

The BG hierarchy is not fixed. It can be reconfigured at any time so that subsystems within vehicles can be replaced, or vehicles can be reassigned to different chains of command whenever required.

Note that in Figure 4 there are surrogate nodes for the Section, Platoon, and Battalion echelons. These enable any individual vehicle to assume the role of a section, platoon, or battalion commander. Surrogate nodes also provide each vehicle with higher echelon plans, models, goals, rules of engagement, and priorities during those periods of time when the vehicle is not in direct communications with its supervisor. Surrogate nodes in every individual vehicle enable it to cooperate effectively with others, and act appropriately in teams, even when contact with supervisor nodes is interrupted.

Similarly most, if not all, human brains contain the potential to assume the role of head of an immediate family, an extended family, or tribe. And every mature adult has the ability to cooperate effectively and act appropriately, even when not under the immediate supervision of an authority figure. It can be conjectured that higher echelon nodes in the brain are dedicated to planning and controlling behavior of the individual relative to (i.e., in collaboration with or in competition with others) family, friends, and larger organizations.

3.5 Perception

The role of perception is to build and maintain an internal model of the external world with range and resolution that is appropriate for behavior generating processes at every echelon of the BG hierarchy. Perception is accomplished by interactions between SP and WM processes that provide attention (i.e., windowing), segmentation and grouping, computation of attributes, filtering and confirmation of grouping (i.e., recursive estimation), and classification. At each level in the SP hierarchy, patterns in the sensory input are grouped into entities and events. For each entity or event, pointers are established that define relationships to other entities and events, and to the regions in space or time that comprise them.

The diagram in Figure 5 shows how bottom-up SP processes that operate on sensory input are influenced by top-down information from *a priori* KD knowledge, and BG representations of tasks and goals. These interactions occur at almost every level in the SP hierarchy. At the bottom left of Figure 5, subentity images enter a SP level to be processed. At the lowest level, a subentity image is simply an array of pixels from a camera or a retina. At higher levels, a subentity image is the output from a lower level SP process. The SP process of windowing operates to mask out regions of the image that are without behavioral significance, and focus SP-WM resources on regions that are important to achieving behavioral goals. The SP process of segmentation separates subentities that belong to entities of interest from the background. Grouping clusters subentities into entities based on some gestalt hypothesis (e.g., proximity, similarity, good continuation, symmetry, etc.) Grouping labels the segmented subentities with the name of the entity to which they be-

long. At various levels in the SP-WM hierarchy, grouping yields entity images of edges, boundaries, surfaces, objects, or groups of objects. The result of each grouping operation is a hypothesized entity image wherein each pixel in the entity image is labeled with the name of the group to which it belongs. Each grouping operation generates pointers from subentities to entities, and vice versa. This establishes links between labeled regions in the iconic representation, and named entity or event frames in the symbolic representation.

Once segmentation and grouping have been achieved, SP and WM computation processes can then compute the value of entity attributes (e.g., size, shape, color, and texture) and state (e.g., position, orientation, and motion) for each segmented region of interest in the entity image. Next, SP-WM recursive estimation processes generate predicted entity attributes to be compared with observed entity attributes. When predicted attributes match observed attributes, confidence in the gestalt grouping hypothesis is increased. When the confidence rises above threshold, the grouping hypothesis is confirmed.

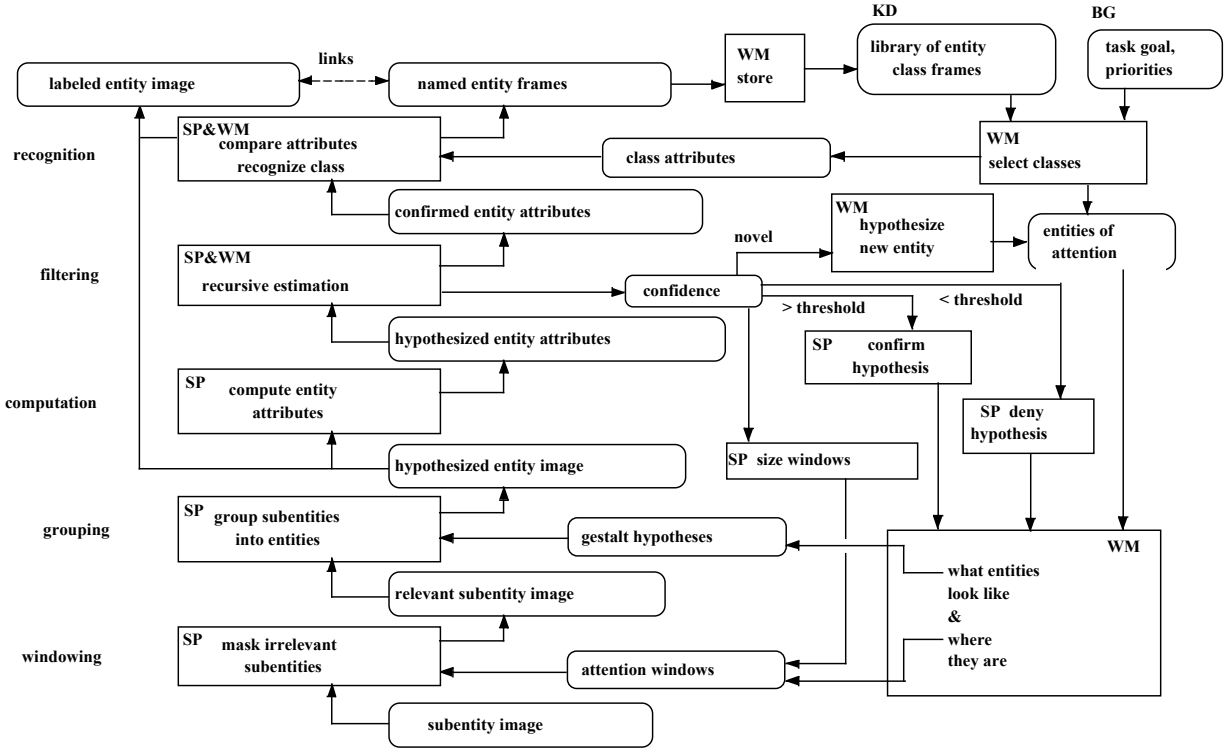

Figure 5: A data flow diagram of interactions between SP and WM processes in a typical SP-WM level. Functional operations are shown in boxes with square corners. Data structures are shown in boxes with rounded corners (from [2]).

During the recursive estimation process, small differences between predictions and observations are used to update the model. Large differences may cause the level of confidence in the grouping hypothesis to fall below threshold. When this happens, the hypothesis is rejected, and another gestalt hypothesis must be generated. If a suitable hypothesis cannot be found, then the

observed region in the image is declared novel, and worthy of inclusion on the list of entities of attention as a region of interest.

Once the grouping hypothesis is confirmed, the list of attributes in the confirmed entity frame can be compared with the attributes of stored entity class prototypes in the KD. When a match is detected, the entity is assigned to the matching class. This establishes a class pointer from the entity frame to the name (or address) of the class prototype frame. Each pixel in the entity image can then inherit additional class attributes through its link with the entity frame. Finally, a VJ process determines whether or not the classified entity is of sufficient importance to be stored in long-term memory. If so, then a WM process will enter the classified entity frame into long-term memory in the KD.

Top-down information from BG processes about task goals and priorities enter Figure 5 at the top-right. This top-down information enables a WM process to select a set of entity classes that are important to the task from a library of prototype entity class frames that resides in long-term memory. This set of important entities is prioritized and combined with bottom-up information about novel regions of interest. The result is a list of entities and regions of attention. This list is used by WM processes at the bottom right of Figure 5 to generate expectations and predictions regarding where these entities and regions of attention should be expected to appear in the image, and what they should be expected to look like.

What entities are expected to look like is defined by the attributes in the prototype entity class frames. *What* information provides guidance to the heuristic selection of gestalt hypotheses that will be used to control the grouping of subentities into entities. *Where* entities can be expected to appear in the image can be computed from the state-variables in the entity class frames. *Where* information provides guidance to the heuristic processes that define windows of attention to be used to control sensory pointing, tracking, and segmentation operations.

More detail about the RCS reference model architecture is available in [2] and [1].

4 Mapping the Model onto the Brain

We turn now to the task of mapping the RCS model onto the brain, and vice versa. It is assumed that the brain is first and foremost a control system, consisting of an integrated society of computational modules that has evolved through natural selection to enable the self-organism to survive and propagate in a complex, uncertain, and often hostile world. It is assumed that the machinery of the brain is organized so as to sense and perceive the external world, to build and maintain an internal model of the world, and generate and control behavior in pursuit of goals. Goals are defined as desired states to be achieved or maintained. Some goals are generated internally in response to perceived needs, drives, motives, and beliefs. Other goals are generated exter-

nally by requests from peers or subordinates, or by commands from family, tribal, religious, political, or military authorities.

The anatomy of the brain is as stereotypical as the anatomy of the body. The brain is organized front-to-back such that sensory processing (SP) and the world model (WM) processes that support SP are located in the back. Behavior generation (BG) and those portions of the WM that support the kinematics and dynamics of planning behavior are located in the front. The brain is organized top-to-bottom as a hierarchy of BG echelons and SP levels. At each level of SP and each echelon of BG, there are connections with the limbic system. These correspond to value judgment (VJ) processes.

VJ processes mimic the function of emotions. Patricia Churchland defines emotions as "the brain's way of making us do and pay attention to certain things. They are assignments of value that direct us one way or another." [12]. The emotions are the outputs of the parts of the brain that assign value, estimate cost, risk, and benefit, and define what is important. At the most fundamental level, the emotions are feelings and urges that motivate self-preservation and gene propagation. The portions of the brain that perform these functions belong to the limbic system. The limbic system is a collection of regions near the center of the brain that surround the thalamus. The most primitive parts of the limbic system are located in the region that processes smell and taste. The VJ processes include the hypothalamus where the drives of thirst, hunger, and sexual arousal are computed. They include the amygdala where feelings of fear and rage are generated, and the emotional importance of objects and events is computed. They include the system of sensors and processing modules that generate feelings of pain, both physical and emotional. The VJ processes also include the reticular activating system, the pain-pleasure centers, and the elements of the autonomic nervous system that report on bodily health and physical fitness.

4.1 Back-to-front

Back to front, the brain is organized such that, for the most part, sensory processing (SP) is in the back (i.e., posterior regions), and behavior generation (BG) is in the front (i.e., anterior regions.) In the spinal cord, sensory information enters through the dorsal roots, and motor neurons exit through the ventral roots. In the cerebral cortex, the primary sensory processing areas for vision, hearing, and somatosensory data are located behind the central sulcus, and behavior generating areas are located forward of this central dividing line.

World modeling (WM) and knowledge data-structures (KD) are split between the back and front. Those parts of WM and KD that support sensory processing are located in the back. Those parts of WM and KD that support the BG processes of kinematic and dynamic planning are located in the front.

Two-way communication pathways connect SP, WM, and BG processes throughout the human nervous system. Information flowing from SP through WM to BG closes feedback control loops at every BG echelon. Information

flowing from BG through WM to SP provides behavioral context and top-down expectations for perception at almost every level in the SP hierarchy. Sensory input is compared against model-based predictions at many levels of abstraction. At higher SP levels, iconic images and maps are linked to symbolic representations, and vice versa. Both iconic and symbolic representations are used to facilitate interpretation of sensory data. Both iconic and symbolic representations are also used for planning and visualization of alternative courses of action.

At each echelon in the BG hierarchy, WM information maintained by SP processes in the posterior brain supports BG planning and control functions in the frontal regions. At each level in the SP-WM hierarchy, input from BG echelons influence sensory processing. In short, each level of SP affects motor behavior, and each echelon of BG influences sensory processing.

Communications between the frontal cortex and the parietal and occipital cortices are accomplished by fibers in the superior longitudinal fasciculus, the arcuate fasciculus, the inferior occipito-frontal fasciculus, and the cingulum. Communications between the frontal cortex and temporal cortex are accomplished by fibers in the unicate fasciculus. Long fibers in the cingulated and arcuate fasciculi connect high levels in the multi-modal sensory processing hierarchy with the prefrontal cortex. Medium length fibers connect mid levels of the uni-modal sensory hierarchy with premotor cortex. Some of these medium length fibers in the unicate fasciculus connect Wernike's speech comprehension area and Broca's speech generation area. Shorter fibers provide tight coupling between the primary sensory cortex and corresponding regions in the primary motor cortex. There are few connections between the highest level behavior generating regions in the prefrontal cortex and the primary sensory cortex.

Value judgment (VJ) processes required for focusing attention and situation evaluation at all SP levels, and decision-making and planning at all BG echelons, reside in the limbic system and related structures. The limbic system includes the cingulate gyrus of the cortex that runs from front to back in the midline fissure between the two sides of the brain, plus the subcortical structures of the amygdaloid complex, the septal nuclei, the hippocampus, and the hypothalamus. The anterior cingulate provides high level estimates of cost and benefit of long-range plans that are formulated in the prefrontal cortex. The amygdaloid complex computes appropriate levels of fear and rage. The hippocampus computes degrees of importance for remembering objects, events, and situations. The reticular activating system is a midbrain level VJ function that detects novelty and generates alerting signals. VJ processes in the spinal cord are provided by pain sensors.

VJ processes in the posterior brain evaluate and assign value to entities, events, situations, and episodes; and assign confidence to grouping hypotheses and classifications in support of SP-WM processes. VJ processes in the frontal brain evaluate success or failure of behavior; and estimate the cost, risk, and benefit of simulated results of hypothesized plans in support of BG-WM processes.

4.2 Top-to-bottom

Top-to-bottom, the brain is organized hierarchically. Fuster describes two cortical hierarchies: one for sensory processing in the posterior cortex, and the other for motor control in frontal cortex. This is shown in Figure 6.

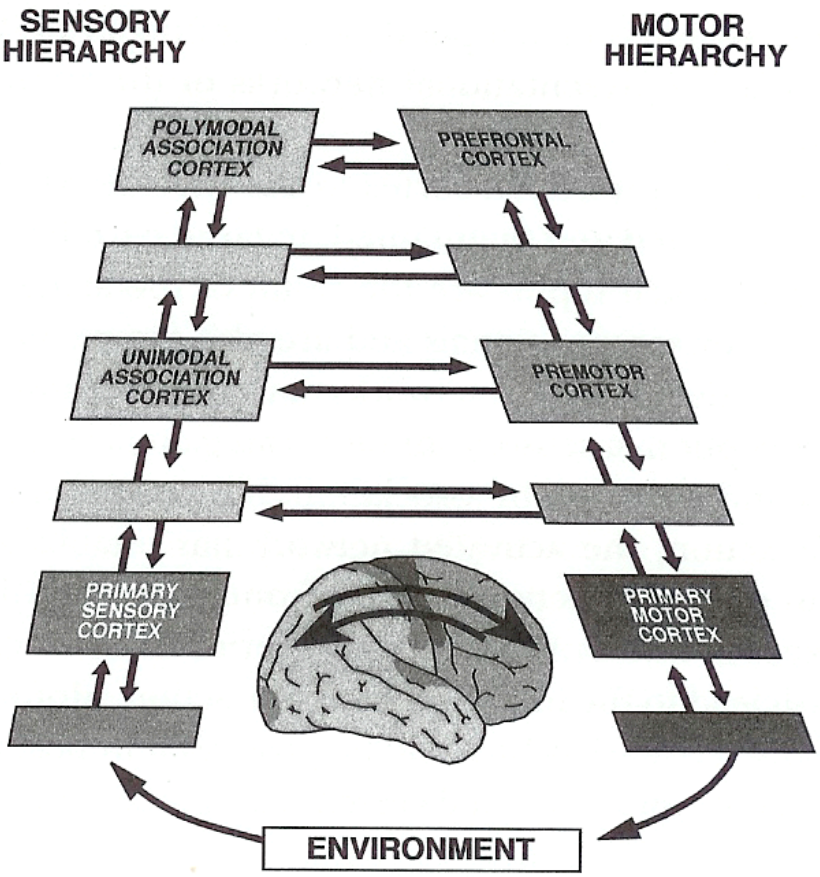

Figure 6: Two hierarchies in the cortex: one for perception in the posterior cortex, and one for behavior in frontal cortex. Two-way communication pathways between posterior and frontal cortices link the two together. (from [18] by permission.)

4.2.1 The Motor Hierarchy in the Brain

At the top of the motor (BG) hierarchy in the brain, high-level decision-making and long-range planning occur in the prefrontal cortex. Mid-level

decision-making and mid-range planning functions are computed in the premotor regions. Simple behaviors are selected and sequenced in the primary motor cortex. Computations of coordinated dynamic forces and motions necessary to maintain stability and generate coordinated dynamic movements are performed in the basal ganglion, cerebellum, and midbrain motor centers. Servo control of muscles for moving eyes, head, arms, hands, fingers, torso, legs, and feet are performed in midbrain and spinal motor centers. At each echelon, BG processes are supported by WM processes that simulate plans, and VJ processes that evaluate predicted results.

The motor hierarchy is defined by the decomposition of tasks and goals into subtasks and subgoals. At each echelon in the motor hierarchy, higher echelon goals and priorities are decomposed into lower echelon plans and patterns of action.

It is widely recognized that the lateral prefrontal cortex sits at the top of the behavior generating hierarchy where it plays an overarching role in decision making and planning of behavior. The prefrontal cortex is occupied with high-level goals, priorities, and plans – not with details of action. The prefrontal cortex plays a prominent role in the formation of novel, complex, and temporally extended behaviors. Some have called it the "organ of creativity." [19]. Others have identified it as the center of consciousness [27].

The premotor cortex sits just beneath the prefrontal cortex in the behavior generation hierarchy. Anticipatory discharge of premotor neurons generally begins after that of prefrontal neurons and as much as a few seconds before that of primary motor neurons. Single unit studies show that, in general, motor representation in the premotor cortex is not defined in terms of particular muscles, or muscle groups, but in terms of global movement, trajectory, or target. Premotor neurons are activated in anticipation of a purposive movement to attain a particular goal, not in anticipation of a particular motion of the body.

The primary motor cortex (M1) resides below the premotor cortex in the BG hierarchy. Neurons in M1 fire immediately preceding the movement of particular muscle groups, and are selectively tuned to dynamic loading. Motor representation is more somatotopically organized and more related to dynamic and kinematic aspects of movement.

The basal ganglia lie beneath the cortex and include the putamen, caudate, globus pallidus, substantia nigra, and the subthalamic nuclei. In lower vertebrates, the basal ganglia represent the highest level in the behavior generation hierarchy. In humans, the function of the basal ganglia in human brains is dramatically influenced by reentrant loops from the frontal cortex and thalamus. This suggests that the basal ganglia in humans have been co-opted by BG processes in the frontal cortex to provide WM simulation of body dynamics in support of the planning of behavior.

The midbrain motor centers include the cerebellum, the vestibular nuclei, and the red nucleus among others. The vestibular nuclei provide WM information about linear and rotary accelerations of the head and body, and provide a sense of which direction is up. The red nucleus receives WM information from the globus pallidus, the substantia nigra, the cerebellum, and BG commands from the primary motor cortex. It projects to the spinal motor centers and to the inferior olive, which is involved in cerebellar motor learning.

At the lowest level in the BG hierarchy are the final motor neurons that are located in the spinal cord and midbrain. These activate muscles to produce observable behavior.

4.2.2 Sensory (SP) Hierarchies in the Brain

At the bottom of the SP hierarchies in the brain are sensory neurons that respond to stimuli from the external world and from internal states within the body. Signals from these sensors are processed in a sensory processing hierarchy that interprets the sensory data stream. At each level in the SP hierarchy, lower level entities and events are segmented and grouped into higher-level patterns that can be named. For each named pattern, attributes and state can be computed, worth can be ascribed, class can be assigned, and characteristic behaviors can be anticipated. Named patterns can be linked to form relationships, semantic nets, and grammars. The segmentation of pixels and signals into entities, events, episodes, and situations takes place in small steps through a hierarchy of SP levels, each of which has a limited field of regard in space and time.

Segmentation and grouping are what define levels in the SP hierarchy. Segmentation is a gestalt process that partitions an image or map into figure and background. Segmentation separates pixels (or subentities) that lie on (or belong to) an entity, from pixels that do not. Grouping is the process of labeling each pixel with the name of the entity to which it belongs.

Grouping enables the computation of entity attributes and state. For example, it enables the computation of size, shape, texture, orientation, range, position, velocity, average color, and average temperature of an entity labeled x. Attributes of entity-x can be stored in the data structure named x that is located in memory at the address x. Entity-x attributes and pointers can then be recalled into working memory by sending the name x to an address decoder.

There are five major modes of sensory input: (somatosensory, vision, hearing, smell, and taste), and many hierarchical levels of SP-WM processing for each mode.

The Somatosensory Hierarchy

Somatosensory and proprioception information flows through two levels of sensory processing in the spinal cord and midbrain before entering the ventral posterior lateral (VPL) nucleus of the thalamus on the way to the anterior

parietal cortex in Brodmann's area 3 (a.k.a. S1.) Here it forms a map of the body surface. Area 3b contains somatosensory information from touch sensors in the skin, while 3a contains proprioceptive information from muscles and joints about the position and orientation of the body surface. Output from areas 3a and 3b are combined in Brodmann's area 1 (a.k.a. S2) where tactile features such as edges and texture are perceived. Output from S2 proceeds to S3 where the size, shape, and position of objects relative to the body are perceived. Output from S3 proceeds to the posterior parietal cortex where it is combined with spatial knowledge derived from vision [26].

The Visual Hierarchy

At the bottom of the vision SP hierarchy, there are photosensitive rods and cones in the retina. There are two main types of sensory neurons in the retina, large (or magna) cells that give rise to the so-called *where* pathway which is most sensitive to position and motion, and small (or parva) cells that give rise to the so-called *what* pathway which is most sensitive to color and shape. Two levels of image processing take place in the retina. These extract spatial and temporal gradients before visual information is transmitted via the optic nerve to the lateral geniculate nucleus (LGN) of the thalamus, and to the pretectum, and superior colliculus.

The visual pathway through the pretectum closes a reflex arc on the pupillary reflex to control image brightness. The visual pathway through the superior colliculus closes a visual feedback control loop that enables the eyes to smoothly track moving targets, and to saccade rapidly from one fixation point to the next. The majority of visual information goes to the LGN of the thalamus on its way to the primary visual cortex (V1). SP processes in LGN and WM processes in V1 work together to group pixels into simple entities such as edges, lines, and blobs, and compute entity attributes such as magnitude, orientation, and color.

It is estimated that at least 32 different WM representations of the egosphere reside in the visual cortex [17]. These 32 WM representations are arrayed in hierarchical levels, starting with V1 and proceeding through V2, V3, V4, and V5 [26], [15]. For each of these 32 WM representations, there is a corresponding SP process with looping interconnections to the underlying pulvinar nucleus of the thalamus. As visual information ascends this hierarchy, more complex patterns are segmented, classified, and linked in spatial-temporal relationships that represent situations and episodes. Along the way, visual information splits into two channels: one, a *where* channel that ends up in the posterior parietal cortex, and the other, a *what* channel that ends up in the inferior temporal cortex.

Hearing

At the bottom of the auditory SP hierarchy, there are acoustic sensors consisting of hair cells that detect vibrations in the cochlea. Signals from sensory neurons in the ear are processed in the cochlear nucleus, the superior olive,

and nucleus of the lateral lemniscus. These are low-level motor SP-WM modules that enable reflexive BG response to sounds. The auditory signals then proceed upward to the inferior colliculus where information is represented in a form roughly comparable to a sonogram (a map of frequency vs. time) and the direction of origin of sounds are represented on the head egosphere.

Connections between the inferior and superior colliculi enable spatial registration of sound with the visual coordinate frame in the superior colliculus. This closes a tight loop between hearing and vision for detecting the location of objects in space, and it enables the eyes to saccade to points in space from which sounds are detected.

From the inferior colliculus, the auditory signals enter the medial geniculate nucleus of the thalamus, and then to the primary auditory cortex in Brodmann's area 41 in the posterior temporal lobe. From the primary auditory cortex, the auditory information ascends through a number of layers of processing in the superior temporal cortex where it is processed into the fundamental elements of spoken language, i.e., phonemes, words, and phrases, and merged with visual objects, events, and relationships.

Smell and Taste

Sensors for smell and taste are located in the nose and tongue. Processing for smell and taste are somewhat different from the other senses. Somatosensory, visual, and auditory senses provide information about the geometry and dynamics of the external world. Smell and taste provide information about the chemical properties of things in the vicinity of the nose and mouth. The senses of smell and taste enter the limbic system and provide input for the most primitive form of decision making – whether to eat something, or not. It can be conjectured that these decision-making capabilities have evolved into much more sophisticated value judgment functions.

4.2.3 Communication between SP-WM levels

Two-way communication occurs both up and down between levels in the SP-WM hierarchies. A summary of the forward and retrograde pathways in the visual processing hierarchy is described in [40]. In general, the forward pathway up the SP-WM hierarchy conveys specific information, whereas the retrograde pathway down the SP-WM hierarchy conveys diffuse information. This suggests that the forward pathway carries specific sensory information (i.e., pixel or signal attributes in iconic form, or entity or event attributes in symbolic form), whereas the retrograde pathway carries addresses or pointers for selecting data structures or processing algorithms. In other words, processed images move up the visual processing hierarchy, while context information from higher-level SP-WM processes and from BG echelons moves down the hierarchy. This enables top-down knowledge to select models from the library of class prototypes, to generate expectations that can be compared with observations, or to select algorithms for focusing attention, segmentation, or grouping.

Address information is diffuse because many address lines must synapse on all of the neurons where information might be stored. See for example, the addressing mechanisms in the cerebellum described in [6]. Address lines elicit specific contents only when they convey a specific pattern corresponding to the location where information is stored. Contents of memory are evoked only when the proper pattern appears on the address fibers.

4.2.4 Merging of visual, auditory, and somatosensory information

Output from the unimodal processing hierarchies for visual, auditory, and somatosensory information merge – first in bimodal association cortex, and then in multimodal association cortex [35]. Output from the *what* channel of the visual processing hierarchy merges with the auditory processing hierarchy in the temporal cortex where named visual objects are associated with sounds and words. Output from the *where* channel of the visual processing hierarchy merges with the somatosensory processing hierarchy in the posterior parietal cortex where somatosensory knowledge of the body is integrated with visual knowledge about the location and motion of objects in the world.

Finally, all three space-time sensory modalities converge in the junction of the temporal, parietal, and occipital lobes, in a multimodal association region that includes Wernicke's area and lateral temporal cortex. This is where proprioception, vision, and audition come together to form a unified model the external world relative to the self [35]. Ties to the limbic system overlay this knowledge with emotional values.

The integration of these three sensory hierarchies in the multimodal association regions generates a world model that is much more than simply a reproduction of the sensory input. The world model in the posterior cortex is represented in exquisite detail, focused on the point of attention, and referenced to the inertial egosphere [3]. Entities and events are segmented from the background, overlaid with attributes, sorted into classes, assigned worth, and linked in spatial and temporal relationships. Situations, places, and episodes are endowed with emotional significance. Thus, what is presented to the frontal cortex is an integrated montage of named objects and events that are related to each other in patterns, situations, and episodes with form and meaning.

All of these perceptual processes occur subconsciously, in real-time, without apparent effort, and hence, without any conscious appreciation of the complexity of the computations that enable these phenomena.

4.3 Cortical columns

The fundamental computational unit in the cortex is the cortical column. The neocortex consists of a large two-dimensional array of several million cortical columns that are strikingly similar in form, and presumably in function as well. This huge array of cortical columns provides the massive parallelism

that underlies much of the computational power of the brain. The two-dimensional structure of this array suggests that much of the computation in the neocortex is performed in the iconic domain, or at least is closely linked to the iconic domain. Evidence for this can be seen in agnosia caused by lesions in the posterior parietal cortex. These are manifest as a deficit in body image and in the perception of spatial relationships expressed in body egosphere coordinates. Unilateral lesions in specific regions of posterior parietal cortex result in patients that not only ignore specific regions of their body, but specific regions of the outside world, both in the observed world and in the world visualized in the imagination [26].

There are two types of cortical columns: micro-columns that contain about 100 neurons each, and hypercolumns that are compact collections of about 100 microcolumns [18]. Microcolumns consist of neurons that detect the presence or absence of a particular attribute, such as the orientation of a line or edge within the region on the egosphere covered by the parent hypercolumn. A hypercolumn consists of a set of microcolumns that provide information about a particular pixel or region of space or time.

4.4 Thalamocortical loops

Each cortical column is tightly coupled with an underlying thalamic nucleus in a thalamocortical loop. However, the thalamocortical loops in the frontal cortex are significantly different from those in the posterior cortex. This is illustrated in Figure 7. Loops in the posterior cortex implement windowing, segmentation, grouping, recursive estimation, and classification. Those in the frontal cortex implement planning and control of behavior.

4.4.1 Posterior Thalamocortical Loops

In the posterior cortex, all sensory input reaching the cortex flows through the posterior thalamus first. Visual information from the eyes flows through the lateral geniculate nucleus (LGN) of the thalamus before entering the primary visual cortex. Auditory information from the ears flows through the medial geniculate nucleus (MGN) before entering the primary auditory cortex. Somatosensory information from the skin, joints, tendons, and vestibular system flows through the ventral posterior lateral nucleus (VPL) on its way to the primary somatosensory cortex. After reaching the cortex, sensory information loops back to the thalamus. V1 loops back to the LGN. The somatosensory cortex loops back to the VPL. The auditory cortex loops back to the MGN.

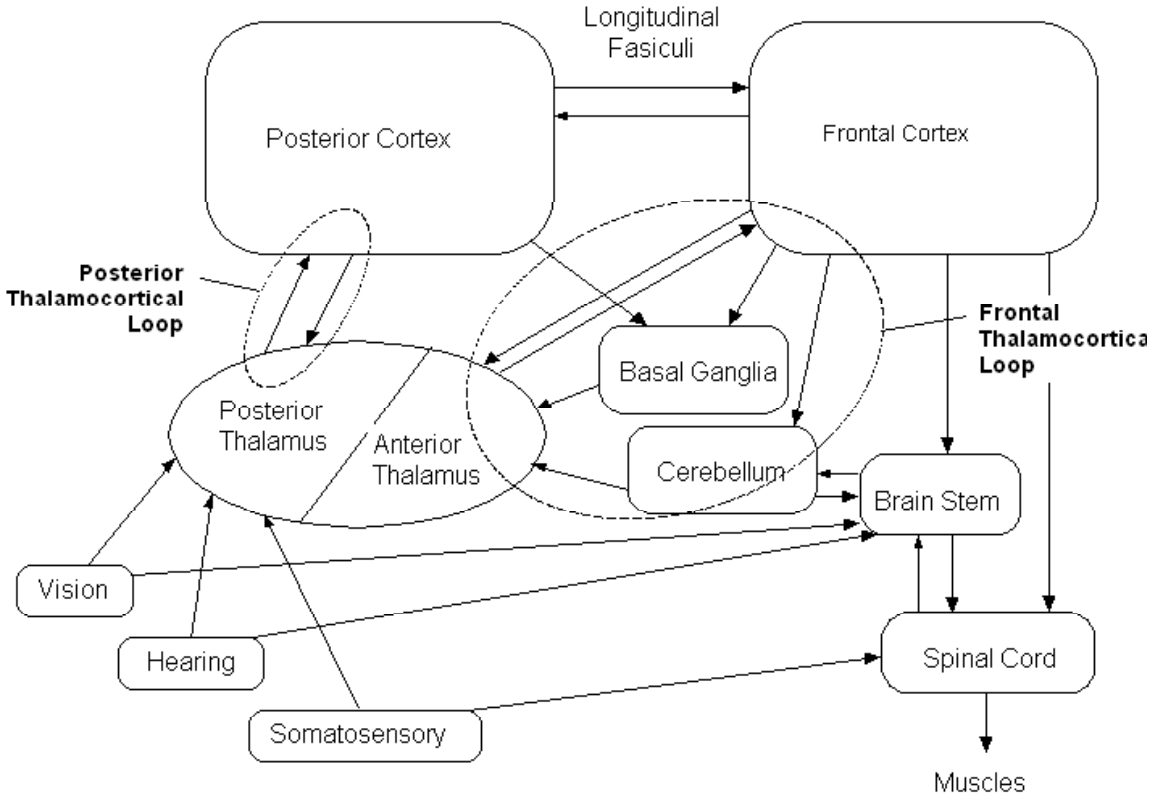

Figure 7: Thalamocortical loops. (redrawn from [22] and [26]).

However, it is not just the primary sensory cortex that loops back to the thalamus. At all levels throughout the occipital, parietal, and temporal cortices, in both bimodal and multimodal association cortices, there are massive looping connections between the thalamus and the cortex. In the higher levels of the SP-WM hierarchy, thalamocortical loops involve the pulvinar, the VPL, the dorsal medial, the lateral dorsal, and lateral posterior nuclei of the thalamus [22], [26], [24].

This looping circuitry enables windowing, segmentation, grouping, recursive estimation, and classification at all levels in the SP-WM hierarchy. Observed signals, images, entities, and events are compared with predicted signals, images, entities, and events, so that estimated models in the posterior cortex can be updated every thalamocortical loop cycle.

Thalamocortical loops involve four fundamentally different types of connections between the thalamus and cortex. The first derives from "core" cells in the thalamus. These receive data input from the sensory periphery (or lower levels in the SP-WM hierarchy) and send specific topographic projections to the cortex. These specific projections activate neurons largely in layer IV and the lower part of layer III in the cortical columns. This activation radiates from the middle layers (IV) to the superficial layers (II and III) and then to the deep

layers (V and VI) [24]. It is hypothesized that these core thalamic neurons provide data in the form of images, attributes, and state variables.

A second set of connections from thalamus to cortex derives from "matrix" neurons in the thalamus. These matrix neurons are driven largely by feedback from deep layers (V and VI) in the cortex [16]. Matrix neurons project diffusely to layer I of the cortex where they terminate on the apical dendrites of neurons in layers II, III, and V. The nature of these connections strongly suggests mechanisms by which the thalamus can address data structures in the cortex to select models for recursive estimation and classification. It is hypothesized that the matrix neurons provide addresses in the form of pointers to locations where models and algorithms are stored in the cortex.

The third type of connections between thalamus and cortex returns specific topographic projections from the cortex back to the same core cells from which the specific inputs originated. This feedback path enables comparison of model-based predictions with sensory observations. It is hypothesized that this feedback contains predicted attributes and state variables to be compared with observed attributes and state variables. Specific feedback from cortex to thalamus could also be used to define windows of attention.

A fourth type of connections returns diffuse projections from the cortex to the thalamus. It is hypothesized that this feedback provides addressing information whereby the cortex can select processing procedures and algorithms for comparing predictions with observations, and for segmentation and grouping. Granger [22] suggests a process by which the thalamocortical loops can perform both grouping into general categories, and segmentation into subcategories.

In short, it is hypothesized that the SP processes that perform windowing and grouping are located in the thalamus, along with SP processes that compare WM predictions with sensory observations; and that the KD knowledge structures that store attributes, state, worth, class, relationship pointers, and expected behavior of entities and events are located in the posterior cortex along with the WM processes that generate model-based predictions. It is further hypothesized that model-based predictions generated in the cortex are returned to the thalamus to be compared with observed data from sensory input. Variance between predictions and observations is forwarded to the cortex via specific inputs to update the model in the cortex. Small differences are returned to the cortex as specific inputs to update the model. Large differences reduce the confidence level to the point where the cortex rejects the model and searches for an alternative model. The degree of similarity between predictions and observations can be used to assign levels of confidence in the related VJ processes. Comparison between stored class prototypes and estimated entity attributes enables classification.

At all levels in the SP-WM hierarchy, interactions between the cortex and the thalamus play a similar role. Predictions generated by WM processes in the cortex are compared in the thalamus with input from lower level SP-WM

processes. Thus, what is known about posterior thalamocortical loops maps well onto the RCS model of recursive estimation described above in section 3.4.

4.4.2 Anterior Thalamocortical Loops

A different type of looping connection exists between the thalamus and the frontal cortex. In the primary motor and premotor cortical regions, thalamocortical loops travel from the cortex to the basal ganglia (caudate, putamen, globus pallidus, and subthalamic nuclei), the substantia nigra, and the cerebellum, before returning through the thalamus (ventrolateral, anteroventral, or center median nuclei) to the cortical regions from which they originated. [19], [26].

These loops appear to be involved in decision-making and planning. It is hypothesized that goals (i.e., desired states of the world) and hypothesized plans to achieve them are generated in the frontal BG cortex based on knowledge of the world provided by connections with the posterior WM cortex, and knowledge of internal drives and states generated by the hypothalamus and the autonomic nervous system. BG goals and plans are then sent to the basal ganglia, substantia nigra, and cerebellum for dynamic and kinematic modeling of the body. The predicted results are then transmitted to the thalamus where they are compared with the desired results provided by direct input from the cortex. Predicted results are also sent to the limbic system for evaluation of cost, risk, and benefit. Differences between the desired results and predicted results are computed in the thalamus and returned to the cortex for improving or modifying the plan. If the differences between desired goals and predicted results are small, and the VJ evaluation is positive, the plan is accepted and sent to the next lower echelon in the BG hierarchy. If the differences are large, or the VJ evaluation is negative, the BG cortex may generate an alternative plan to achieve the goal, or select a different goal because the predicted result is not worth the cost or risk of achieving the goal. This maps well onto the RCS planning processes described in section 3.3.

There are similarities as well as differences in the relationships that exist between the cortex and the thalamus in the posterior and anterior regions of the brain. Similarities are that both posterior and anterior cortices generate models of the world. In the anterior cortex, the models generated are *desired* states of the world. In the posterior cortex, the models are *estimated* states of the world. In both regions, the thalamus compares models with situations, either observed or predicted. The variances between models and situations are returned to the cortex where the models may be revised.

Differences between thalamocortical loops in posterior and frontal cortices are related to the source of the lower level input to the thalamus, and in the meaning of the measured variance. In the posterior regions of the brain, the lower level input to the thalamus consists of data flowing up the sensory processing hierarchy, and the variance is used to update the world model. In the frontal regions of the brain, the lower level input to the thalamus consists of pre-

dicted results of planned actions simulated in the basal ganglia, and cerebellum. Here, the variance is used to search the space of possible futures for a good plan.

In both posterior and anterior regions, VJ processes in the related limbic system evaluate the goodness or badness of the model, and the level of confidence that should be assigned to it. In the frontal brain, VJ processes evaluate the costs and benefits of planned behavior, and the variance between what is desired and what is predicted. In the posterior brain, VJ processes evaluate the worth of the estimated state of the world, and the variance between what is observed and predicted.

Within the cortex itself, addresses from higher level cortical regions provide top-down information needed for coordination and control. In the frontal cortex, top-down addresses can take the form of commands that select procedures and parameters for task decomposition and planning. In the posterior cortex, top-down addresses provide context needed to select algorithms for focus of attention, segmentation, grouping, and classification. In both cases, variance between top-down and bottom-up data can be used to update the model.

5 Modifications Required in the RCS Reference Model

Although many features of the RCS reference model map directly onto the architecture of the human brain, there are at least three important ways in which the drawing in Figure 4 needs to be modified to map well onto the brain.

First, the relative size of the nodes near the top of the hierarchy in Figure 4 gives the impression that the sensory-motor hierarchy is a classical pyramid that converges to a small localized command center at the top. Quite the opposite. To the extent that the SP-WM-VJ-BG hierarchy in the brain is a pyramid, it is inverted, with the largest number of neurons at the top [19]. The prefrontal cortex, which resides at the top of the BG hierarchy, is one of the largest and most complex structures in the human brain. The prefrontal cortex is much larger than the premotor cortex, and the premotor cortex is larger than the primary motor cortex. The midbrain BG centers are smaller still, and the BG processes in the spinal cord contain by far the smallest number of neurons.

Similarly in the SP-WM hierarchies, the number of sensory neurons for touch, proprioception, vision, and hearing is less than the number of neurons in the primary sensory cortex. The primary cortical sensory areas are small compared to the unimodal association areas, and the unimodal association areas are small compared to the multimodal association areas. The primary sensory areas for taste and smell are small compared to the subcortical limbic centers, and these are small compared to the limbic cortex. Thus, the relative size of the RCS nodes at the uppers levels should be enlarged to properly represent the architecture of the brain.

Second, the RCS diagram in Figure 4 needs to show the WM as physically separated into two parts above the first two or three levels: one part supports SP in the posterior cortex and models the external world, while the other part supports BG in the frontal cortex and models the internal dynamics of the body. Figure 4 suggests that SP-WM-VJ-BG processes are physically adjacent at all levels and echelons. This is a useful theoretical concept, but in the brain, it is physically true only at the lowest levels in the spinal cord, and to a lesser degree in the midbrain and primary sensory-motor cortex. Above the primary sensory-motor cortex, the sensory processing and behavior generating processes migrate further and further apart, and the WM processes that support them do so as well. The top of the BG-WM hierarchy lies all the way in the front in the prefrontal cortex, while the top of the SP-WM hierarchy resides near the back around the junction of the parietal, temporal, and occipital cortices. Thus, the SP-WM-VJ-BG nodes at the higher levels and echelons in the brain are not compact, but are distributed over anatomically distant parts of the brain. To map well onto the back-to-front architecture of the brain, the upper level RCS nodes need to be split into frontal BG-WM and posterior SP-WM components.

Communications between BG-WM and SP-WM are maintained at all levels through massive front-to-back axonal communication pathways of the longitudinal fasiculus, the arcuate fasciculus, the occipito-frontal fasciculus, and the cingulum. Thus, the close functional relationship between SP-WM-VJ-BG processes implied in the RCS nodes is conceptually valid, but the neural structures in which these processes occur are not physically located close together except in the spinal cord, midbrain, and primary somatosensory cortex.

Third, Figure 4 needs to be modified so as not to imply a one-to-one correspondence between SP-WM levels and BG-WM echelons. The number of SP levels is not necessarily the same as the number of BG echelons (neither in RCS nor in the brain.) In the brain, as in RCS, the one-to-one correspondence between SP levels and BG echelons holds only at the lowest two or three levels in the hierarchy. In fact, the SP-WM hierarchy is in some sense orthogonal to the BG-WM hierarchy. BG-WM is a hierarchy of task decomposition, while SP-WM is a hierarchy of spatial-temporal grouping. The number of SP-WM levels is not equal to the number of BG-WM echelons, and the number of SP-WM levels is not the same for all sensory modalities.

To address these issues, Figure 4 has been redrawn in a form that more closely maps onto the human brain. The result is shown in Figure 8.

In the posterior cortex, the relationships between SP and WM processes are drawn as cylinders, to indicate the tight thalamocortical looping interaction between them. It is hypothesized that the posterior WM processes reside in

the cortex and the SP processes in the underlying thalamic nuclei. Thus, the posterior SP-WM cylinders are shown with WM on the top.

In the frontal cortex, the relationships between BG and WM are also drawn as cylinders, with BG on the top to indicate that the BG processes reside in the cortex. Three levels of BG-WM are shown in the prefrontal cortex to indicate that there may be multiple levels of abstract thought that take place in this large and poorly understood region. In the prefrontal cortex, WM support for long-range planning may reside in the prefrontal association areas and in the

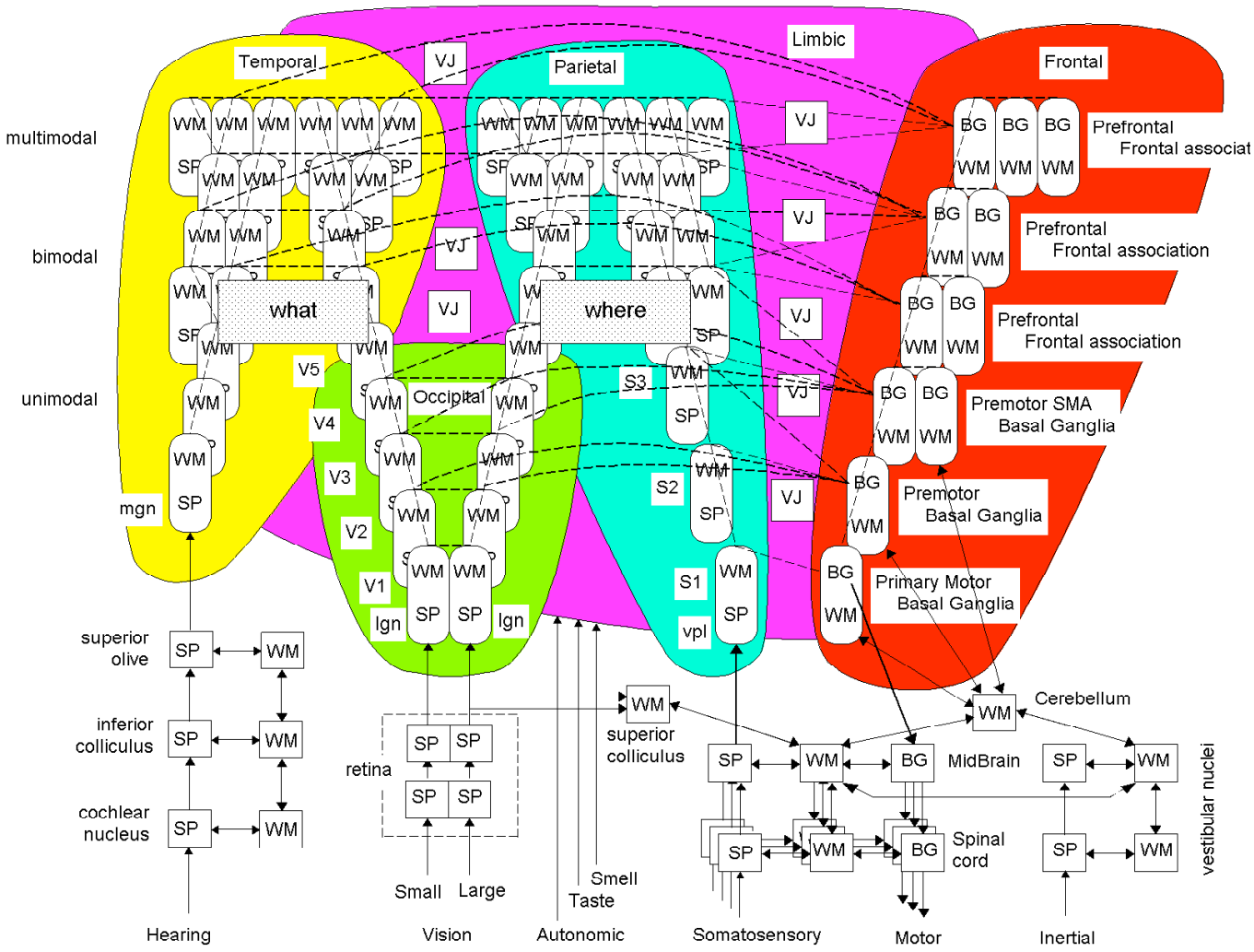

Figure 8: RCS redrawn to map onto the anatomy of the central nervous system.

posterior cortex. High-level decisions and plans generated in the prefrontal cortex are transmitted to the premotor cortex where they are decomposed into mid-level commands to the primary motor cortex. Both premotor and primary motor cortices use the basal ganglia and cerebellum for kinematic and dynamic planning. All areas of the cortex are supported by the underlying thalamic nuclei.

Task decomposition continues in the midbrain and spinal cord until low-level commands are issued to the muscles. Subcortical SP-WM-BG processes are represented in a more conventional RCS format at the bottom of Figure 8.

The massive front-to-back axonal communication pathways that connect the external world model in the posterior cortex to the decision-making and planning processes in the frontal cortex are shown in Figure 8 as dotted lines. At all levels, in both frontal and posterior cortex, VJ processes provide evaluations. In the posterior brain, VJ processes evaluate situations and episodes. In the frontal brain, VJ processes evaluate plans and behaviors.

Sensory input enters Figure 8 from the bottom. Smell, taste, and input from the autonomic nervous system are shown entering the limbic system at the bottom center of Figure 8. Vision, hearing, and somatosensory data streams undergo two or three levels of SP-WM before reaching the thalamus, and several more levels of SP-WM in the unimodal association cortex before entering the multimodal association areas. Somatosensory and visual data streams merge in the posterior parietal cortex where representations of space are generated. This is the *where* channel. Visual and auditory processing streams merge in the anterior temporal cortex where pointers are established that link visual entities to auditory events, and vice versa. This is the *what* channel. Finally, visual, auditory, and spatial representations are merged in the junction of the parietal, temporal, and occipital cortices where a fully annotated model of the world is represented. Output from all of these multimodal association areas is forwarded to the frontal cortex and basal ganglia.

In both posterior and frontal cortical regions, the cortex stores and uses models of the world. In the frontal regions, the models are used to select goals, make decisions, construct plans, and control behavior. In the posterior regions, the models are used to focus attention, segment entities and events, estimate state-variables and attributes, define relationships, and sort things into classes. In both front and back, the basic function performed by the cortex is the same – namely to build, maintain, store, and compute with models of the world. In the posterior brain, the underlying thalamic nuclei compare sensory input with world model predictions, and forward the differences to update the model. In the frontal brain, the underlying thalamic nuclei compare desired results (goals) with predicted results of hypothesized plans, and send the variance to VJ for plan evaluation. In both front and back, the thalamus may provide the timing required to synchronize addressing mechanisms in both cortex and thalamus.

At each level, the somatosensory, visual, and hearing systems build an internal model that is both iconic and symbolic. The model is iconic in that objects occupy regions on the egosphere. The model is symbolic in that entities on the egosphere are segmented from the background and given names and assigned attributes such as shape, size, position, orientation, velocity, color, and texture. Named entities can then be sorted into classes, assigned worth, associated with behavior, and linked into syntactic and semantic networks. The resulting world model can be used to predict how the environment will

evolve and how named objects will behave, both in the short term for recursive estimation and in the longer term for decision-making, and planning.

6 Summary

In this paper, the RCS reference model architecture for intelligent systems has been described and mapped onto the physical structure of the brain. Both the RCS architecture and the brain are hierarchical, with layers of interconnected computational modules that generate functionality of sensory processing, world modeling, value judgment, and behavior generation. At the lower layers, these processes generate goal-seeking reactive behavior. At higher layers, they enable perception, cognition, reasoning, imagination, and long-term planning. Within each hierarchical layer, the range and resolution in time and space is limited. At low layers, range is short and resolution is high, whereas at high layers, range is long and resolution is low. This enables high precision and quick response to be achieved at low layers over short intervals of time and space, while long-range plans and abstract concepts can be formulated at high layers over broad regions of space and time.

The RCS reference model consists of a set of computational processes that are well understood and well defined. These can be implemented by a variety of methods, including differential equations, finite state automata, production systems, Bayesian logic, predicate calculus, semantic nets, linguistic grammars, computer algorithms, matrix operations, and arithmetic formulae. They store information in data structures that can be modeled by addressable memory locations that contain state-variables that can be organized into vectors, strings, arrays, lists, frames, objects, classes, agents, commands, schema, and plans; and these can be linked by pointers to represent relationships, situations, and episodes. Each of these computational processes and data structures can, in principle, be implemented by neural networks.

The mind consists of a set of computational processes that are less well understood and less well defined. The brain computes using neurons, dendrites, synapses, and active membranes as computational components. It employs axons, action potentials, transmitter chemicals, and hormones for communication. These computational mechanisms are organized through genetic design, maturation, and learning experiences into a hierarchical web of modules, each of which contains networks, recurrent loops, arrays of cortical columns, and subcortical nuclei, each of which contains hundreds or thousands of neurons. The brain consists of a hierarchy of massively parallel computational modules and data structures interconnected by information pathways that enable analysis of the past, estimation of the present, and prediction of the future.

The mapping of RCS onto the brain provides many insights into computational mechanisms in the brain. It suggests a number of testable hypotheses regarding where the computational mechanisms of sensory processing, world modeling, behavior generation, and value judgment reside in the brain. It suggests how various forms of knowledge could be represented in the neural

architecture, and hints at what types of messages might be conveyed over the various neural pathways.

For example, it is hypothesized that thalamocortical loops in the posterior cortex provide focus of attention, segmentation, grouping, recursive estimation, and classification of entities and events. Exactly how this is achieved by the known axonal pathways and synaptic connections remains a topic for research. Communications with the limbic system provide for assessment of worth, attractiveness, and emotional value. Exactly how these state variables are encoded in the neuropile is as yet unknown. In the frontal cortex, thalamocortical loops provide computational mechanisms for decision-making, planning, and control of behavior. Connections with the limbic system provide for evaluation of cost, risk, and benefit of current and future actions. Diffuse fiber pathways convey addresses. Specific fiber pathways convey data. The mapping of RCS reference model onto the brain suggests how these computational processes might be integrated into an intelligent system that is aware of itself and its situation in the world. But this is only a hypothesis, which hopefully can be tested in the near future.

RCS theory predicts that what is perceived is far richer and more complex than what is sensed, i.e., that most of what is perceived is a hypothetical model of the world constructed by cognitive processes to explain and predict the sensory input. RCS theory predicts that those areas of the lateral prefrontal cortex that are thought to give rise to consciousness perceive the world model in the posterior cortex as a hi-fidelity real-time diorama that is focused at a point of attention, referenced to inertial space, and filled with dynamic patterns that are segmented into named entities and events. These entities and events are associated with attributes, state, and worth; and they are sorted into classes with prototypical properties and behaviors. These are linked by a network of pointers that denote spatial, temporal, and causal relationships that comprise places, situations, and episodes. RCS theory predicts that this entire representation is projected back onto the visual and tactile images, so that entities and events in the world appear to have identity and meaning.

The RCS reference model architecture provides a theoretical framework and an engineering methodology for building and testing computational models of mental processes. Some of these models are testable with current technology. All should be testable in the foreseeable future given the rate of technological progress in electronics, computer science, and neuroscience. No claim is made that RCS is the only architecture, or the only methodology, or even the best. Simply, that it is has been used by a number of engineers and researchers for building complex intelligent systems.

7 Conclusions

One might ask, “If this computational model is capable of explaining the phenomena of mind, why does it not act like a human being? Why is it not as good as a two year old child in face recognition? Why can’t it tie a shoe? Why can’t it carry on a conversation in natural language?”

The answer is two fold: First of all, a computational model of mind does not yet exist –except as a goal. This paper does not offer a computational model capable of explaining the mind. It only asserts that such a model is possible, and suggests a research direction, which if followed might someday lead to that goal. Secondly, even after a computational theory does exist, an enormous engineering effort will be required to achieve anything close to human levels of performance in complex tasks of perception, dexterous manipulation, and conversational language.

Human hand-eye coordination is one of the marvels of nature. Each hand has about 26 degrees of freedom controlled by thousands of muscles. It contains thousands of sensors that measure position, velocity, and tension in muscles and tendons. The hand is covered with skin that contains tens of thousands of sensors that measure touch, pressure, vibration, temperature, and pain. The hand is attached to a wrist, an arm, a shoulder, a neck, and a head that contains two eyes. To tie a shoe, the visual cortex must have image processing mechanisms that enable the brain to interpret a particular subset of the millions of neural impulses flowing in the optic nerve as an *object* that is a member of a *class* labeled "shoelace_visual." The visual world model must contain data structures that represent relationships between the shoelace, the shoe, and the fingers. In the somatosensory cortex, there must be mechanisms that interpret a particular subset of the millions of neural impulses flowing in the spinal cord as an *object* that is a member of a *class* "shoelace_tactile." Then there must be computational mechanisms in the posterior parietal cortex that fuse the tactile world model with the visual world model. Finally, there must be behavior generating mechanisms in the frontal regions of the brain that access libraries of motor skills related to shoelace tying, combine these with situational parameters from the integrated tactile/visual world model, and generate strings of commands to the muscles in two arms, two hands, and ten fingers to manipulate the shoelace while tracking what is actually happening with the eyes.

Even after a scientific theory is formulated and tested, it typically requires an enormous investment of funding and engineering talent to implement what is theoretically possible. Recall that the fundamental theory of celestial mechanics was developed by Issac Newton in the 1600s, and the theory of rocket propulsion was developed in the 1920s and 30s by Robert Goddard. But it took an investment of $20 billion dollars (in 1960 currency) and a national effort over a decade to place a man on the moon. It may take a comparable effort to create a human-equivalent robot here on earth.

We are at least a decade from a widely accepted computational theory of mind. Some argue it will take many decades, perhaps even centuries. Others claim that the mind will forever elude scientific explanation. And even after a widely accepted theory of mind is achieved, it will require many years and huge investments in engineering development to achieve intelligent machines that rival human capabilities in dexterous manipulation.

In only three technology areas –video cameras, computers, and vehicles– have the requisite investments been made. In one application area –unmanned vehicles– a few hundreds of millions of dollars have been invested, mostly by the military. It is here that some promising results have begun to emerge. Air, ground, and undersea vehicles require the control of only a few degrees of freedom, and the theory required to do this is mature. With the advent of LADAR cameras, the image processing technology required for ground vehicles to perceive the road and to detect and track other vehicles and pedestrians on and near the roadway is within reach. As a result, automatic cruise control and collision avoidance systems are on the drawing boards, and probably will appear on military, commercial, and private vehicles within a few years.

However, building a computational model of the mind is a much more ambitious goal. The mind is a phenomenon that is observed only in the brain. The human brain is a massively parallel, hierarchical structure with multiple loops, made up of more than 100 billion tiny computers, each of which performs a complex non-linear computational operation on thousands of synaptic inputs a few hundred times per second. Each neuron produces an output that is conveyed to many hundreds of other neurons at a rate of several hundred bytes per second. The brain receives input from millions of sensors, and computes output for hundreds of thousands of muscles. There are more than a million photosensors in each eye. There are hundreds of thousands of sensors in the ears and vestibular system. There are many tens of thousand sensors in the skin, muscles, and tendons. There are thousands of sensors in the smell and taste organs. There are many thousands of motor neurons that enervate thousands of muscle groups.

I doubt that machines can achieve human level cognitive capabilities in perception, cognition, and dexterous manipulation until they can emulate the level of complexity, sophistication, and massive parallelism that exists in the human brain. I believe we must at least understand the computational processes that take place in the human brain – at scale and in real-time – before we can hope to understand the mechanisms of mind.

Within a decade or two, computational power of small, moderately priced computers will enter the realm of hundreds of teraflops. Knowledge of the structure and function of the brain will increase significantly. Understanding of cognitive architectures will improve, and sensor technology will enable construction of real-time spatial-temporal models of entities, events, relationships, situations, and episodes in the world. If adequate funds are directed toward research and development, we may see computational systems in which the outlines of what can only be called "mind" will emerge.

In the mean time, we can predict that research on intelligent systems will yield important insights into the phenomena of attention, gestalt grouping, filtering, classification, visualization, reasoning, communication, intention, motivation, and subjective experience. We currently know how to build systems that pursue goals, simulate the future, make decisions, formulate plans, and react to

what they see, feel, and hear, smell, and taste. It is not unreasonable to expect that at some point, when engineered systems begin to approach the sophistication and complexity of the human brain in sensing, perception, cognition, reasoning, planning, and control of behavior, at least some elements of mind will emerge. At that point, machines may begin to behave as if they are sentient, knowing, individuals motivated by hope, fear, pain, pleasure, aggression, curiosity, and operational priorities.

Certainly there is much about the mind that will remain a mystery for decades, perhaps even centuries. Aspects of mind such as a sense of justice, honor, duty, reverence, beauty, wonder, and religious experience may remain the purview of philosophy for a very long time. But there is much that is yielding to the computational approach. This is a frontier that is open to scientific inquiry, and if this frontier is aggressively pursued, it seems likely that progress will be made along the road toward a computational theory of mind. And surely, practical applications of importance to military, commercial, and personal, users will fall out along the way.

References

[1] Albus, J. S., et al. (2002) "4D/RCS Version 2.0: A Reference Model Architecture for Unmanned Vehicle Systems," NISTIR 6910, National Institute of Standards and Technology, Gaithersburg, MD, 2002.

[2] Albus, J. S., Meystel, A. (2001) *Engineering of Mind: An Introduction to the Science of Intelligent Systems*, Wiley, New York

[3] Albus, J.S. (1991) "Outline for a Theory of Intelligence," *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 21, No. 3, pgs. 473-509, May/June

[4] Albus, J.S. (1975a) "A New Approach to Manipulator Control: The Cerebellar Model Articulation Controller (CMAC)," *Transactions of the ASME Journal of Dynamic Systems, Measurement, and Control*, September pp. 220-227

[5] Albus, J.S., (1975b) "Data Storage in the Cerebellar Model Articulation Controller (CMAC)," *Transactions of the ASME Journal of Dynamic Systems, Measurement, and Control*, September pp. 228-233

[6] Albus, J.S. (1971), "A Theory of Cerebellar Function," *Mathematical Biosciences*, 10, pp. 25-61

[7] Albus, J. S., and Barbera, A. J. (2004) "RCS: A Cognitive Architecture for Intelligent Multi-Agent Systems," Proceedings of the 5th IFAC/EURON Symposium on Intelligent Autonomous Vehicles, IAV 2004, Lisbon, Portugal, July 5-7

[8] Anderson, J. R. (1993) *Rules of the Mind*. Erlbaum, Hillsdale, NJ.

[9] Arkin, R.C. and Balch, T. (1997) "AuRA: Principles and Practice in Review", *Journal of Experimental and Theoretical Artificial Intelligence*, Vol. 9, No. 2, pp. 175-189.

[10] Brooks, R.A. (1986), "A Robust Layered Control System for a Mobile Robot", *IEEE Journal of Robotics and Automation*, RA-2, 14-23

[11] Brooks, R.A. (1999), *Cambrian Intelligence: The Early History of the New AI*, MIT Press, Cambridge, Mass.

[12] Churchland, P. S. (2002) *Brain-Wise: Studies in Neurophilosophy*, MIT Press, Cambridge, Mass.

[13] Churchland, P. S and Sejnowski, T. J. (1992) *The Computational Brain*, MIT Press, Cambridge, Mass.

[14] Damasio, A. R. (1999) *The Feeling of What Happens*. Harcourt Brace, New York

[15] DeYoe, E. A., and Van Essen, D. C. (1988) "Concurrent processing streams in monkey visual cortex." *Trends in Neuroscience* **11**:219-226

[16] Diamond, M., Armstrong-James, M., and Ebner, F. (1992) "Somatic sensory responses in the rostral sector of the posterior group (POm) and in the ventral posterior medial nucleus (VPM) of the rat thalamus." *J. Comp. Neurol*. **318**:462-476

[17] Felleman, D. J., and Van Essen, D. C. (1991) "Distributed hierarchical processing in primate visual cortex. *Cerebral Cortex*, **1**, pp. 1-47

[18] Fuster, J. M. (2003) *Cortex and Mind: Unifying Cognition*, Oxford University Press, Oxford

[19] Fuster, J. M. (1995) *Memory in the Cerebral Cortex: An Empirical Approach to Neural Networks in the Human and Nonhuman Primate*, MIT Press, Cambridge, Mass.

[20] Ghallab, M., Nau, D., and Traverso, P. (2004) *Automated Planning: Theory and Practice*. Morgan Kaufmann, Boston

[21] Gazi, V., Moore, M.L., Passino, K.M., Shackleford, W.P., Proctor, F.M., and Albus, J.S. (2001) *The RCS Handbook: Tools for Real Time Control Systems Software Development*, John Wiley & Sons, NY

[22] Granger, Richard, (2005) "Engines of the brain: The computational instruction set of human cognition." *AI Magazine*

[23] Grossberg, S., Finkel, L., Field, D. (eds.) (2004) *Vision and Brain: How the Brain Sees*. Elsevier, London, UK

[24] Hawkins, J. (2004) *On Intelligence*. Times Books, Henry Holt & Co., New York

[25] Hayes-Roth, B. (1995) "An architecture for adaptive intelligent systems." *Artificial Intelligence*, 72

[26] Kandel, Eric; Schwartz, James; Jessell, Thomas (1995) *Essentials of Neural Science and Behavior*, McGraw-Hill, New York

[27] Koch, Christof (2004) *The Quest for Consciousness: A Neurobiological Approach*. Roberts & Co. Publishers, Englewood, CO

[28] Kortenkamp, D., Bonasso, R. and Murphy, R. (eds.) (1997) *AI-based Mobile Robots: Case Studies of Successful Robot Systems*. MIT Press, Cambridge

[29] Kurzweil, R. (2006) *The Singularity is Near*. Viking Press

[30] Laird, J., Newell, A., and Rosenbloom, P. (1987) SOAR: An Architecture for General Intelligence, *Artificial Intelligence*, 33, pp. 1-64

[31] Latombe, J. C. (1991) *Robot Motion Planning*. Kluwer Academic Publishers

[32] Madhavan, R., Messina, E., and Albus, J. S. (Eds.) (2007) *Intelligent Vehicle Systems: A 4D/RCS approach*, NOVA Books

[33] Marr, D. (1969), "A Theory of Cerebellar Cortex", *Journal of Physiology*, (London), 202, pp. 437-470

[34] McHale, J. (2006) "Robots are fearless." Military & Aerospace Electronics, June http:// mae.pennnet.com/ Articles/ Article_Display.cfm? Section=ARTCL& ARTICLE_ID=258264& VERSION_NUM=2&p=32&pc=ENL.

[35] Mesulam, M. M. (1998) "From sensation to cognition." *Brain* 121:1013-1052

[36] Miller, W. T. (1994) Real-time application of neural networks for sensor-based control of robots with vision. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 19, pp. 825-831

[37] Moravec, H. (1999) *Robot: From Mere Machine to Transcendent Mind*. Oxford University Press, New York

[38] Nilsson, N. J. (1998) *Artificial Intelligence: A new synthesis*. Morgan Kauffman: San Francisco

[39] NIST web sites http://www.isd.mel.nist.gov/projects/rcslib/, http://www.isd.mel.nist.gov/ projects/ rcs/index.html, http://sourceforge.net/projects/usarsim, http://sourceforge.net/ projects/moast.

[40] Pandya, D. N., and Yeterian, E. H. (1985) "Architecture and connections of cortical association areas." In: *Cerebral Cortex*, Vol. 4, pg. 3-61, (ed. A. Peters and E.G. Jones) Plenum Press, New York.

[41] Shoemaker, C., Bornstein, J., Myers, S., and Brendle, B. (1999) "Demo III: Department of Defense testbed for unmanned ground mobility," SPIE Conference on Unmanned Ground Vehicle Technology, SPIE Vol. 3693, Orlando, FA, April

[42] Smith, P. J., Geddes, N. D., (2002) "A cognitive systems engineering approach to the design of decision support systems." In *The human-computer interaction handbook: fundamentals, evolving technologies, and emerging applications*. (Ed. Jacko, J. A., and Sears, A.) Erlbaum, Mahwah, NJ

[43] Stanford Encyclopedia of Philosophy (2003) http://plato.stanford.edu/entries/computational-mind/#notions

[44] Truex, R. C., and Carpenter, M. B. (1969) *Human Neuroanatomy*. Williams and Wilkins Co., Baltimore

[45] Volpe, R., Nesnas, I., Estlin, T., Mutz, D., Petras, R., and Das, H. (2001) "The claraty architecture for robotic autonomy." Proceedings of the 2001 IEEE Aerospace Conference, Big Sky, Montana, March

[46] Winston, P. H. (1992) *Artificial Intelligence*, Addison-Wesley, Reading, Mass.

[47] Albus, J.S., Bostelman, R.V., Hong, T.H., Chang, T., Shackleford, W., Shneier, M.O. (2006) The LAGR Project. Integrating learning into the 4D/RCS Control Hierarchy, Proceedings of ICINCO 06 International Conference in Control, Automation and Robotics, Setubal, Portugal, August 2006.

[48] Pinker, S. (1997) *How the Mind Works*, W.W. Norton, New York

Acknowledgements

The material represented in this paper is the result of thirty years of support from a variety of sources. The Manufacturing Engineering Laboratory of the National Institute of Standards and Technology has provided the staff, facilities, and the freedom to pursue this large subject over the entire three decades. The Army Research Laboratory, Robotics Office under Mr. Charles Shoemaker and Dr. Jon Bornstein have supplied continuous funding support for applying RCS to unmanned ground vehicles for more than 20 years. During the 1980's, the Navy Manufacturing Technology program provided funding for the Automated Manufacturing Research Facility, Mr. Jack McInnis, PM. DARPA has provided support through several programs over the years, starting with the MAUV (Multiple Autonomous Underwater Vehicle) program, Dr. Tony Tether, PM. DARPA funding has also been provided for Submarine Operational Automation Systems, Dr. Ed Carapezza, PM; the MARS (Mobile Autonomous Robot Software) program, Dr. Mark Swinson and Dr. Doug Gage, PMs. Current DARPA support is from the LAGR (Learning Applied to Ground Robotics) program, Dr. Larry Jackel, PM; and the Persistant Ocean Surviellance program, Dr. Ed Carapezza, PM. The

immediate impetus for writing this manuscript was provided by the DARPA BICA (Biologically Inspired Cognitive Architectures) program, Dr. David Gunning, PM.

The Mind as an Evolving Anticipative Capability

Ron Cottam, Willy Ranson and Roger Vounckx

The Evolutionary Processing Group, Vrije Universiteit Brussel

Abstract

René Descartes is habitually associated with the fundamentality of a *categorical* distinction between *mind* and *matter* [1]. Contrarily, Terrence Deacon has described our self-experience, not as a (static) *category* but as a *process*: as "*what we should expect an evolutionary process to feel like*" [2]. 'Modernistic' Darwinism would maintain that the primary character of evolution is genetic-mutational randomness. But where, then, does the mind's apparently directed causality of free will come from? Is evolution indeed random? In the light of early 21st century genetics we will question the attribution of environmentally-isolated randomness to evolutionary mutation. We submit that evolution has itself evolved from 'Darwinian' atemporal randomness towards anticipative awareness, auto-catalyzed by *Anticipative Capability*, which both drives the evolution and bounds it. We consequently argue that the evolutions of survivability, anticipation, consciousness, intelligence, wisdom, evolution *itself*, and indeed *the mind* are broadly equivalent. We reject the anthropomorphically convenient categorical separation of entities into 'living' and 'non-living', and note that the manifestation of 'life' indicates a continuity of evolvability and *Anticipative Capability* between blind inanimate dependence on Newton's Laws and human technological control. We derive definitions of *intelligence*, *sapience* and *wisdom* from the multiscalar properties of *birational* hierarchical information-processing, and point out the relevance of mirror neurons and empathy to anticipation. Overt anticipatory behavior depends on just those hyperscalar properties of neuronal networks which are responsible for the evolution of *the mind* through self-observation. We explain how *Anticipative Capability* in the absence of self-observation is unlikely; that self-observation in the absence of scalar development is impossible; that emergence of scale corresponds to the emergence of a 'theory of self' in infants; and that the attainment of 'wisdom' in humans is associated with the development of cervical hyperscalarity. We conclude that *both* the historical development of the mind *and* its ongoing evolutionary nature can be best characterized by 'survival of the adequately anticipative'.

1 Introduction

In this paper we will adopt the radically simplifying supposition that it is possible to refer to the multitude of individual animals similar to ourselves as a *species*, a *group*, or a *society* to which we can attribute common mental states, points of view or thoughts. While this is not particularly controversial with

respect to the word *species*, it becomes far more so for the words *group* or *society*. We[1] are writing this paper from within a specific room, university department, society, country and continent, and at each level of this quasi-localization, as for other categorizations, it could be argued that generalization automatically leads to misunderstanding and falsehood. However, generalization is central to our daily lives and our survival, and its absence renders communication impossible. Its use and influence provides much of the motivation for the objective matter we will describe here, but it will also be more or less evident that it automatically permeates our subjective descriptions. Communication involves not only the transmitter, but also the receiver. Given a choice between saying nothing, thus avoiding imprecision, and commenting nevertheless under the supposition of vigilant appraisal, we select the latter.

A minor difficulty with the argument we wish to present is that it interweaves a multiplicity of very different strands, from theoretical biology to neuroscience and philosophy via crystallography and system theory… Our fervent hope is that the reader will not be inadvertently sidetracked into a comprehensional *cul-de-sac* by any lack of clarity of our description or by the direction we will take at logical bifurcations along the route.

1.1 Anticipation and Anticipative Capability

The capacity to predict the future is a major constituent of the mental platform from which humans observe and judge their surroundings, both as individuals and as groups. Predictably, with this statement, we are now unintentionally embroiled in prophesy and soothsaying! As usual, further characterization can resolve the problem we have created: some things can be predicted, some can not. When driving along a motorway, and observing that the carriageway in the opposite direction is completely blocked by an accident, it is reasonably easy to predict to some extent the short-term future of drivers traveling in the opposite direction on their own side of the motorway (assuming that there is no intervening motorway exit; that the time it takes them to reach the blockage is greater than the time it takes to clear it; et cetera; et cetera; et cetera; …). Any generalization that 'the future is unpredictable' has, itself, a contextual dependence: it just depends what is being talked about! Consequently, the capability to anticipate future[2] situations or events is not a purely technological 'set it and forget it'[3] ability, it is at the very least both a contributor *to* and consequence *of* the relationship with the environment which is usually referred to as *intelligence*. We will argue in this paper that

[1] … at which point, if not before, we, the authors, are of course immediately immersed in the problem itself !

[2] We beg the reader to excuse this tautology, which has been inserted in an attempt to reinforce clarity!

[3] A phrase lifted from the computer program 'Diskeeper' [3].

Anticipative Capability (*AC*) is central to the evolution which has led to our species' mental development, and that the evolutions of survivability, anticipation, consciousness, intelligence, wisdom, evolution *itself*, and indeed *the mind* are broadly equivalent.

1.2 Anticipation and life

A large part of the differentiation we attribute to distinct parts of our environment relates to the presence or absence of anticipation in an entity's behavior. Entities which exhibit anticipation *are* alive: the rest *are* not. Is it necessary to distinguish between '*are* alive' and '*seem* alive'? Well, it just depends – but is this important? In supposing that our prime aim is to survive, it does not make much difference. Killed by a tiger and killed by an automobile are much the same (assuming equality of our reactions to the two!). It is not ontology which is of concern, but where 'a difference makes a difference'[4] to the way things are viewed and to *state of mind*. If a computer passes the Turing Test [5], it may as well be treated as a person (or rather, in Turing's own construction, as a woman). Further consideration is, like philosophy, for people whose bellies are already full!

Within the localization of specific room, university department, society, country and continent we noted above, our bellies are usually suitably convex, and we find it necessary to disagree with the usual self-referential differentiation between 'what is alive and what is not'. In experience, 'the alive' do indeed exhibit anticipation, but is it really absent from everything else's 'behavior'? Billiard balls are clearly not alive according to the usual differentiation, and we would not expect to be attacked by them (always supposing that their 'normal' state is one of *rest* relative to ourselves, and not of *rest* relative to our galaxy…). But why do they bounce off each other? In doing so they maintain their *identity* – they never merge, and their constituents remain distinct. Are the grounds for this Newtonian reactivity categorically *different* from anticipation, or are they the exposition of a minimal yet relevant degree of regard for the future – of mental activity? 'Rubbish' we hear – yet is the position which generates this reaction a scientific one, or simply the outpouring of a concern to maintain our position in this world as 'something special'[5]?

1.3 The argument

In this paper we will begin by supposing that there is no categorical difference between 'what is alive' and 'what is not', and that the observable distinction we make between the two emerges from their contextually-created properties, and not from our own prejudices. In doing so, we will clearly be accepting

[4] Lifted from Gregory Bateson's [4] description of information.

[5] The authors nominally apologize for mistakenly using here the word 'something' in place of 'someone'.

that there *is* no fundamental difference, even though it may be useful for us to build the distinction into our psyches at a very low level.

Our task is now to account for the evolution of *Anticipative Capability (AC)*, as a route towards shedding light on *the mind's* emergence. We will first address more carefully our self-centered conceptual distortion of Nature, then follow this with a consideration of the traditional compartmentalization of evolution and question its key assumption of 'purely' random mutation. It is difficult from outside an organism to distinguish between *anticipation* and its *simulation* – as it is between *mind* and *mechanism* – but we can clearly establish the reality of *AC* in ourselves. We propose that *AC* has most likely evolved towards its high-level implementation through the intermediacy of its simulation, and that this is consistent with the paper's initial rejection of a categorical distinction between *alive* and *not alive*. We note the inevitability of *scale* in organisms, and place it in a universal ecosystemic context, which then leads us to draw conclusions about the nature of *intelligence*, *sapience* and *wisdom*, and their role in *the mind's* evolution. Anticipation requires extensive environmental and inter-organism information to operate effectively, and in this context we will address the importance of recently discovered *mirror neurons*.

Armed with these tools we will conclude by addressing both the central character of *AC* and its peripheral effects, and suggest that biological information-processing systems operate primarily through self-observation, supported by the partial isolation of their structural scales. We will indicate a manner in which awareness may evolve, and propose that this self-observation is the central characteristic of *the mind's* neural architecture. In all of this, *AC* raises its head as an evolutionary facilitator, not only in directing evolution, but in supporting the currency of awareness, consciousness and identity. As the paper's title suggests, we not only describe the evolution of *AC*, and its function as an evolutionary process, but will depict the *mind's* emergence as a consequence of *evolving Anticipative Capability* itself.

2 Anthropomorphism

The traditional *Homo-sapient* view of Nature maintains that objects and organisms are essentially different. The history of this viewpoint's development is very complex, and given the historical complexity it is often easier to accept that 'things are as they are usually described' rather than to poke around in descriptions which are part of the understanding of our own nature. This 'acceptant option', however, is inadmissible in our current exercise! The ancient Greek philosophers bore witness to attempts to de-mystify our surroundings, most obviously in Aristotle's [6] replacement of Plato's [7] deistic explanations by human experience and definition. By the nineteenth century, *man* – as a generic term including, of course, *woman* – had come to see *himself* as something of God's equal, in *his* newly found engineering capabilities and unfettered horizons, but *he* was not yet sufficiently self-confident to dethrone *his* deistic *sibling* and fully embrace atheism. Not so in the twentieth century, most particularly as a result of the abrupt rise in technology engendered by

the global conflicts between 1930 and 1945: science now began to take its 'rightful place' in the scheme of things, thus demonstrating that *man* was self-sufficient, no longer needing to rely on divine influence to control *his* fate.

One important aspect of *man's* rise to dominance – at least, as seen by *himself* – was the historical establishment of a view of Nature which presupposed a preliminary separation of 'distinguishable entities' into 'the living and the inanimate' (not, of course, forgetting the 'inplantate'[6]), followed by the interesting (!) construction of a hierarchy of 'the living' which included an abrupt differentiation between *man* at the summit (created, of course, in God's image) and 'the rest' (where *man* did not always include 'animals resembling the auto-referential *man*, but being of different colors', or even *woman* – explicitly included above). A natural consequence was the wholesale adoption of anthropomorphic descriptions of natural phenomena. Reasonably, human nature could take no other course: if *man's* sibling God had created *man* in his[7] own image, then it behooved upon *man* to explain Nature in *his* own image!

The progressive demise of a belief in God during the twentieth century created instability in the transfer of *man's* quasi-deistic character to his descriptions of Nature, and resulted in a radical rethinking of his place in the scheme of things.

2.1 ... or anti-anthropomorphism

Given the apparently general applicability of system theory, was *man's* reliance on an overarching anthropomorphic position sustainable? Clearly not. The end of the twentieth century witnessed its violent rejection, and any connection between humans' and Nature's characteristics was expunged. The result of this *anti-anthropomorphism*, however, was arguably even worse, as it completely decoupled *man* from his environment! One consequence was the associated automatic acceptance that evolution was uniquely 'directed' by random processes: that is to say, it was not directed *at all*!

2.2 Dethroning anthropomorphism

The common aspect of *anthropomorphism* and *anti-anthropomorphism* which concerns us here is their common categorical separation of 'distinguishable entities' into 'what is alive' and 'what is not'. We see no reason why this categorical separation should be maintained. Our own view is that 'life' is a label which humans stick on some entities and not on others, without looking any further into whether it is a defining characteristic or not. We have published

[6] … and not, of course, forgetting the other three 'kingdoms' of life – the prokaryotic *monera*, and the eukaryotic *prostita* and *fungi* – and not, of course, forgetting *mimivirus et al.*

[7] … or, in *her* own image… or maybe even better, in *its* own image.

elsewhere [8] arguments in favor of non-self-referential 'definitions' of life (for example, in terms of joint digital-analog coding, as proposed by Hoffmeyer and Emmeche [9]), but the primary aim here is to understand the 'whys and wherefores' of life, and not to be able to state that 'this entity is alive, and that one is not'. Ego is a powerful mover: *Homo sapiens* will have to dethrone itself from its presumed position at the pinnacle of evolution to address these questions. Over the past 50 years, evidence from anthropological studies has been growing that the 'exclusively human' characteristics which are held most dear can also be found in other species. The long-held supposition that 'lower animals feel no pain' is now 'on its last legs'; Thompson and Ogden [10] have impressively demonstrated that while macaque monkeys and pigeons cannot use analogy as a tool, chimpanzees can and do on a regular basis [11]; the widely published video of *Betty the crow* manufacturing a hook from a piece of wire to get hold of food [12] has been a shattering revelation!

The dethronement of *both* anthropomorphism *and* anti-anthropomorphism leaves a descriptive vacuum to be filled. Fortunately, within the same time period that anti-anthropomorphism took centre stage, the environmental movement and 'the ecosystem' were 'born', arguably dating from the publication in 1962 of Rachel Carson's book 'Silent Spring' [13], and the stage was set for a yet another revolution related to anthropomorphism – that of *man* as *a part of* Nature! [8]

3 'Roll back driver'…

A common problem encountered while updating computer software is that for some reason a newly installed version of previous code causes problems. While 'traditional' computer systems offer the possibility of 'uninstalling' the new software, this is rarely carried out cleanly, and various bits and pieces of code are left lying around to trip up the user or crash the system. A similar situation can be found within the conceptual framework from which we, as humans, view our environment. Ideally, it should be possible to 'uninstall' fallacious segments of belief and start again from before their manifestation, but unfortunately the human 'uninstall' process is no more reliable than its software analogue, and it is difficult to be confident of success. In the computer domain, the lack of confidence in un-installation procedures has been addressed by the use of a new term – that of 'rolling back' an installation. Even if the uninstall procedure remains exactly and erroneously the same, this

[8] The reader could reasonably argue that the stance we have adopted here does nothing beyond replacing 'reasoning derived from a presumed deity' 'by reasoning derived from a universality of Nature'. While this is indeed the exchange we have made, it results in quite a different view of our environment, which rather than being based on 'transferred deism', thus engendering human ego, is grounded in a humility of similarity, eliminating the 'egotistical requirement' for humans to be 'above' the rest of nature.

at least *sounds* as if an un-installation is the exact reversal of the preceding installation.

In the context of this paper we wish to perform an analogous operation. We do not wish to *recombine* the fallaciously-differentiated categories of 'alive' and 'not alive': we wish to *roll back* our conceptualization and remove from our discussion any predetermining sense of their validity. The Microsoft Windows Hardware Device Manager offers the recovery option of 'Roll Back Driver' [14]. A *driver* couples a device to its host by appropriately organizing information for it to operate effectively. We believe that the usual ungrounded categorical differentiation between 'alive' and 'not alive' incorrectly organizes the information which is needed to understand life and evolution.

Our ostensible purpose here, therefore, is to 'roll back' that differentiation and install a new 'driver' which can make sense of the apparent conflict between natural ecosystemic unification and its conventional categorical distinctions. *Homo sapiens* is a part of Nature. Quantum mechanics has indicated that observation without influence is impossible. It is natural to expect that a useful *description* of the environment will mirror the characteristics it exhibits, much as Robert Rosen [15] developed a validating coordination between formal models and real systems. The *abstract* driver which is needed to understand evolution will correspond to the *practical* driver which evolution itself requires. The prime candidate for this dual role is *Anticipative Capability* – the ability to 'see' where descriptions are leading; the capacity to 'direct' evolutionary change – without which both understanding and evolution would drown in a combinatorial explosion of possibilities and likelihoods. However, this then leaves no option other than to address the question: 'where does *AC* reside?' But is that a sensible question? *Does* it 'reside'? *Can* it be localized? Does it only manifest itself in the convoluted cortices of the 'higher' species? Or is it a property of *every* differentiated entity? Or is it 'everywhere'… a ubiquitous sea in which both 'alive' and 'not alive' drift, survive and evolve?

We suggest that it is all of these. But not equally, everywhere. High-level implementation of *AC* necessarily implies extensive information-processing, but it is not simply the *quantity* of processing which must be maximized, it is the spatial information-processing *density* [16]. So, even if *AC* is a universal property, its evolution should be most evident in dense information-processing networks. The central hypothesis of this paper is that highly *evolving Anticipative Capability* and conceptual *mind* are indistinguishable.

3.1 New driver not required…

The first 'roll-back' to be performed is that of the attractive analogue we introduced above. It is *not* necessary to install a 'new driver' to make sense of Nature – a suitable one is already 'provided' in the kernel of the 'Operating

System'. Evolution does not require *AC* to be 'pre-fetched'[9]. Evolution and *Anticipative Capability* are both[10] mutually catalytic and dependent. *AC* is concurrently a precursor of evolution and one of its outcomes. The development of a primitive eye enabled past organisms to survive while more complex vision evolved. Similarly, the evolution of evolution corresponds to a refinement of mutational or developmental directivity, supported by a degree of capability to visualize the consequences of change, and resulting in enhanced visualization.

4 Evolution or evolution?

Woe betide he (or equally, of course, she) who suggests that there is anything other than a linguistic relationship between evolution, as the temporal advancement of a chemical reaction or a planetary system, for example, and Evolution, as the sacrosanct practice of Life. Darwin [17] proposed that the Evolutionary process which led to species differentiation depends on natural selection, itself supported by variation and reproduction. But did Evolution spring into being, just as Darwin [17] later described it, at precisely the moment two carbon chains first encountered each other in a primeval soup? Surely not[11]. We contend that Evolution is a product of evolution, and that its categorical process-compartmentalization into *mutation*, *reproduction* and *selection* is an evolutionary artifact similar to the condensation of matter into spatially-compartmentalized forms such as protons, neutrons and electrons.

Mendel's [19] experiments indicated that Evolutionary inheritance is digitally controlled via genetically transmitted material. Genomic mutation is traditionally supposed to be entirely random, even if randomness in the other contributions to Evolution is more difficult to quantify. We postulate that in general the random aspects of evolution have been progressively modified during its evolution towards more directed forms, and that it is in this way that evolution itself has evolved into the apparently compartmentalized set of Evolutionary operators described by Darwin. The first issue here is not whether Darwin's ideas make sense or not, it is whether the conventionally categorical compartmentalization of the Evolutionary 'process' into *mutation*, *reproduction* and *selection* is sufficiently valid. The second issue is whether Evolutionary mutation is truly random – undirected – or whether some sense of internal or environmental consequence can control mutation in a … - $genotype_{(x-1)}$ - *geno-*

[9] 'Pre-fetching' is a computer technique which speeds up applications and services by anticipating their procedural requirements and loading them into memory in advance.

[10] … where the textual ambiguity is appropriate.

[11] It is worth noting that recent research places the origin of self-reproducing entities at the level of RNA, where two enzymes have been demonstrated to perpetually 'cross-replicate' [18].

type$_{(x)}$ - *genotype*$_{(x+1)}$ - … Evolutionary sequence. The third is whether any presumed Evolutionary directivity can be the author of its own enhancement.

4.1 Evolutionary[12] anti-compartmentalization

The categorical compartmentalization of natural processes is a convenience which is anticipated[13] to aid in understanding, although, as we hinted in the Introduction to this paper, generalization often results in *mis*understanding. Categorization is *always* approximate, but *successful* categorization is only minimally so. Unfortunately, our overridingly blind belief in the *linearity* of cause-effect relationships leads us to suppose that a 'good approximation' is always sufficient – that '99% will do'. While this may be so at the heart of 'equilibrium physics' it is wildly *in*sufficient in the 'far-from-equilibrium' world of life and evolution, where microscopic differences may bring about catastrophe[14]. The basic dichotomy of Nature is that its 'invention' of survival-assisting simplicity has concurrently generated survival-endangering loop-holes. The evolutionary adaptability of life is Nature's 'war horse' in its balancing act between these extremes, and process-compartmentalization aids system stability by 'capping' deleterious sub-systemic excursions.

We challenge the central compartmentalist dogma of a conventionally held view of evolution that the genome is isolated from environmental influence. The traditional interpretation of Darwin's [17] proposals is that there is *no* Lamarckian [20] 'injection' of environmental information into a genome, which would suggest that evolution cannot *itself* evolve. The translation of this interpretation into a contemporary view of genetics suggests that modifications to the structure of DNA *always* correspond to the enactment of *entirely random* molecular mutations, whose structural survival over generations follows Spencer's [21] principle of 'survival of the fittest'. However, an important realignment of this interpretation maintains that a more apposite characterization would be 'survival of the barely adequate'[15]. Although Spencer's dictum may well have suitably characterized survival within early ecosystemic populations which only exhibited a small number of genes, every individual gene does not critically influence organism survival. An organism with a single genetically-defined characteristic would be far more susceptible to environ-

[12] For the remainder of this paper we will dispense with any formal distinction between use of the words Evolution and evolution.

[13] … which of course requires the application of AC!

[14] In many cases, for example, the mutation of a single DNA base-pair can result in an organism's death.

[15] There are many and varied modified forms of Spencer's [21] 'evolutionary principle', from Pietikainen's [22] "*Potential non-survival of inherited features not aligned with potential survival of the inherited features that provide better survival in the current environment*" to Norman's [23] "*Whoever dies with the most toys wins*"!

mental extinction than one where a vast array of characteristics is defined by a complex multiply-connected genetic network. We suggest that a progressive broadening of the genome 'pool' from 'survival of the fittest' to 'survival of the merely adequate' has paralleled the evolution of more and more complex organisms, as illustrated in Figure 1. For a complex organism this has permitted multi-stage multi-generational genetic evolution to take place without intermediate genetic products being forced to provide instantaneous competitive advantage.

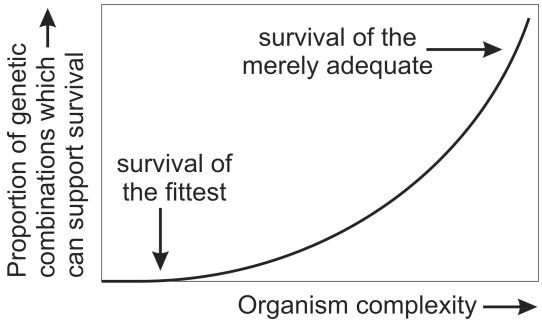

Figure 1: Hypothetical spreading of the genome pool with organism complexity, from 'survival of the fittest' to 'survival of the merely adequate'.

The last century's interpretation of evolution[16] presupposed that genotypic DNA contains complete instructions for the creation of a representative phenotype, that sexual reproduction implements the results of random mutations between the DNA of two phenotypes, and that the selection of surviving genotypes is enacted by the natural environment. Central genetic dogma maintained that expression of the minute proportion of DNA which codes for the production of proteins is sufficient for embryogenesis. Long before the identification of DNA, however, argument raged over the relative importances of 'Nature' or 'nurture' in phenotypical development, and the observation of radical differences in morphology, character and disease sensitivity between identical twins has raised serious questions as to the central genetic dogma's sufficiency.

4.2 Genetics versus gene-protein mapping

Recent research has begun to shake the foundations of contemporary belief in the unique importance of 'one gene, one protein' to embryogenesis[17], indicating that even *complete* knowledge of the human genome would be insufficient to determine human fabrication. The Human Genome Project [25] began with the presumption that complete knowledge of the protein-coding genes in

[16] More precisely, this statement should only refer to an extended middle-to-late period of the 20th century. The interpretation it refers to is even now difficult to 'roll back'.

[17] … leaving aside here the added complications of 'one gene, more than one protein' which are associated with alternative splicing (see, for example, [24]).

DNA would reveal the entire 'human blueprint', but it now appears that vital parts of this 'blueprint' are located *outside* those genes. Protein-coding genes only account for some 2% of human DNA[18], and the remaining inter- and intra-genetic sequences have long been dismissed as irrelevant evolutionary artifacts, or 'junk DNA'. But is the 'junk' DNA, *really* junk? While there does not appear to be clear correspondence between a species' complexity and the number of its protein-coding genes

"… the amount of noncoding DNA… does seem to scale with complexity" *[26].*

While large parts of DNA may not code for proteins, they do produce active RNA, which can *directly* influence cell behavior [27]. There is now an extensive known family of RNA variants whose revealed functions range from environmental sensing to gene suppression: pseudogenetic-RNA (e.g. [28]); antisense-RNA (e.g. [29]); double-stranded-RNA (e.g. [30]); riboswitch-RNA (e.g. [31]); more than 150 different micro-RNAs (e.g. [32]); … Research into the family of protein-coding genes has always been facilitated by their easily recognized standard 'start' and 'stop' codes. RNA-only genes, however, do not exhibit such general characteristics, making their structures and functions far more difficult to determine. The current state of discovery, however, does indicate that RNA-only genes constitute a hitherto unsuspected layer of information and control in the genome.

The comparison of orthogolous base-pair sequences between different species indicates that many long-range non-protein-coding genetic regulators have been conserved throughout long periods of evolution. Surprisingly, a high proportion of these are found in the ~98% of DNA which is conventionally assumed to be irrelevant – in the 'junk' DNA [33, 34]. Experimental results indicate that widespread regulatory changes may have contributed to uniquely human features of brain development and function, again weakening the primacy of 'one gene, one protein'. In addition, Segal *et al.* [35] have reported the discovery of a pattern embedded in the organization of the ~30 million protein-spool nucleosomes around which a DNA chain is wrapped. Segal [36] has suggested that transcription factors may only recognize sequences which lie *between* nucleosomes, and that those which occur in parts of the DNA which is wrapped *around* the nucleosomes may be inaccessible. If this is correct, it would provide

"*… a 'real quantitive handle' on exploring how the nucleosomes and other proteins interact to control the DNA*" [36].

[18] There is no generally agreed value for this percentage, which is variously quoted with values between 1% and 5%. In any case, protein-coding sequences only make up a minute fraction of the some 3 billion base-pairs of human DNA.

More significantly in our current context, the nucleosomes provide yet another mechanism by which 'external' control is exercised on the conventionally presupposed uniquely 'internal' workings of DNA.

4.3 Randomness, Baldwin, epigenetics, individuals and societies

In the light of the examples given in Section 4.2, and of other recent results, we feel justified in questioning the attribution of perfect environmentally-isolated randomness to genetic mutation. Even if we were to maintain such a belief, it would necessarily only apply to the meaninglessly unreal *abstraction* of DNA from its local environment. It now seems likely that a major part of reproductive and embryogenetic control is *not* exercised by the small percentage of DNA which *seemingly directly* and digitally controls protein synthesis. As Mattick has suggested:

"*… what was damned as junk because it was not understood may, in fact, turn out to be the very basis of human complexity*" [26].

Many of the other DNA-related processes now coming to light depend on a multiplicity of more-or-less locally-environmental *analogue* effects. This 'insertion' of *analogue* influences into the previously supposedly *digital* transcription-synthesis route between gene and protein has *enormous* consequences. A prime attribute of digital or quantized interactions is their insensitivity to 'noise' and to small-scale locally-environmental influences. This simplifying isolation breaks down when analogue influences come into play, leaving the door wide open to environmental pressures on genetic mutation.

There are yet other low-level mechanisms which impact on the simplistic random view of genetic mutation, and this even at the *digital* level of the genes themselves. For example, the randomness of mutation which leads to an 'independent assortment' of genes only really applies to genes located on different chromosomes. If genes are close to each other on the *same* chromosome they have an increased chance of being inherited jointly – a phenomenon referred to as *gene linkage* [37].

While Darwin [17] espoused random variation, Lamarck [20] defended environmental causality: the Baldwin effect [38, 39, 40], however, provides a 'masking' mechanism somewhere *between* random variation and the facilitation of anticipation. Terrence Deacon [41, 42] has proposed that complexes of genes can be integrated into functional groups when environmental changes mask and unmask selection pressures. Many animals synthesize vitamin C, but in anthropoid primates the crucial gene for this endogenous synthesis is nonfunctional. Deacon [41] has suggested that the loss of functionality was linked to the evolution of color vision, which promoted the adoption of a diet of fruit rich in vitamin C, and that this masked the effect of the gene.

If all this were not enough, the last decade has witnessed the 'emergence' of a far greater influence on the pathway from genome to phenotype – that of epi-

genetics[19]. Epigenetic codes are far more susceptible to environmental influence than are genetic ones. Epigenetic control is exercised either by changes in the proteins (histones) that package DNA into chromatin, or through modification of the DNA itself (methylation). Epigenetic influences can result in phenotypical characteristics which have traditionally been considered to be purely genetically determined. Waterland and Jirtle [44] have demonstrated, for example, that change in the diet of a pregnant mouse can completely change the color of her young.

But do these various effects which permit environmental DNA modification constitute or provide evidence of anticipation? Well, not necessarily on their own, but they do provide mechanisms contrary to traditional genetics through which anticipation may be exercised. Even if the mutations associated with reproduction are randomly defined, the selection of a mate for reproduction, and therefore the selection of which genome takes part in mutation, is dependent on individual choice. But individual choice is always modified and sometimes strictly controlled by social pressures [45] or enforcement [46]. While Darwin [47] proposed that the frequency of genetically acquired individual traits depends on the sexual attractiveness of their possessors, the abstractness of many higher-level social mores and attitudes provides a pathway to genome manipulation through intention and fashion. In societies' most extreme rejection of randomness, the inconvenience of human evolutionary mutation is currently under attack from medicine and genetic engineering. The ubiquitous anticipation of death has provoked extensive research into its causes, and a major target of scientific endeavor is the elimination of cancer.

The particulate objects of Newtonian physics are restricted to 'externally' defined interactions in their temporal evolution. The particle-like entities of Quantum Mechanics are presumed to follow 'externally' imposed rules which depend on the randomness of probability. Primitive low-complexity organisms appear to follow a rule-based evolution which depends on random mutation, while high-complexity mammals develop complex social structures which modify the randomness of their evolution. Human societies are currently developing through genetic engineering the means of *excising* randomness from their evolution though anticipation. Are these apparently different cases just that – different – or are they different stages in the continuous development of *Anticipative Capability* through evolution, from the simply inanimate to the complex animate? And is this apparent progression towards *AC* externally directed, or is it the recursive enhancement of *AC* itself: is *Anticipative Capability* auto-catalytic?

[19] Conrad Waddington's [43] term for "*the interactions of genes with their environment, which bring the phenotype into being*."

4.4 The auto-enhancement of *Anticipative Capability*

Directed action presupposes some degree of anticipation in its planning, even if this is only 'that there will be a next achievable state', and the success of directed action depends on how well all possible intermediate eventualities will have been anticipated (or it depends, of course, on a lucky guess!). Logical anticipation relies on pre-established internal models, both of possible eventualities and of their dependence on the balance between predictive vagueness and predicted causality, but its mainstay is a belief in itself, based on past success. As an auto-catalytic learning process, anticipative success builds confidence in *AC*, and supports the anticipative exploration of the environment's 'future phase space'. But this is *only* probable if its executor is *aware* of both immediate and past successes.

Anticipation appears in at least two different guises in our account. One is the real-time anticipation of an entity's movements or intentions, for example the path a falling rock will take, or whether a hunter's prey will jump to the left or to the right; the other is the integration in an organism's DNA of genetic modifications which appear likely to enhance survival. The first of these two is comparatively simple, in that it depends only (!) on the generation of sufficiently general internal models and their intelligent application[20]. The second appears to be extremely complex, as it raises the question of correspondence between the phenotypical modification of its genotype's survival and the genotypic modification of a phenotype's survival.

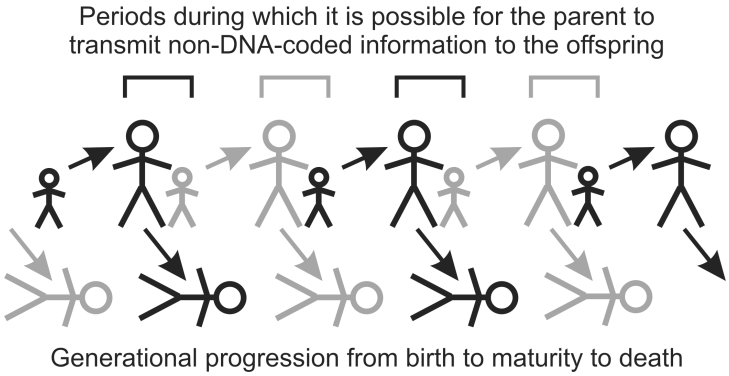

Figure 2: The transmission of non-DNA-coded information from parent to offspring during the period between offspring birth and parental death.

Let us take as an example the usual instantiation of Lamarckian [20] evolution: the giraffe's neck. In an environment where a multitude of smaller animals consume leaves which are low down on the trees, a long neck promotes survival by permitting access to leaves which would otherwise be out of reach. Darwin [47] proposed that animals and birds choose their mates on the basis

[20] We address the association between anticipation and intelligence in Section 6 of this paper.

of a non-random judgment of fitness, and numerous more recent studies [48] have confirmed his hypothesis. While our giraffe is constrained to mate with another one that concurrently *survives* – which only implies that it has sufficient *randomly*-generated 'fitness' – it must still *choose* which one to mate with: genome inheritance depends on *both* randomness *and* directivity. Consequently, although our giraffe only addresses its own current needs, its actions have future repercussions: the phenotype inadvertently modifies the heritable genotype, and in doing so unconsciously enhances the chances of survival of future phenotypes. This appears to be a process which is independent of anticipation, but an intermediate part of the story is missing. Our giraffe has offspring, but these are not just 'left to survive on their own': they pass an extended time with their parents, being *intentionally* taught, and learning techniques which enable them to survive and later to breed. Although their 'fitness' initially depends on random mutation and inbuilt instincts, it also depends on our giraffe's transfer of information to its offspring and the resulting replacement of short-term instincts by long-term goals [49]. Figure 2 illustrates the evolved progression of generational interactions through which the 'hard-wired' DNA pool of offspring's instinctive capabilities can be modified or broadened by the transfer of more abstract *non-DNA-coded* capabilities during the period within which parents and their descendents co-exist[21]. In human families, parents educate their children through *anticipation* of their future needs[22]: is this reasonably *entirely* absent from our giraffe's behavior?

The first of anticipation's two apparently different guises in our account – the real-time anticipation of events or actions – is none other than the initial stage of our giraffe's history. Both guises are part of evolution: real-time anticipation leads directly or indirectly to long-term change. The 'memory' which supports transfer of the implications of current action to future survival is *the environment itself*. Evolution consists of two coupled systems: a coded carrier – the genotypic DNA – and its logical ecosystem – the phenotypic environment: evolution is *birational*[23].

We propose that the historical evolution of evolution has constituted a progressive change from the almost pure randomness of Quantum Mechanical

[21] The authors leave to the reader any contemplations of the relevance of *memetics* to the inter-generational transfer of *non-DNA-coded* capabilities.

[22] Given extended lifetimes, it is not only *parent* and *child* that can co-exist and interact, but also *grandparent*, parent and child.

[23] More carefully, we would here suggest that evolution consists of *at least* two coupled systems – but that would overly complicate our description in this context. This would clearly then make Nature *more than* birational. For an extended description of natural birational systems, the reader is referred to Section 5.3 of this paper and references [50] and [51].

'social' interactions[24] towards the incipient dominance of directivity. While evolution arguably proceeds in leaps and bounds rather than in a smooth progression, we maintain that it is nevertheless a continuous process, but one which exhibits varyingly-rapid recursion [see 53 for a beautiful illustration of this effect from evolutionary computation].

A number of researchers are now convinced that anticipation is a pre-biological attribute of Nature, for example as a fundamental property of electromagnetic systems [54]. While this automatically justifies the attribution of low-level anticipatory activity to primitive organisms, it does not explain the emergence of the high-level *AC* we ourselves experience. Natural hierarchy theory indicates that *all* scalar emergences include a 'pre-planning' stage [50, 51][25], and we believe that during evolution Nature has first *simulated* high-level anticipation, and then later implemented it. Although anticipation is habitually associated with conscious cognitive processing, *single-celled* amoebas are observed to both orient themselves with infrared light [55] and hunt for prey. Their capacity to direct their actions without neurons implies some degree of sub-cellular cognition, and implies that anticipation predates neural networking. All organisms exhibit some kind of apparent anticipation, although it is often difficult to distinguish between the evolutionarily-early *simulation* of high-level anticipation and its later *implementation* as an *aware* strategy. Trees lose their leaves 'in anticipation' of winter. Some bacteria, for example *Vibrio fischeri*, only emit light when sufficient of their own number are present: 'anticipative' *quorum sensing* [56] depends on the presence of a threshold concentration of bacterial-emitted signaling molecules. The carnivorous Venus flytrap plant *Dionaea muscipula* 'anticipates' the sustenance it will gain by closing its trap-like leaf on a fly. Moths avoid flying into dark shadows, whose presence they apparently associate with predation – they survive by 'anticipating' attack. Mammals appear to be *aware* of their use of anticipation. As Dubois points out:

"*A cat jumping to a table is also a good example of an anticipatory system: the cat looking at the table builds in its brain a model of the situation and is concentrated on the final state, not the initial state where it is before jumping. So the cat has a model of itself and its environment and compute(s) its current state (the velocity and direction) in function of the anticipation of its final state*" [54].

Is *awareness* of anticipation necessary for its use? Only if the absence of awareness would negate its effectiveness: the *sufficiency* criterion for organisms is

[24] While individual Quantum Mechanical interactions are mathematically deterministic, their implications in large systems are probabilistic: Antoniou [52] has demonstrated that when classical Quantum Mechanics is extended to large systems the logical completeness breaks down.

[25] Even the implementation of *anticipation* requires *pre-planning anticipation* – it relies on the infinite recursive temporal chronicle of anticipation of anticipation of anticipation of ...

survival and not awareness. The effective *simulation* of anticipation is a computationally economical solution for less complex organisms, and its reduction in humans to an automated reaction in intensively practiced scenarios [49] provides similar advantages. Awareness of the use of anticipation in *directed* action, however, *is* necessary for its auto-enhancement. Even then, the simulation or automation of anticipation still permits 'instinctive' environmental adaptation, for example in the way that a good golfer will anticipatively adapt his swing to the state of the wind. But 'instinctive' action also often exhibits demonstrably inferior characteristics which derive from previously learned 'related' actions or accidental adoptions, and whose elimination demands *awareness*[26].

4.5 The evolutionary emergence of the mind

The thesis of this paper is that a developmental continuity exists between primitive low-level anticipation and its high-level *aware* exposition. Consequently, if we accept that *Anticipative Capability* has auto-catalytically evolved, and that awareness has been necessary for its evolution, then we must also accept that *low-level awareness* is *already* a 'property' of Nature's most primitive organization [57], and that it is *not* a uniquely high-level cognitive emergence. The picture we present is of an evolving natural association between *Anticipative Capability* and awareness – of *aware capability* – in short, of *the mind*. However, while the auto-enhancement of *AC* requires awareness of anticipatory success, it does not presume awareness of its own enhancement – it does not yet presuppose *self*-awareness[27].

The continuous evolutionary scenario we envisage is between the identity-conserving Newtonian anticipation[28] of minimally-aware primitive environ-

[26] A young boy in an isolated Scottish island in the 1970s learned to play an accordion upside down, which makes particular musical sequences extremely difficult to play. His performance was surprisingly good, however, but it could not improve beyond a certain degree until he was made aware of his initial error and corrected it. Small children often look at picture books upside down, until they become aware of which way up they should be – either from received information, or by noticing a correlation between pictures and the entities they represent.

[27] The following exchange appears in a published conversation [58] between David Bohm and Rene Weber. Bohm: "*I would say that the degree of consciousness of the atomic world is very low, at least of self-consciousness*"; Weber: "*But it's not dead or inert. That is what you are saying*"; Bohm: "*It has some degree of consciousness in that it responds in some way, but it has almost no self-consciousness*". It is important to note that Bohm uses here the word consciousness and not awareness. These two expressions are often freely interchanged, but there are clearly distinguishable scale-dependent phenomena which we will differentiate in this paper by the terms awareness and consciousness.

[28] We refer the reader back to Section 1.2 for reference to the identity-conservation of Newtonian entities.

mental differentiations and the high-level awareness of anticipation, of *AC and* of its auto-enhancement – of *self*-awareness – which are all attributable to at least *Homo sapiens*. Although 'continuous', this scenario entails the evolution of higher recognizably different *levels* of awareness, and different *kinds* of awareness, as evolution progresses – the *emergence* of meta-scales of awareness and self-awareness in association with an organism's anticipative information-processing, and the manifestation of *hyperscale* in a birational ecosystemic setting. The low levels of awareness we associate with primitive entities and organisms are *passive* in character – witness the success of describing Newtonian interactions as exhibitions of purely rule-based reactivity. In this paper we will describe as 'consciousness' the higher emerged levels or kinds of 'awareness' from which actions may be *directed*, and within which a degree of 'free will' may be experienced and apparently exercised.

5 The Inevitability of Scale

Scale plays a decisive role in anticipation, as does anticipation in scale. The prevailing conventional image of our universe is one of a countless myriad of differentiated entities – whether of quarks, of electrons *et al*. of molecules, of bacteria, of animals, of humans… Beyond this simplistic picture, we imagine that some kind of cohesion binds numbers of entities together into 'systems' – often small numbers (e.g. bi-atomic $Na^{+}Cl^{-}$); sometimes very large ones (e.g. ordered human DNA, which contains some 3 billion base pairs, each of 30 or so atoms, or the presumably disordered sun, with some 2^{190} components! [59]).

It would, however, be unrealistic to suppose that in all cases systemic cohesive forces are completely satisfied by their current numbers of sub-systemic elements. Ionic bonding – such as that between Na and Cl in $Na^{+}Cl^{-}$ – produces very small systems, evidencing rapid and complete cohesive satisfaction, but gravitational cohesion is responsible for *massive* conglomerations, which *always* accumulate more material if it is available. As usual, the most interesting regime in a polarized scheme is somewhere between the extremes, where countering forces are balanced and small influences can result in large effects[29]: carbon-based covalent-bonded organisms provide an excellent example of a varyingly cohesive system.

[29] Electronic circuits, for example, operate between the polar extremes of oxide insulation and metal conductivity, and the devices which exercise control are made of semiconducting C, Ge, Si from the middle of the periodic table, or their mid-range compounds GaAs, InSb …

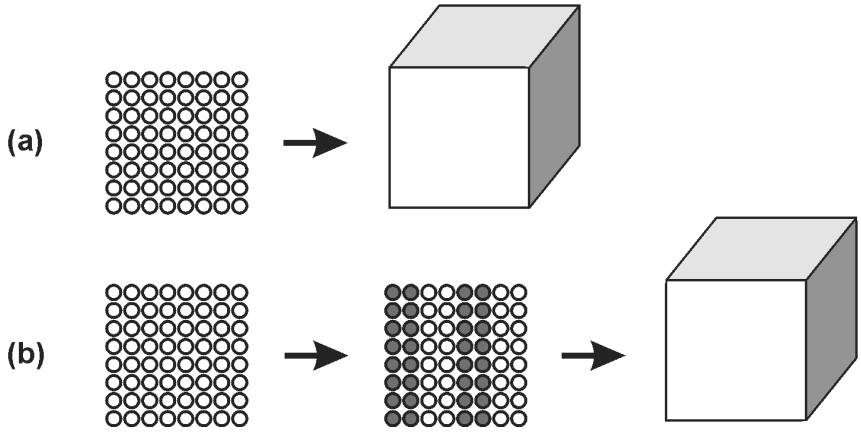

Figure 3: (a) The two scales of a simple cubic crystal, in which the microscopic geometrical arrangement of atoms is mirrored in the macroscopic shape of the crystal itself, and **(b)** The three scales of a cubic super-lattice, where the microscopic arrangement of atoms is first 'mirrored' in a mesoscopic alternation of atomic layers and then in the macroscopic crystal shape. The reader should note that the internal atomic arrangement of each multi-atomic-species quasi-molecular layer in a super-lattice is more complicated than this simplified illustration suggests.

When multi-component or multi-agent systems expand they progressively lose cohesion when this is compared to competing environmental forces, and they become unstable. They then have two 'options': to fragment; or to restructure themselves to remain unified. If, as we have suggested in Section 1.2, Newtonian particles act anticipatively to maintain their identities, then it is reasonable to propose that multi-component systems are driven towards restructuration rather than fragmentation in a similar anticipative manner. Restructuration creates a new kind of organization, where energy is conserved by replacing a proportion of low-level short-range cohesive communication by more efficient higher-level long-range communication, and a new system scale is born[30]. An excellent example of this effect is provided by naturally stable crystals, where the observable physical shape at the macroscopic scale is coupled to and defined by the regular arrangement of atoms at the microscopic one (see Figure 3(a)).

5.1 Scalar fragmentation

Ultimately, restructuring a system's cohesion leaves the way open for new expansion, and instability soon raises its head once more, but now it usually appears at the newly emerged scale. It may be that the 'option' of restructuring presents itself once more, and that stability can be re-established through the creation of yet another new, and again higher scale. An example of the

[30] … or to use a more common expression, 'a new scale *emerges*'. For simplicity we have made no reference here to boundary conditions or external interactions: we have described the instantiation of scale *as if* it were a result of a hypothetical 'self-organization' rather than the consequence of ecosystemic influences, without which there would be neither tendency to cohere nor to fragment.

resulting appearance is provided in the crystallography domain by artificially-created super-lattices – for example of AlGaAs and GaAs – which exhibit three different scales of organization rather than the two of simpler crystals (see Figure 3(b)), although here the developmental sequence depends on technological artifice rather than on the resolution of internal instability. The *natural* development of multiple scales is commonly restricted to carbon-based covalent-bonded systems – to living organisms.

A first systemic restructuration takes place *from* a population of coexisting differentiated entities in the scientifically familiar *spatial* dimension[31] *into* a differentiated entity in the less familiar 'dimension'[32] of *scale*. Further yet-higher scales *also* emerge into the same scalar 'dimension', where a population of differentiated scales may coexist. But in a naturally developing system, these scales are not independent: their multiplication corresponds to the emergence of a differentiated entity in the yet higher dimension[33] of *hyperscale*.

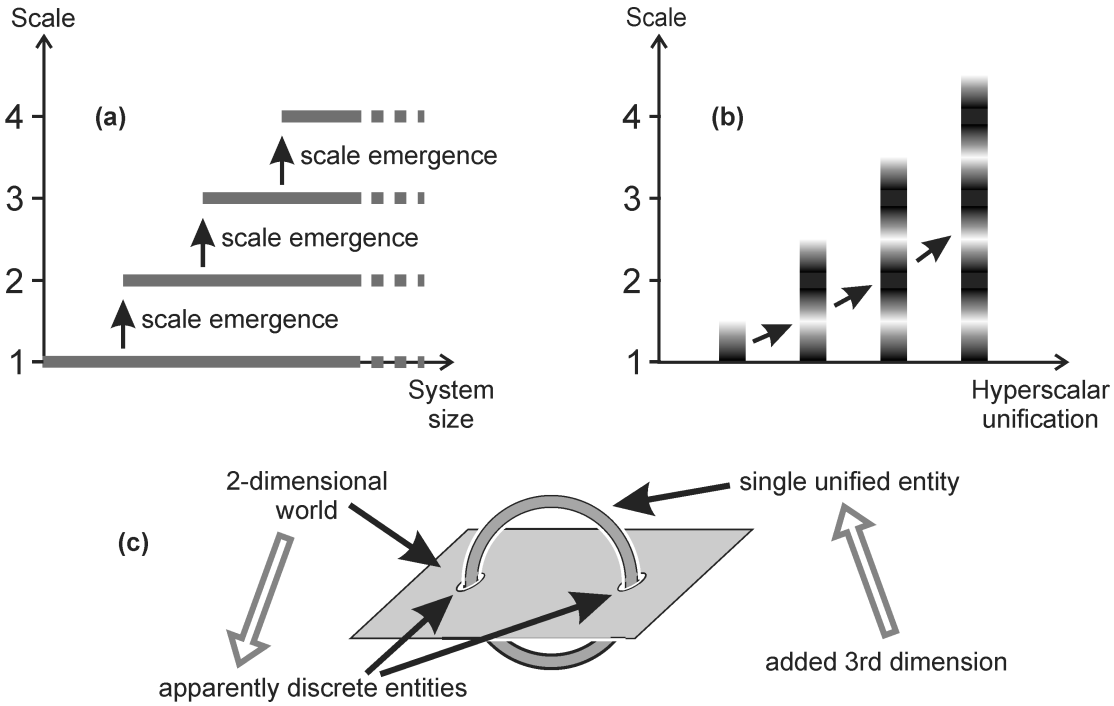

Figure 4: (a) Fragmentation of a natural system in the scalar dimension corresponds to **(b)** Unification and the maintenance of identity in the hyperscalar dimension, and **(c)** Two discrete entities in a two-dimensional 'world' may correspond to a single entity when viewed in, or from three dimensions.

[31] … or dimensions.

[32] … or 'dimensions'. It is notable that although this sentence is written from an external observer's point of view it applies equally well to the system itself, which would probably also find the spatial dimension more familiar than the scalar one!

[33] … or 'dimensions'.

The 'choice' made by an expanding system to retain its identity by *not* fragmenting in the *spatial* dimension consequently apparently leads to fragmentation in the *scalar* dimension through the emergence of multiple scales (Figure 4(a)). Its identity, however, is concurrently maintained in the *hyperscalar* 'dimension'[34] (Figure 4(b)), in a manner analogous to the way in which multiple apparently discrete entities viewed in *two* spatial dimensions may correspond to a single *three* dimensional entity, as illustrated in Figure 4(c). It should be noted that Figure 4 is presented here merely as an aid to visualization: it illustrates the barely realistic simplest form in which the complex relationships between spacial extension, scale and hyperscale of a quadri-scalar unified entity can be portrayed.

The individual scales of a natural system are very different from those which may be imposed on an artificially established information-processing assembly. A digital computer is often described in a hierarchical form, which corresponds to its simplification in a bid to promote understanding, but there are no real scales present. If we modify the descriptive boundary of one 'enclosure' – for example 'the processor' – to include another – e.g. 'the processor *plus* memory' – it makes no difference at all to the computer's operation: the apparent scales are only descriptive, for the sake of convenience. The scales of a *natural* system *cannot* be subjected to the same kind of 'mix and match' operation. Any one scale's internal character depends on all the others through a combination of emergence and slaving, and this partially isolates it from them.

Figure 4(b) implies somewhat simplistically that system unification, if not cohesion, increases with the number of extant scales. This is partially a result of inter-scalar slaving, but it is also a consequence of the inter-scalar transmission of order [50] illustrated in Figure 3. Here again, crystal structures provide a useful, if more complex comparison. Single (monoscalar) cubic crystals of carbon are far more cohesive, or stronger in general terms, than are both their hexagonal mono-crystalline and polycrystalline counterparts, but polycrystalline metals can be *either* weaker *or* stronger than their monocrystalline siblings. If a polycrystalline metal's dislocation-accumulating crystallite boundaries are more resistant to deformation than the crystallites themselves, then the metal will be weak and ductile; but if the crystallites are more resistant than their boundaries the metal will be very strong, but it will then also suffer brittle fracture. It remains a moot point whether polycrystalline materials are multiscalar in the same sense as living organisms, where biology has apparently focused on scale-multiplication as an aid to individual survival and therefore as an evolutionary advantage.

[34] This concurrent apparent fragmentation and retention of identity appears to be *a*, or even *more* than a, principal characteristic of Nature: it corresponds evolutionarily to the concurrent splitting-up and unification of the universe through the influences and results of Einsteinian relativity.

5.2 Reunification and hyperscale

Although it is disarmingly easy for us to create a problem in an artificial domain which has no locally-indicated solution, in Nature this appears to be very rare. In some apparently magical manner, Nature never creates a problem without its resolution being close at hand. Possibly this is the clearest distinction which can be made between *natural* and *artificial* systems. Natural systems, for example, may expand until the generality of communication-restriction degrades the stability of their coherence, but fragmentation can then be avoided by recourse to scalar restructuring. Similar resolution is unavailable to an artificial system, where relationships and boundary conditions are externally imposed and therefore uncorrelated.

In a natural multiscalar system where the different scales are partially isolated it may be difficult to see how they could be reunified, but the key lies in a tradeoff between completeness of access and completeness of representation. The partial isolation of a scale does not mean that it cannot be accessed; just that access will depend on the degree to which its internal structure and processes are required to be represented externally. Consequently, it may well be possible to represent a scale very well if this requires information which is non-critical to the scale *or to its intentions*. '*Intentions*'…? Intention, of course, implies anticipation; anticipation implies intention. In Section 1.2 we related Newtonian reactivity to the maintenance of identity – to the *intention* of preserving identity through anticipation. Is it unrealistic to attribute *intention* to the individual scales of a multiscalar system? As we pointed out in Section 5.1 above, "*The natural development of multiple scales is commonly restricted to carbon-based covalent-bonded systems – to living organisms*".

Hyperscale constitutes a 'dimension', or 'dimensions', within which a unified representation of *all* of a system's scales and their interrelationships may be found. In human terms, it is the personal domain from which we view our world, and within which we can refer not only to 'a system' or 'a network' but also to its components or constituent processes. More controversially, from within hyperscale we can directly influence our environment without consciously descending through a sequence of various representational levels to the action potential of muscular activity: we *live* in hyperscale [60], and it provides the 'virtual world' within which, and from which we anticipate events and play out the consequences of our actions [61]. The reader can him- or herself create an excellent example of hyperscalar 'bridging' between *intention* and *action* by repeatedly drumming the fingers of one hand sequentially on a table. Observation of the outside of the upper forearm then reveals similarly sequential but most probably previously unnoticed contractions of the muscles responsible for the fingers' movements. Our awareness is transferred to the target of our intentions, whatever and wherever that may be.

Hyperscale *is* the correlation of multiscalar properties: an entity *is* its hyperscale. The relationships between it and an individual scale decide that scale's survival or demise, so it is in a scale's 'own interest' to interact cooperatively

with hyperscale, even if its interaction is somewhat restricted. A corollary is that the hyperscalar representation of an individual scale will be to some extent imprecise. This tradeoff between completeness of access and completeness of representation is the generic form of Heisenberg's Uncertainty Principle, which then applies to all inter-entity and inter-scalar observations[35].

5.3 Birational ecosystemics

Quantum Mechanics (QM) does not *replace* Newtonian Mechanics (NM): it *complements* it. The two of them form an ecosystemic pair, within which each is the logical ecosystem of the other [50]. The NM of classical physics is reductive towards *localization*: QM is reductive towards *nonlocality* [64]. It is tempting to propose that the two are related in a manner which is analogous to that of an organism with its ecosystem, but more correctly the analogy is in the opposite sense: it is an organism's relationship with its ecosystem which is a partially degenerate analogue of the generic coupling between a natural rationality and its ecosystemic complement – for example between NM and QM. The familiar logic of a NM system-description conceals its complementary QM description. Consequently, each NM scale of a system is associated with a QM scalar complement.

Transit between scales in a multiscalar Newtonian system is problematic. A very simple example is provided by 1 + 1 = 2. There is no general contextual manner in which the two 1's can be unerringly merged into a single 2. In a normal arithmetic context the 2 is a formally instituted result of the operator of addition, but in the less-than-formal context of scalar emergence the implied reduction in number of degrees of freedom is multifractally complex [65]. Consequently, the various recognizable scales of a natural multiscalar system, which take the form of Newtonian 'potential wells', are separated from each other by complex interfacing regions. It is only from within hyperscale that inter-scalar transit can be easily, if virtually[36] effected. The set of scales of a natural system make up a correlated hierarchy, whose global identity corresponds to its hyperscale. More surprisingly, the set of inter-scalar complex regions *also* make up a correlated hierarchy, whose global identity corresponds to a second *complementary* hyperscale. It appears that the 'virtual' inter-scalar transit inherent in the correlatory 'construction' of hyperscale depends on a generic form of *quantum error correction*, where limited scalar information is supplemented by 'hidden' relatedly-scalar complementary information to provide a complete description of the system [51].

---

[35] This process is closely related to the dual-channel quantum teleportation proposed by Bennett *et al*. [62], and for a complete system it corresponds to Feynman's 'summation over all paths' [63].

[36] … although the bridging example given above suggests that hyperscalar 'inter-scalar' transit belies this use of the word 'virtual'– as is the case for *all* truly complementary complex systems, where logical *existence* is a *derived* property and not a primary one [66].

The ultra-high level Newtonian-complex hyperscalar exchanges of a natural birational hierarchy ensure the *physical* viability of an *information*-based entity. An *organism's* survival depends primarily on the temporal stability of its environmental evaluation, but the inter-scalar consequences of perceptional errors in a *monorational* processing architecture generate instability. *The* most notable quality of a *self-correlating* Newtonian-complex architecture is that imprecision or error in either one of its *birational* assemblies can be resolved by the other. The brain and the body of an animal have both evolved from the conglomeration or multiplication of smaller organisms which exhibit *both* metabolism *and* information-processing. Collier [67] has pointed out that with time and evolution the different parts of an organism become functionally differentiated through an exchange of autonomies. The brain has seemingly ceded its major metabolic functions to the body so that it may better operate as an information-processor, and in the interests of survival the body has ceded higher-level information-processing to the specialized brain. *The mind's* essential function is to sustain the body's *physical* viability through the medium of its 'neural substrate', and stability is ensured through its high level hyperscalar exchanges. As humans, we habitually make a psychological distinction between *logic* and *emotion*, and make use of each of them to resolve the other's erroneous dead-ends and maintain mental stability [50]. The authors believe that the two high-level constituents of natural information-processing – the 'logical' Newtonian hyperscale and the 'complex' hyperscale – are the primitive precursors of this binary psychological distinction – of *logic* and *emotion*. For the remainder of this paper we will consequently adopt the terms *logical sapience* and *emotional sapience* to distinguish between the two hyperscalar primitives of *wisdom*.

5.4 Surmounting scale

It should not be presumed that the two interleaved hierarchies referred to above are necessarily *objectively* asymmetric in character. They are, however, *subjectively* asymmetric. It is worth noting that if a component C is removed from a real complementary system S, the remaining system is *not* equivalent to (S - C), as the point of view and consequently the rationality associated with C is different from that associated with its ecosystem. Removing the rabbits from a countryside environment does not simply result in the original countryside minus the rabbits! We may well make the removed rabbits very happy by protecting them from the countryside's foxes, but the foxes are unlikely to be similarly delighted. Real birational hyperscalar information-processing systems are similarly subjective in character. From a point of view which corresponds to the Newtonian 'potential-well' hyperscale, each individual scale has a 'normally' logical and approximately complete internal character, and adjacent scales are separated by Rosennean[37] fractally-complex regions.

[37] "A system is simple if all its models are simulable. A system that is not simple, and that accordingly must have a nonsimulable model, is complex." [68] [for an overview of Rosennian complexity, see references 69, 70].

Somewhat surprisingly, the view from the 'complex-interface' hyperscale will be identical. Each individual scale (now from a 'complex' point of view) has a 'normally' logical and approximately complete internal character, and adjacent scales are separated (now from a 'complex' point of view) by incompletely apparently-complex regions which correspond to the Newtonian wells[38] [50]. Straightforward monorational navigation between the different Newtonian scales of such a complementary system is precluded, as even the simplest inter-scalar route necessarily passes through at least one complementarily-rational complex region. This is the difficulty with 1 + 1 = 2 referred to in Section 5.3.

Our own human experience, however, is that *it is indeed* possible to conceptually move between different scales of an internally represented environment: so how do we do it? Hyperscalar 'bridging' clearly has an important role to play in a pre-established stabilized system, but as usual in any constructive environment its primarily 'top-down' nature requires a 'bottom-up' counterpart. These expressions – *top-down* and *bottom-up* – are habitually used in creative environments to define constructive directions, but reliance on either of them on its own is of little use. Uniquely *top-down* construction of a system brings with it a myriad of possible sub-systemic variations and the combinatorial explosion of yet more sub-sub-systemic and even more sub-sub-sub-systemic possibilities. In any practical environment this explosion is tempered by injecting a *bottom-up* sense of 'what kind of construction' the system should have[39]. Similarly, uniquely *bottom-up* construction could well produce *any* kind of system other than the one required – where the notion of 'the one required' corresponds to the injection of a *top-down* component. It should be immediately noticed that *both* of these constructional injections anticipatively couple together current action and future aim.

5.5 Intelligence, sapience and wisdom

If hyperscalar 'bridging' is to be established, we first of all need to take account of 'what kind of inter-scalar transit' we will have recourse to. As was indicated in Section 5.3, it appears that the inter-scalar transit inherent in the correlatory 'construction' of hyperscale depends on a generic form of quantum error correction, where scalar information is supplemented by a 'hidden' complement to provide complete description of the system [51]. Initiation of this kind of inter-scalar process is itself anticipative, and low-level scalar correlation can consequently be associated with the concept of *intelligence* [71].

[38] Note the differences between 'normal' NM logic and QM logic.

[39] A classic example, provided by Liane Gabora, is that if your target is to build a garage door opener, you do not start by considering chemistry!

Many different definitions of *intelligence* have been formulated. James Albus has proposed a representative definition specifically in terms of *intention* and *anticipation* that intelligence, as a 'property' of *the mind*, is:

"*The ability to act appropriately in an uncertain environment; appropriate action is that which maximizes the probability of success; success is the achievement or maintenance of behavioral goals; behavioral goals are desired states of the environment that a behavior is designed to achieve or maintain.*" [72].

It is virtually impossible to make sense of this definition, or indeed of the majority of others, without presupposing at least a degree of *internal awareness* of its constituent parts – of *appropriateness*, of *success*, of *environment*, of *intention*, of *anticipation*, and of *awareness* itself [58]. The stated central thesis of this paper is that a developmental continuity exists between primitive low-level anticipation and its high-level *aware* exposition, and that highly evolving *Anticipative Capability* and conceptual *mind* are indistinguishable. In the light of James Albus' credible definition, it is clear that *intelligence* is intimately associated with the anticipative achievement of goals through attention to environmental detail, and that both *Anticipative Capability* and *the mind* are active participants in the evolution of inter-scalar processes.

As inter-scalar correlation builds up to include *all* the scales of a system we approach the *global* correlation of hyperscale, and we will associate *multi-inter-scalar* correlation with the concept of *sapience* [71]. We should now remember that there are *two* interleaved hierarchical structures in a natural birational information-processing system, and that each of these will generate its own rationally distinct hyperscale. Consequently, we must take into account that there will be *two* distinct *intelligent* rationalities, and *two* distinct *sapient* rationalities. The two interleaved hierarchies, however, are ecosystemically interdependent, and the highest conceivable level of systemic correlation will therefore be *between* the two sapient rationalities, resulting in a *single* highest level, which we will associate with the concept of *wisdom* [71].

As inter-scalar regions are always multifractally complex [65], the *intelligent* process which leads to inter-scalar correlation is similar to the *sapient* process which leads to global scalar correlation, and an equivalently-rational *intelligence-sapience* pair sets lower and upper boundaries for a continuous range of *Anticipative Capability*. In general, *intelligence*, *sapience* and *wisdom* are *all* to some extent associated with *any* degree of information-processing, and in our further discussions of *AC* we will often refer to '*Intelligence*, *Sapience* and *Wisdom'* as a unified 'property' by the acronym '*IS&W*'.

Evolution is closely associated with the survival of a species through that of its individuals. But what advantages do *IS&W* confer on an individual? How does *the mind* of an individual help? Are *IS&W* of immediate use in responding rapidly to environmental threats? Well, not obviously. Do we really want to wait blindly until something untoward occurs, and then start trying to find a solution? If so, what is the evolutionary point of memory? Organisms sur-

vive by building up experience and assembling it into internal models of their environmental relationships. This provides two main advantages. Firstly, it effectively 'pushes' a large part of information-processing 'into the past'– into memory – by making 'ready-to-wear' threat-responses available 'off the shelf' for *IS&W* to use [73]. Secondly, it supplies extensive ready-correlated information which can be used both in anticipating the results of its own actions and those of its opponents, and in generating complicated plans of action. These prospective advantages to an organism depend not only on the availability of previously constructed internal models; they require multiply-scaled internal threat-responses to be mutually-correlated in a 'data-base' – or, rather, a 'model-base' – whose internal structure matches that of the organism's environment [74].

6 Evolving Anticipative Capability

The title of this section has (at least) two very different meanings: one refers to the historical development or evolution of the attribute of *AC*; the other targets the attribution of an *AC* which is itself evolving. The paper addresses both of these aspects, and we would maintain that the former has only been possible *because* of the latter. It is important to note that anticipation alone is insufficient to ensure survival. Gunderson and Gunderson have indicated that *intelligence* and *capability* are very different 'beasts':

> "*… one can have significant intelligence and lack capability, or vice versa*" [75],

but they also pointed out that intelligence and capability will naturally evolve hand in hand:

> "*… the intelligence and the capacity to use that intelligence must develop together*" [75].

The situation for *anticipation* and *IS&W* (*Intelligence*, *Sapience* and *Wisdom*) is similar. Neither is much use on its own, but it is also possible for *either* one *or* the other to dominate, so that awareness of future events may be insufficiently supported by the capacity to act suitably, or suitable actions may well be feasible but may remain unimplemented through lack of awareness of their value. This 'hand-in-hand' developmental nature, and the autocatalytic interdependence of anticipation and awareness, suggests that the evolutions of survivability, anticipation, consciousness, intelligence, wisdom, evolution itself, and indeed *the mind* are broadly equivalent: there is only *one* 'evolutionary process' which depends on an evolving directivity, itself mirrored in *Anticipative Capability*.

It is understandably easy to grossly underestimate the practical requirements for successful anticipation of events or conditions. Scientific education strives to persuade us that the association of a small amount of isolated data with formal models can deliver precise and accurate prediction – as in many prede-

termined temporally linearly-dependent contexts it can. Unfortunately, these successes encourage the acceptance of a conflation of *precision* and *accuracy*, whose presumed equivalence collapses in more realistic contexts. In a natural context, where *complex* multiscalar organisms are embedded in a *complex* multiscalar environment, the achievement of reasonably precise *and* accurate anticipation demands internal model-structures which reflect this bipartisan complex multiscalar character. It was pointed out in Section 5.2 that tradeoff between completeness of access and completeness of representation within a hyperscalar context is the generic form of Heisenberg's Uncertainty Principle: the involuntary conflation of precision and accuracy corresponds to a Newtonian unawareness of the generality of this tradeoff. Hierarchical correspondence between a natural environmental and its internal representation provides fertile ground for the emergence of anticipatory information [76], whose precision and accuracy rely on the quality of this 'mirroring'. The establishment and long-term evolution of suitable internal models depends ultimately on the survival of their hosts, and therefore on the availability of *Anticipative Capability*. Consequently, *the mind* has co-evolved with hyperscale, as its facilitator.

6.1 Anticipation and probability

In the lack of a deterministic description of our environment we fall back on *probability* for protection. Will 'this' event take place, or not? If its eventuality has a probability of 99% then we presume that it will. But where does the probability distribution we are using *originate*? We cannot justify its use formally – only numerically, from individual measurements of past events. Nor can we entirely separate the individual events from the probability distribution, much though statisticians would love to do so: a probability of 0 or 1 implies that event and distribution are strongly coupled; their maximal decoupling is when the probability is ½. It *could* make more sense to consider that *a particular event* and *the probability distribution* are *always* to some extent coupled. Unfortunately, we then end up with *probabilities* of *probabilities* of *probabilities* of… and we appear to be no further ahead. The most important limitation is that probability *itself* and our *belief* in it are inseparably intertwined. An announcement that the type of aircraft we are about to board is known to crash once in every ten million flights is not so likely to dissuade us from travelling. If, however, this statistic is changed to one crash in every *five* flights, we may well lose our desire to board the aircraft!

If we accept that the *probability* of an event's occurrence indicates that it will *really* take place, then the *probability distribution* can be used to direct 'suitable action'. For this to be successful the probability itself must be well constructed, which implies that sufficient intelligence has access to enough environmentally-defining information. So, given the intelligence to use *all* available knowledge, definition of the *probability* of an event and its *anticipation* are very closely related, if not the same! Probability distribution is an emergent meta-scale: it exists at a higher level than the individual events whose occur-

rences it represents. While a probability distribution is constructed from *past* events, anticipation can make use of it in conjunction with *current* and anticipated *future* states to verify the suitability of specific actions. Anticipation injects *expected future information* into a *future's past information* through *current* decision-making, thus adding to, or changing the population of events that the probability distribution represents.

The combination of *awareness of action* and *anticipation of consequences* which is available through the evolution of *Anticipatory Capability* makes it possible to manipulate future probability distributions by injecting preferred bias into the otherwise-presumed independence of individual events. Survival of an individual or species depends on the development of just this capacity – the capability of *the mind* to have power over potentially threatening contexts.

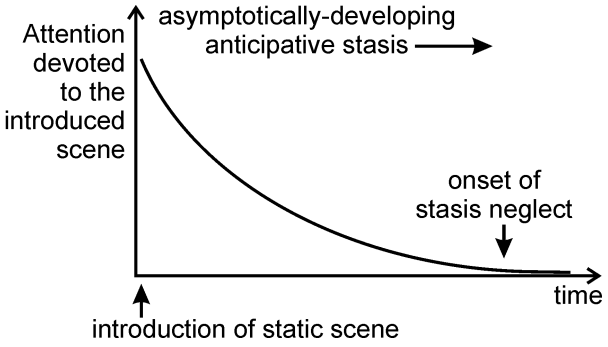

Figure 5: The asymptotic progression of anticipation of a static or repetitive scene, from heightened awareness on the scene's introduction to the establishment of stasis and the scene's 'disappearance'.

6.2 Asymptotic anticipation, dynamic awareness and stasis neglect

Anyone who has sat waiting at a red traffic light will probably have noticed that if you just sit staring in front of you the red light slowly fades and disappears – until you move your head, and there it is back again – *stasis* fades easily from our awareness. Similarly, if you sit looking out of a train window, the repetitive passage of electric pylons slowly drifts out of your attention, leaving space for other thoughts. Any model of dynamics is static, as Zeno pointed out [77]: repetitive movement is similar to stasis in its lack of impact on our attention. This has a direct bearing on the attentive relationship between observation, anticipation, *awareness* and *consciousness*, in which consciousness is *generated* from awareness through anticipation. If an entire scene surrounding us is static, or is dynamically 'static' in Zeno's sense, then its every modification can be predicted on the basis of a previous observational chronicle, and anticipation reduces to historically-based repetition. *IS&W* are then entirely irrelevant (as they are, for similar reasons, in a digital computer). More precisely, in this situation, a possibly multiscalar observational structure effectively collapses into a single algorithmic or quasi-algorithmic recursion, de-

void of mystery or surprise. Robert Rosen [15] pointed out that this is the very antithesis of life: it entails neither *IS&W* nor any incentive for their evolution.

Figure 5 illustrates what happens when we are exposed to a new, but static or repetitive scene. As time goes on our information-processing reduces asymptotically to a realization[40] of stasis: the odds on anything happening progressively shrink towards zero. Initially our 'processor' will track the scene carefully, as there is no prior chronicle upon which to base any supposition, but after a while it will conclude that there is nothing of immediate interest and replace *attention* by *neglect*. This phenomenon of 'stasis neglect' has the important effect of making it possible for repeated actions to be transferred out of regions of the brain which *consciousness* 'keeps track of' and for them to become quasi-automatic [49]. Many of our bodily functions continue quite happily without (normally) disturbing our awareness – for example the heart's beating – although a degree of attentive modification may be feasible[41]. A major advantage of this 'automation' of learned actions is that it leaves space in consciousness for exploration and evolution, and for events or actions whose immediacy is more important.

But what happens to the static or repetitive aspects of our surroundings which we neglect – those which are dropped from our awareness? Do they disappear?[42] Presumably not, 'in reality'. But, also, not necessarily from our retinal-neural system. The traffic signal's red light has still arrived at our eyes, even if we 'choose' to ignore it, and in many cases we can recall details of earlier events which we were not aware of at the time. More reasonably, attention does not exist 'in a vacuum' – at least, not from the 'point of view' of our mind. It is part of a complementary system, where the 'spotlight of awareness' is focused on a limited region of a far wider tapestry. In accordance with the precepts of Section 5.3, attention is partnered by its 'ecosystem' of 'currently neglected stasis', which makes up a pool of globally relevant but locally inaccessible information. This ecosystemic association is reminiscent of a wide range of 'phenomena', notably:

- The existence of 'hidden variables' behind the explicit components of physical models.

[40] Although the word 'realization' could apparently be better replaced here by 'establishment', or even 'non-realization' (!), this is a very difficult point, as it relates to the possible scalarity of awareness, and of its hypothetical hyperscalarity!

[41] This relates to depictions of consciousness as a 'supervisor' which 'keeps track of' neural activities. Consciousness certainly permits a degree of structured active control, while awareness is more passive in character – c.f. the implications of Charles Peirce's 'firstness', for example [78].

[42] … a question which is somewhat related to 'If a tree falls in the forest when no one is there, does it make a sound?'!

- The 'background dimensions' which take part in quantum error correction.
- Freud's 'unconscious mind', which is fed by unresolved events and information.

Our capacity to focus on a chosen part of our environmental complexity is critically grounded in the assumption that 'we don't have to bother about the rest', or, more precisely, in our anticipation that 'the rest' will not substantially diverge from stasis while we deal with the current objects of our attention. This *complementary* nature of an anticipative attentive/neglective system and its ecosystemic grounding has momentous implications for any account of the *self* and its environmental relationships, as we shall see.

Stasis degrades attentiveness through its effect on anticipation. Our *awareness* of stasis depends on the adaptive anticipative generation of a recursive observational chronicle: awareness and consciousness depend on *Anticipative Capability*. Stasis can wipe out the attentive 'scalar-local' application of intelligence, but its effect on 'global' sapient processing must surely be even greater. Boredom at a single scale is incomparable with the possibility of boredom at *every* scale! Anticipation is vital for survival, but it is the *necessity* for anticipation which drives the evolution of *IS&W* [71].

Terrence Deacon has famously suggested that:

"*Our self-experience of intentions and 'will' are not epiphenomenal illusions. They are what we should expect an evolutionlike process to feel like*" [2, p.458].

If we may, we would modify his conclusion to read:

"*Our self-experience of intentions and 'will' are not epiphenomenal illusions. They are what we should expect the anticipative evolution of IS&W to feel like*".

Similarly, we find it reasonable to modify Descartes:

"*I think, therefore I am*" [79]

to read:

"*I anticipate, therefore I am*".

6.3 Mirror neurons, autism and empathy

Anticipative Capability depends on the availability of internal models – not only of an organism's surroundings, but also of *itself* and its relationships with the environment. Nature has avoided major phenomenological conflicts due to relativity by its evolution of a 'physics' which as closely as possible 'mirrors' global effects in local ones. Newtonian physics presumes that the local and the global are mutually consistent [50]. While this may provide sufficient support for anticipation in an environment consisting uniquely of low-complexity entities, the introduction of informationally-intense network-based organisms

presents anticipative difficulties. Recent neurophysiological research [80, 81, 82, 83, 84, 85, 86, 87], however, has detected features of the mammal brain – *mirror neurons* – which enable them to 'mirror' another organism's neural complexity. This imparts the capacity to learn through imitation – one of the most important ways in which children, for example, assimilate practical and social skills.

A huge amount of interest is currently focused on mirror neurons – first described by Fadiga *et al.* [80] following their observation that a specific subset of macaque monkey motor-skill neurons are activated, not only when an action is performed, but also when watching another monkey performing the same action. Kohler *et al.* [81] have reported that even the *sound* of the action being performed is sufficient to trigger activation. Oberman *et al.* [82] have proposed that mirror neurons are vital to the development of social skills, as they provide the means of learning by example, and their importance in the development of *IS&W* and in the operation of anticipation cannot be exaggerated. *Individual* neurons, however, do not 'mirror' whole segments of cognition – they contribute to a neural sub-system, usually referred to (in the singular) as the Mirror Neuron System (MNS). Williams *et al.* [83] have linked the empathy-debilitating occurrence of *autism* to MNS defects, and there is now extensive documentation of involvement of 'the MNS' in a wide range of cognitive phenomena. It is not at all clear whether *the same* MNS is involved in all of these observations; indeed, it starts to look as if the function we refer to as 'mirroring' is a fundamental neural strategy, rather than the property of a single limited sub-system. Molnar-Szakacs and Overy have suggested that

"*… musical experience involves an intimate coupling between the perception and production of hierarchically organized sequential information, the structure of which has the ability to communicate meaning and emotion*" [84],

and that this may be mediated by the (or *a*) MNS. Which begs the question: are there a number of differently oriented MNSs in the brain, or only one 'audio-visual' MNS, or is 'neural mirroring' the mistakenly-specific interpretation of a more general cognitive strategy?

Fadiga *et al.* observed 'mirroring' effects in region F5 of the macaque premotor cortex – comparatively early on in the vision-perception-planning-motor sequence of action control. There appears to be a close resemblance *in this sense* between 'neural mirroring', 'dream-mechanisms' and the techniques which are habitually used to train Artificial Neural Networks (ANNs). Rock [88] has pointed out that many of the currently proposed mechanisms of dreaming assume that either random or structured inputs from different parts of the brain [89] are 'injected' early on into the visual processing chain, where they replace the 'normal' retinal input. Both visual 'mirroring' and dream scenarios may be influenced by input from other senses, however, for example by sounds or smells. Similarly, ANNs are trained to operate 'correctly' by replacing their 'normal' environmental input by artificially-constructed but 'sufficiently representational' case studies or templates – care being taken to

reduce any consequent inadvertent injection of unrepresentative (i.e. 'other-sensory') information. Are all these – 'mirroring', dreaming and ANN-learning – examples of a very general neural strategy which is related to *scale*? Multiscalar systems are not only characterized by their 'upward' *emergence*, but also by their 'downward' *slaving*[43] [90]. *Slaving* forces the properties of sub-systemic elements into mutual correspondence, and stabilizes their emerged or emerging higher scale[44]. Is this the 'mechanism' which underpins 'mirroring' and dreaming? – it most certainly *is* the mechanism which underpins ANN-learning. If so, then our understanding of cognitive processes will depend on assimilation of the scalar properties of hierarchical systems[45] into *any* theory of *the mind*.

A system is grounded in the functions of its various sub-systems or elements, but its operation is usually described at the system level, and not in terms of its constituent functions. Which leaves us with a problem: 'where' do the system-level properties we refer to 'reside'? A computer processor, for example, consists of a large number of interconnected primitive logic gates, *and nothing else*. So 'where' are its system-level properties? Well – nowhere – except in our own minds – unless we wish to deconstruct them to their 'equivalents' of sets of gate operations[46]. This problem is not, however, restricted to logical processors: it is a very general property, of both 'machines'[47] and organisms. An automobile engine is built up from numerous sub-systems. 'Where' is the 'engine'? And no, we do not mean 'where is the metal out of which it is constructed?', we mean 'where is the *unified* entity which performs the function we normally associate with an engine?'. Well – again – nowhere! And 'where' is the cognition of an animal? Again – 'nowhere' – except in our own minds – or, in *its own* mind… So, to cut a long story short, we can reasonably (!) expect *elemental* functions to be embodied, but *not* higher-level concepts. Unfortunately, this now poses us with a problem in relation to 'neural mirroring'. A

[43] … also referred to by Stan Salthe, amongst others, as 'downward causation'.

[44] As an example of *slaving* we may cite the properties of the 'elementary particles'. There is little reason to suppose that in the immediate aftermath of the 'Big Bang' all of the 'electrons' resembled each other to the extent that Physics now describes. It is more reasonable that the restriction of their properties to a small set of quantum numbers is a consequence of the contemporary multiscalarity of Nature, and of its resultant downward pressure towards elemental conformity [91].

[45] It should be noted that neural processing does not necessarily depend on hierarchical relationships, but it is unclear whether the brain *necessarily* operates in a more heterarchical manner, or whether the deviations from Nature's and its own generally hierarchical character are evolutionary artifacts.

[46] … and even then there is a problem, as the gate's operation 'itself' suffers from the same difficulty!

[47] We use the term 'machine' in a very loose way in this location, and not with the precise sense defined by Rosen [15].

number of current authors of NMS-related papers [e.g. 85, 86] focus on the possibility that the usual philosophical position, that *concepts* are necessarily abstract and symbolic, *may be wrong*. Gallese and Lakoff, for example, argue that

"*… a disembodied, symbolic account of the concept of grasping would have to duplicate elsewhere in the brain the complex neural machinery in three parietal-motor circuits, which is implausible to say the least*" [85]

and that

"*If all this is correct, then abstract reasoning in general exploits the sensory-motor system*" [85].

So, the discovery of mirror neurons has not only produced an upheaval in the biological and medical aspects of neuroscience, but it even impinges on philosophy. The idea of 'embodied concepts' is foreign to neuropsychology and neurophysiology – and, of course, to the simulated argument we ourselves presented in the last paragraph. *But not to a bi-sapient rendering of the brain's information-processing structure*. Within the birational description of neural processing which lies at the heart of this paper, *every* property and phenomenon is *both* derived from *and* embodied in the neural material itself: the complementary ecosystem of information-processing *is* its material substrate. As we will see, the inter-correlation of *bi-sapience* provides a very satisfying account, both of empathy and of its autistic absence, and this leads us to question whether 'neural mirroring' occurs *only* between different social actors, or whether its generic form *also* characterizes *every* internal process in an individual brain.

6.4 Bi-sapience and empathy

Empathy is vital to *Anticipative Capability*. Its presence in a birational information-processing system contributes much of the grounding from which anticipation operates. However, before we can address its importance we must first look at its relationship to bi-sapience, to Rosen's (M,R) model of an organism [15], to the neural embodiment of concepts [85, 86], and to the phenomenon of *bonding*.

Birational information-processing generates a complementary pair of cross-scalar correlations we have called *sapiences* – one derived from the assembly of 'normal' Newtonian scalar levels, the other from the assembly of their inter-scalar complex interfaces. These two interact and create a *singular* unification of the entire system, which we have referred to as *wisdom*. Each sapience functions as an informational ecosystem for the other, and consequently the bi-sapient unification process *resembles* 'neural mirroring' – but with one vitally important difference. 'Neural mirroring' is usually described within a *monorational* paradigm, where the only 'mirroring' possible is between *logic* and *logic* – thus the word 'mirroring'. In a *birational* correlation, however, each sapience

– *logical* or *emotional* – references its *complement* – *emotional* or *logical* respectively – and the word 'mirroring' is inappropriate.

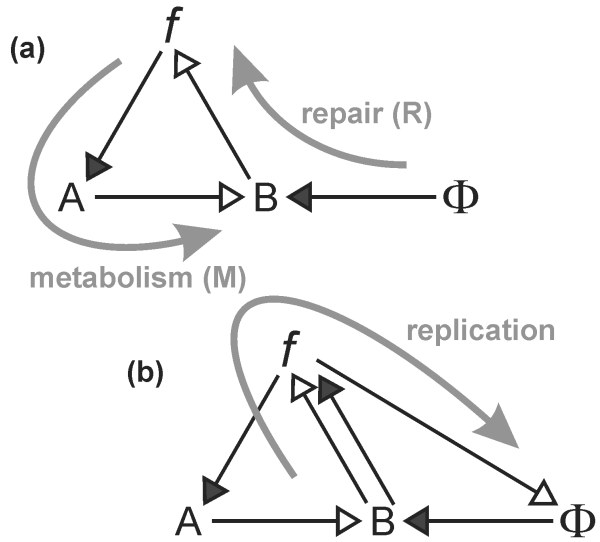

Figure 6: (a) Rosen's Metabolism and Repair (M,R) model of an organism, and **(b)** Its extended (M,R,R) [48] form when Replication is added into the relational scheme.

At first sight, therefore, there appears to be a major difficulty in explaining empathy, in that the *logical* subject of 'neural mirroring' addresses the observed form, or *logical* content of another's actions, and not directly the *emotional* content. So why does use of the expression 'empathy' refer to *emotional* and not *logical* inter-personal equivalence? In any case, emotional content is ostensibly externally inaccessible, in which case we can apparently only observe or take account of *signals* or *indications* of another's emotion – we can only capture the *logically* interpreted content. A possible explanation in terms of bi-sapience would be that inter-personal 'logical mirroring' is directly related by an observer to his or her own internal emotional state, and then conclusions are drawn about the other's emotion by supposing that both people are 'similarly constructed'. Unfortunately, although this hypothetical process does make use of a presumed *internal* logical-emotional complementarity for both observer and subject, *external* inter-personal 'mirroring' is presupposed to be mono-rationally *logical-logical* and not birationally *logical-emotional*. Granted, this description could explain a restricted form of empathy, where we *assume* another's emotional state without having any confirmation, but to find a more suitable bi-sapient explanation we must first look elsewhere – at Robert Rosen's (M,R) model of an organism [15]: we must take into account that both observer and subject are *alive*.

Rosen [15] carefully constructed his general (M,R) model of an organism in terms of mathematical mappings which represent Metabolism (M) and Repair (R). His *metabolic* functor *f* (see Figure 6(a)) maps from environmental-input

[48] (M,R,R) is these authors' extension of Rosen's (M,R) nomenclature to include Replication.

set A to set B, and then the *repair* functor F maps from set B back to the metabolic functor *f*. Further development added a third mapping to represent Replication, and he concluded that in certain circumstances this could be derived from the already-present Metabolism and Repair, without introducing a new separate mapping, as illustrated in Figure 6(b). The required condition has been described by Louie as

"*… the abstract version of the one-gene-one-enzyme hypothesis*" [92]49.

Critical analysis [93, 94] of Rosen's model reveals that it can be reformulated as a figure-of-eight circulatory system (Figure 7(a)) between a *mechanism*, as defined by Rosen [15, p. 203], and its *complement* (Figure 7(b))[50]. The central region of the model then exhibits a four-fold interactive process between *software flow*, *hardware flow*, the *hardware induction of software flow* and the *software induction of hardware flow* (Figure 7(c)). The last of these four components – the *software induction of hardware flow* – is completely alien to inorganic physics, and it *only* appears in a living system [94]. The four-fold interactive process itself is *also* unique to living systems.

Careful comparison between the properties of Rosen's (M,R) model and those of a bi-sapient information-processing system leads to the conclusion that an organism is an intimate complementary coupling between a *mechanism* and its *ecosystem*. Although Rosen did not explicitly address the question of *scale* in organisms, his model is already birational [94]. It is important to note that the complementarities of both Rosen's (M,R) model and bi-sapient information-processing systems are between the material-induced manipulation of information and the information-induced manipulation of material – between information-processing and its substrate of material-processing: between *cognitive processing* and its *embodiment*. The *embodiment of concepts* proposed by Metzinger and Gallese [86] and Gallese and Lakoff [85] is a characteristic property of bi-sapient information-processing!

[49] We refer the reader back to Section 4.2: 'Genetics versus gene-protein mapping'.

[50] It should be noted that the complement of a mechanism is *not* an organism: the complement of a mechanism is the ecosystem within which it operates [94].

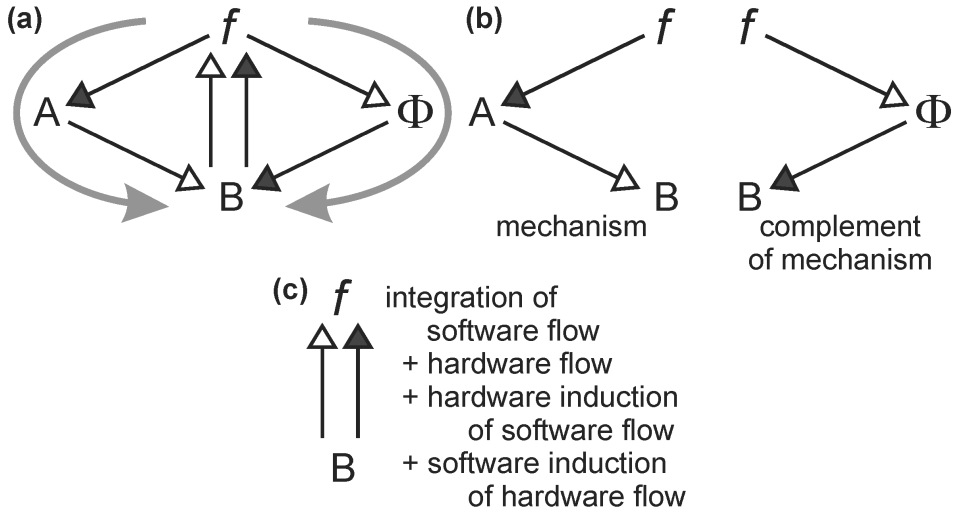

Figure 7: (a) The authors [93, 94] 'figure-of-eight' reformulation of Rosen's (M,R,R) model, **(b)** A comparison between mechanism and its complement in terms of Rosen's functor diagrams, **(c)** The interactive four-fold mid-region of the 'figure-of-eight' reformulation [94] of Rosen's (M,R,R) model.

Until now we have only referred to bi-sapient information-processing within a *single* organism. But empathy occurs between *different* organisms: it is a possibly asymmetric common property of *at least two* notionally independent systems, grounded in their (inter-) communication. So, to start with, how can we characterize *communication*? Unfortunately, the word communication is applied very loosely to a wide range of different situations without being consistently defined. The transmission of information by television, for example, is usually classed as communication. However, communication *must* be bidirectional, but television is not bidirectional *per se*. Imagine a mother on the beach when her small son begins to run towards the sea. She calls after him to come back, but unless he reacts to her call or replies in some way she cannot know if he has heard: *both* directions of information-transmission are required to affirm communication. Communication is consequently a *meta-description* of its underlying individual directions of information-transmission.

Let us now picture communication between a pair of bi-sapient systems, as illustrated in Figure 8(a). The two are separated by a communication-medium which inhibits direct interaction. Each system implements internal bi-sapient processing, and their vision permits observation of the *logical* content of the other's actions, body language and facial expressions. Up to this point in our description the two systems are both nominally and functionally independent of each other, but we also know that this is not necessarily the case – either for humans or for a large number of other animals which form lasting strongly-bonded pairs – and that an individual's actions may contradict the logic of personal survival[51]. How in our current 'model' can we represent *bonding* – the

[51] A relevant example of personal sacrifice is provided by the hypothetical reactions of soldier in battle, threatened along with his comrades by an enemy machine gun. Intel-

extreme realization of empathy? Bonding is not a personal 'possession': it is a phenomenon 'in common' – a *meta-logical-emotional* 'unification', which should therefore operate through bi-sapient correlation. Figure 8(b) suggests how this could occur. Each individual correlates its *emotional* sapience with its interpretation of the other's logical content, creating a reinforcing correlatory circulation. The resulting inter-personal figure-of-eight process operates in the same way as the bi-sapient re-mapping [94] of Rosen's (M,R,R) model, *creating a new living entity*: a 'meta-person' – a bonded pair.

Empathy supports not only the social coupling between first-person logic and second-person emotion, but also the establishment of even wider social coherence through less direct routes. The attribution of empathetic failure to Mirror Neuron System defects [83] implies that 'neural mirroring' is a major facilitator of intra-social communication:

"*The mirror neurons … dissolve the barrier between self and others*" [87].

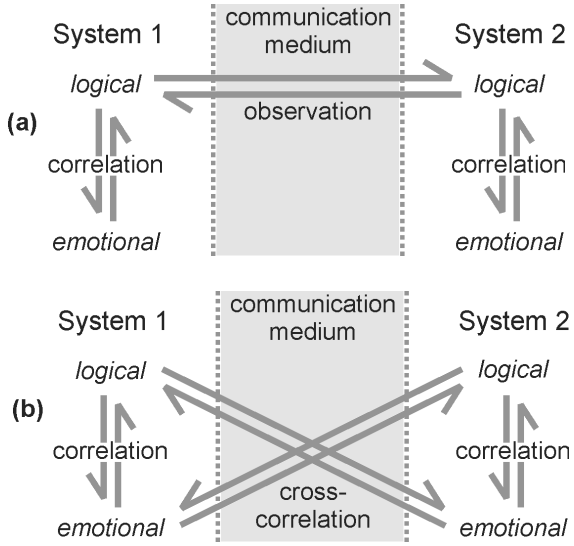

Figure 8: (a) 'Quasi'-empathetic communication between a pair of bi-sapient systems, and (b) 'Real'-empathetic creation of a bonded 'meta-person' through 'figure-of-eight' logical-emotional coupling operating in the same way as Rosen's (M,R,R) model of an organism.

The recognition of another's emotional state, which depends on being able to successfully correlate the logic of action with its associated emotional poten-

ligence may tell him that if he wants to see what is happening he should put his head up and look; his sapient instinct to survive may moderate this choice of action and use his intelligence to keep his head down; his wisdom may well support his sapient conclusion, but even so cause him to rush forwards to disable the gun and thus save his comrades – even though he is aware that in doing so he may lose his own life.

tial, equips us with the ability to 'understand what others feel'. It does not take much imagination to extend the consequences of empathy to conceptual societies of numerous individuals[52], with their multiplicities of more-or-less bonded pairs and groups and their complex interweavings of individuals, alliances, cooperation, contention, rights and responsibilities – but fortunately such an enterprise lies outside the purview of this paper.

It is now easy to see why empathy is vital for successful anticipation. Although Newtonian physics makes it possible to predict with great precision and accuracy many future states of purely inorganic systems, the internal complexity of organisms makes any attempted anticipation of their actions worthless in the absence of an internal model of their emotional states. This most particularly applies to the anticipation of *human* actions, given the extreme complexity of the human psyche. It is notable that members of long-standing human bonded-pairs become extremely good at knowing what their partner is thinking or will do in particular circumstances: their empathy-based inter-personal anticipation is excellent. *Anticipative Capability* can be directly related to the quality of relevant internal models, and its historical evolution maps and coincides with the evolution of neural capability in general. Figure 9 illustrates a hypothetical evolutionary development of *logical* and *emotional Anticipative Capabilities*. *Emotional AC* increases as the evolution of empathy is driven by the progressive increase in neural complexity, and most particularly emotional complexity, while *logical AC* saturates. Where logical anticipation once held sway, *empathy* is now vital in combating the (apparently) comparatively late appearance of *free will*[53].

The communication-medium indicated in Figure 8, along with both the 'transmitting' individual's planning and motor control and the 'receiving' individual's sensing and interpretation, acts as a complex filter, which may reinforce or degrade both the inter-personal 'messages' and the strength of the resultant bonding. If effective operation of the Mirror Neuron System is indeed responsible for the development of social skills and empathy, then *autism* is not a condition which *develops* – it is a natural initial state.

[52] The reader should note with admiration our careful escape from the violent contradiction between our profession of ease and society's horrible complexity, through use of the expression '*conceptual* societies'!

[53] Bruce Edmonds [95] has pointed out that "*our intelligence evolved (at least partially) to enable us to deal with social complexity and modeling 'arms races'*", and that "*There is a clear evolutionary advantage in being internally coherent in seeking to fulfill ones goals and unpredictable by one's peers*".

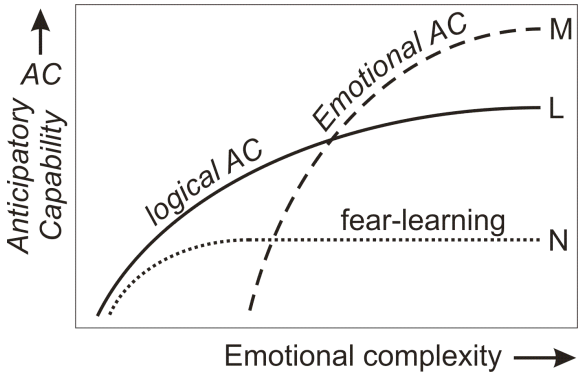

Figure 9: The evolution of logical (L) and emotional (M) Anticipative Capabilities with increase in human emotional complexity. Line (N) corresponds to the early development of emotional 'fear-learning' [96] as a primitive fast but inaccurate anticipative response to incipient danger. Fear-learning still provides an independent emotional route to rapid reaction, which bypasses the normal but comparatively slow processing of the cortex.

Autism is usually progressively recognized when a child's observed inter-personal skills and aptitudes begin to deviate from those associated with 'normal development'. Within a bi-sapient description we would expect autism to be the natural state of a newly-conceived infant, who has not yet sufficient experience of inter-personal feedback to take part in the generation of interrelationships. If this hypothesis is correct we would expect to find evidence of autism in studies of feral children. Opinion in this area is mixed, but it is notable that the Introduction to an Internet site devoted to feral children makes a clear link between feral children and autism: it points out that

"*Itard's detailed records of his work with Victor, the Wild Boy of Aveyron are generally considered to be the first documented account of an autistic child*" [97].

One of the first connections between mother and new-born is the 'mirroring' of a smile [98], within which the baby discovers that it can influence the world which surrounds it and become part of a 'meta-person'. Consequently, we suggest that a progressive increase in empathetic capability and the demise of autism is part of the natural course of a child's growth. Much of the development of empathy depends on successful coordination between the 'transmission of emotion' and its 'reception', where the communicational filtering we referred to earlier plays a vital role. A happy child with a sad face, or an angry mother with a smile, does nothing to help the establishment of bonding, neither does the ill-tempered punishment of an ostensibly helpful, but unsuccessful act. Above all, temporal consistency – whether of stable or unstable action-reaction correlation – is vitally important in establishing trust. In a 'reflective' scenario, it is easy to imagine why the *teacher* has a critical role in learning.

6.5 Auto-empathy and self-observation

The entire set of *Anticipative Capabilities* of a birational information-processing system, and therefore of the brain itself, are embodied in the complementary structure of its various inter-actions and intra-actions; between its local scales and their local ecosystemic counterparts; between its local scales and their associated intelligences; between its multiple scales and their unifying sapiences; between its bi-sapiences and its singularity of wisdom. Experimentation in the field of *neural mirroring* compares an observed *external* event with the neural activations of its *internally*-driven 'image'. We believe that an analogous comparative process takes place *purely internally* in the brain between *all* of its various quasi-independent subsystems, and that it is this which unifies neural activity into the *anticipative singularity of wisdom*.

Every aspect of the brain's information-processing resembles *neural mirroring*, but the 'mirroring' or correlation is not directly between *logical* contents, it is between *one* logical content – an event, or artifact, or a situation – and its complementary *emotional* ecosystem. Experimental measurements of 'mirroring', however, correlate observations of a *logical content* with the *logical 'image'* of its emotional complement. In principle, it should be possible to observe correlated activations in the brain which correspond to those of *neural mirroring* experiments – but there is a problem. *Neural mirroring* experiments reported in the literature compare for a single subject the two measurement pairs {*action* & neuron activation} and {*observed action* & neuron activation}, and the resulting *action/observed-action* 'mirror neuron' conclusions are drawn because neuron activation is measurably the same in both cases. If the *neural mirroring* is purely internal, however, then the two pairs {*one 'thought'* & neuron activation} and {*another 'thought'* & neuron activation} are indistinguishable, as neither the two '*thoughts*' nor the two activations can be independently observed. This makes direct confirmation of internal *neural mirroring* somewhat elusive, as it must (presumably) rely on the subjectivity of an experimental subject's reporting ('I am thinking *this* or *that*') rather than on the more objective response of a measuring instrument (which measures *this action* or *that observed action*).

'Empathy' refers to emotional 'mirroring' between different organisms. If intra-neural correlation is indeed driven by 'mirror-like' processes, then we would expect to find related phenomena between the various neural subsystems – to find a kind of *auto-empathy* – but here would be a *direct* process: literally an *embodied* process. In Section 6.4 of this paper we described the 'figure-of-eight' communicational unification of bonded-pairs, and its extension to societies with all the complex interweaving of individuals, alliances, cooperation, contention, rights and responsibilities that would entail. In a closely related sense, we would expect intra-neural 'mirroring' to result in similarly extensive complexity. Inter-sapient correlation is continuous and nominally infinite in its process of generating the singularity of wisdom, from *logic*, to *emotion*, to *logic*, to *emotion*, …, but in addition to stabilizing the birational system it also arguably supports a sense of continuity, of completeness, of logic

and emotion 'being in tune with each other'. This kind of rational-emotional 'peaceful coexistence' is only feasible within a birational environment. Correlation between the two sapiences permits us to use implicate emotion to extricate ourselves from a logical cul-de-sac, or explicate logic to resolve emotional difficulties.

Auto-empathy is invaluable in steering our actions between the extremes of 'pure' logic and 'pure' emotion. Consequently, it seems likely that the early evolution of a simple and therefore extreme form of *auto*-empathy is at the root of being able to switch between logical and emotional responses in 'fear-learning' (see the reference to independent primitive emotional anticipation in Figure 9). LeDoux [96] describes as an example that if we suddenly see a brown stick resembling a snake on the forest floor, then past 'fear-learning' permits us to rapidly jump out of the way to promote our safety. The amygdalic 'hard-wiring' which facilitates this kind of rapid reaction to possibly-threatening eventualities predates, and now physically bypasses, the comparatively slow information-processing of the cortex. The amygdala is one of the oldest neural subdivisions, and it is associated with the provision of primitive emotional response: it is the brain's primeval anticipative centre. Although we habitually associate human anticipation with conscious action, it is grounded in the *less aware* operation of early neural structures.

The nominally infinite inference of *logic*, to *emotion*, to *logic*, to *emotion*, … in birational correlation is again reminiscent of Rosen's (M,R) model [15]. Rosen has explained in great detail how both Newtonian physics and *life* depend on the 'truncation' of infinitely recursive chronicles. As he describes, Newton's Second Law collapses the state of a particle, which is a nominally infinite series of variables, down to only two – position and velocity. Matsuno [99] has generated a self-consistent view of 'reality' which is based on the interpretation of observation as a mutual measurement, and within which the Heisenberg impossibility of observing quantum particles without influencing them finds a natural home.

Interactions between the two sapiences of a natural hierarchy are an example of mutual observation, and of indirect mutual self-measurement. As we indicated above, their correlation is continuously recursive, from *logic*, to *emotion*, to *logic*, to *emotion*, … corresponding to the system *observing itself observing itself observing itself observing itself*… So how is this apparently inconclusive infinite recursive chronicle related to Rosen's description of truncated sequences? Well, this one becomes truncated *as well*. We have described in Section 6.2 how the anticipatory observation of stasis results in the replacement of *attention* by *neglect*. *Novel* information is carried between the *partially-isolated scalar levels* of a birational system 'on the back of' pre-existent *order* [50, 100]. The result is that the entire birational structure becomes to a large extent frozen, and the *logic-emotion-logic*… self-observational sequence succumbs to *stasis neglect* and self-truncates, leaving behind a quasi-autonomous self-observation: the *self*. Inter-scalar *novel* information – which characterizes 'emergence' in an information-processing system – contributes to compara-

tively slow evolutionary changes in the 'frozen' structure, and in this way the quasi-stability of the *self* is guaranteed. Remaining temporal effects of novelty in the truncated self-observation are reminiscent of Deacon's description of our self-experience of intentions and 'will' as "*what we should expect an evolutionlike process to feel like*" [2, p.458].

Rosen [15] detailed how the 'phenomenon' of Newtonian physics 'resides' in the truncation of the infinity of chronicles which define a particle's state, and how at a higher organizational level the 'phenomenon' of *life* 'resides' in the truncation (or stasis neglect) of the infinite process-circulation of his (M,R) model. Similarly, but at a yet-higher organizational level, the 'phenomenon' of *self* 'resides' in the truncation by stasis neglect of the infinite self-observational sequence of logical-emotional correlations in a birational information-processing system.

6.6 Ecosystemic birationality and the brain

If a generalized 'template' for anticipative information-processing is indeed ecosystemic, how does this relate to our own *Homo sapiens* brain? We saw in Sections 5.3 and 6.5 how interaction between the two hyperscalar sapiences of a birational information-processing system resembles the way we alternate between explicate logic and implicate emotion in extricating ourselves from mental *cul-de-sac*s. We do not maintain that these two sapiences *are* logical and emotional 'intelligences' – to suggest that would be to ignore the scavenging meanderings and cannibalizations of evolutionary development [91] – we propose, rather, that the general birational model provides a guiding template for the evolution of information-processing, much as the signpost at a road junction indicates which is probably the most useful way to go, rather than the precise compass direction of our desired destination. However, we *would* expect to find material structures in the brain which relate, if distantly, to the duality of ecosystemic rationality.

It has long been known that the neural tissue is split into two separate parts – into two hemispheres which display different informational characters. Whilst being far from conclusive, it is more than interesting to note that the two neural hemispheres apparently do tend, in general, towards a bilateral ecosystemic distribution of information-processing which is reminiscent of the two sapiences. While there are exceptions, in general the left neural hemisphere processes information in a linear, sequential, logical, symbolic manner: it is specialized in

"*… verbal skills, writing, complex mathematical calculations and abstract thought*" [88, p. 124].

The right neural hemisphere, in general, processes information more holistically, randomly, intuitively, concretely and nonverbally: it specializes in

"… geometric-form and spatial-relationship processing, perceiving and enjoying music in all its complexity, recognizing human faces, and detecting emotions" [88, p. 124].

The two hemispheres are normally connected together by the largest nerve tract in the brain: the *corpus callosum*, which contains more than 200,000,000 axons [101]. Studies carried out in the 1940s following sectioning of the *corpus callosum* in human patients [102] as a treatment for intractable epilepsy [103] intriguingly indicated that this massive neural intervention resulted in no definite behavioral deficits. Later experiments carried out by Sperry *et al.* [104] provided even more startling results: the 'split-brain' subjects of neural bifurcation provided direct verbal confirmation that the left and right hemispheres afford separate domains of consciousness. This apparent confirmation of conscious duality, however, should be considered in the light of the investigative procedures themselves. Many of the experiments only presented the human subjects with information which related to the presumed operations of the right hemisphere, and the subjects' related (or unrelated!) comment was only elucidated *a posteriori*.

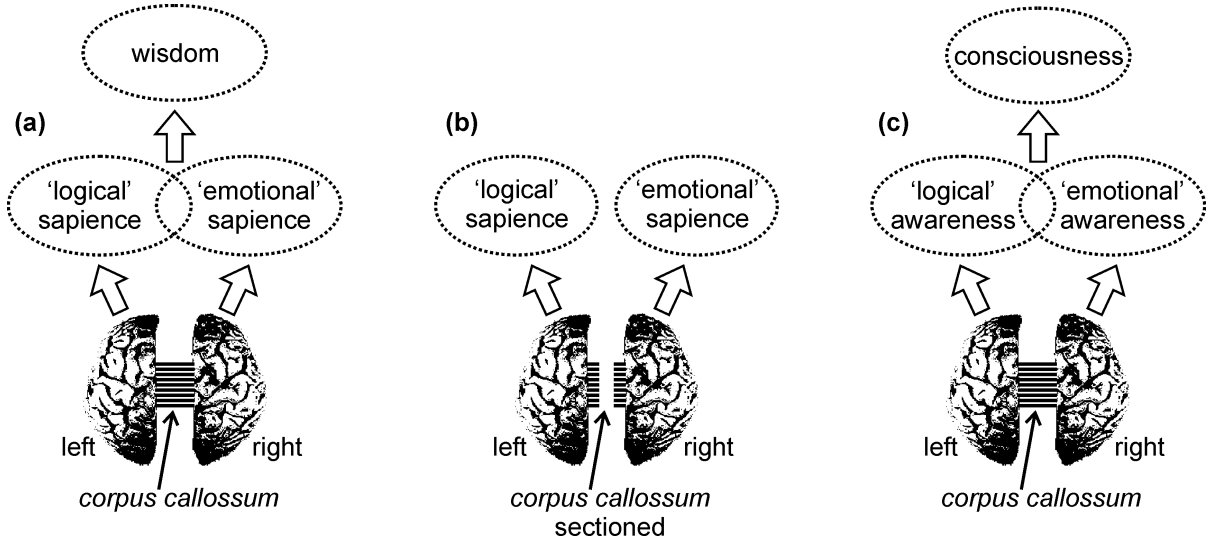

Figure 10: (a) Emergence of the singularity of wisdom through interaction between the two interacting sapiences of a birational system, **(b)** Destruction of the singularity of wisdom when the *corpus callosum* is sectioned, and **(c)** 'Normal' (pre-sectioning) auto-correlation of the two individual hemispheric awarenesses resulting in the singularity of consciousness.

Sperry *et al.* [104] suggested that their results confirmed the attribution of different consciousnesses to the two neural hemispheres, but there appears to be no concrete evidence as to whether these two conscious 'states' were experienced by the subjects sequentially or simultaneously. We question whether the subjects' experiences corresponded to 'normal' unified high-level *consciousness*, or whether in the absence of the *corpus callosum*'s coupling they were related to somewhat lower-level, less abstract *awarenesses* more inti-

mately coupled to the processing biases of the individual hemispheres. This latter conclusion would support the hypothesis that birational processing is indeed relevant to the brain's operation. If, as we have suggested, anticipative information-processing in each of the hemispheres is associated with the one of the complementary sapiences of a birational system, then the *corpus callosum* appears to constitute a 'neural substrate' for the inter-sapient correlations which lead to 'emergence' of the singularity of wisdom (Figure 10(a)). If so, then sectioning the *corpus callosum* should destroy the inter-sapient correlations and consequently extinguish wisdom (Figure 10(b)). Anticipatory sapience requires awareness, and opinion is notably undivided as to the singularity of 'normal' consciousness. Sperry *et al.*s' [104] experiments *do* appear to provide *prima facie* evidence for the existence of the two different sapiences, corresponding to two independent awarenesses which auto-correlate in the 'normal' cross-coupled brain to give a singular experience – that of consciousness (Figure 10(c)).

It is more than interesting to note where the *corpus callosum* is located in the brain. It sits above the brain stem and the associated ancient regions of the brain – for example the *amygdala* which is responsible for the functioning of 'fear-learning (see Section 6.5) – and extends far into the *cerebra* of both left and right hemispheres. So, in addition to 'connecting the two sides of the brain' it separates the bifurcated cerebral regions from the un-bifurcated brain stem. This is curiously reminiscent of techniques which are used in constructing microelectronic information-processors. Inter-elemental communication in early, small, slow computer chips could successfully take place locally, but in later, larger, faster chips long-range communication has been delegated to purpose-built inter-regional communication structures. The net result is that large adjacent processor regions are no longer coupled locally, but communicate *only* through these dedicated structures. In an extreme form of this 'scaling', separate processor regions (for example the processor, cache memory, or more processors) are implemented as separate chips on a single substrate, and inter-chip communication may even be realized using optical fibers. Does this developmental sequence indicate how the brain has developed, and why the *corpus callosum* has evolved?

Section 5 of this paper described how an expanding system must ultimately change its internal communication strategy to survive and to retain its identity, resulting in the *emergence* of higher scale(s) and higher-level inter-regional (intra-scalar) communication. If, as seems inescapable, a similar logic can be applied to the evolution of the brain, we would conclude that the resulting structure would exhibit a spatial transition from the integrated logical-emotional processing of early neural regions (the singular brain stem) to the quasi-independence and consequent approximately specialized autonomy of the higher-level hemispheres (the *cerebra*). So, operation of the materially high-level neural hemispheres *can* possibly be associated with a birational evolutionary template, but this leaves us with an outstanding question. If the material nature of the brain corresponds to a birational information-processing

system, then where are the low-level, or small-scale complements of its neural networks?

Karl Pribram [105] has indicated that the conventional image of a neural network poses a genuine problem. Neurons are usually pictured with a myriad of dendritic inputs – maybe as many as 50,000 – but with only a single output axon. The axon of a real neuron, however, typically splits into a large number of smaller axonites – maybe also 50,000 – which connect it to secondary neurons. The myriad axonites of closely-packed neurons form a tangled mess between them – usually referred to as the 'axonite mesh'. Pribram has pointed out [105] that at their farther extremities the axonites are very thin, and they do not appear capable of carrying localized electrical signals as far as the secondary-neuron dendrites. He suggests that quasi-waves form in the 'axonite mesh' and that these transmit a superposition of the primary-neuron outputs to the secondary-neuron dendrites in a manner which is analogous to the collapse of a superposition of multiple 'possible' QM states to a single 'real' one. Here, finally, is the low-level complement of the brain's neural networks. As we have described, a natural hierarchy may be decomposed into two constituent parts: one related to Newtonian Mechanical (NM) physics, the other to Quantum Mechanics (QM). In the brain the NM assembly is represented by signal summation in the dendrites; the QM assembly is simulated by quasi-wave superposition in the axonite mesh.

Recent experimentation has provided excellent evidence for the low-level complementarity of living systems. Lincoln and Joyce [18] have announced the discovery of a self-replicating DNA enzyme system which exhibits some of the characteristics of life. They report the development of a variety of RNA enzyme pairs which are capable of catalyzing each other's synthesis – a process they refer to as *cross-replication*. Each enzyme, consisting of two oligonucleotide subunits, binds the other enzyme's two subunits to make a new copy of the enzyme itself, and the complementary process continues indefinitely, given sufficient supply of the subunits. The present authors believe that this is the first self-replicating system which has been artificially developed, and it lends weight to the hypothesis that RNA is a precursor of DNA in the evolution of life. The discovery strongly supports this paper's contention that complementary processes are at the heart of living systems.

6.7 Theory of self, theory of mind, self and the mind

A multicellular organism consists of a collection of cells, but it presents itself to the outside world as a unified entity. As we pointed out in Section 5.2 of this paper, an entity *is* its hyperscale: hyperscale provides *identity*, but in a multiply-autonomous organism there are a number of *different* 'identities' related to its different scales and regional autonomies [67]. Restricting ourselves for the moment to a description of the brain, there are at the very least a material or 'structural' hyperscalar identity, which to some extent persists after death, and a 'mental' hyperscalar identity, which so far as we are aware disappears at the point of death. We are entitled to ask what the connection is,

therefore, between our own 'structural' identity and our internal 'mental' identity – our *sense* of *self*. Survival of an organism implies the survival of *both* of these identities, but asymmetrically the mind's demise is subject to the brain's disease or senescence.

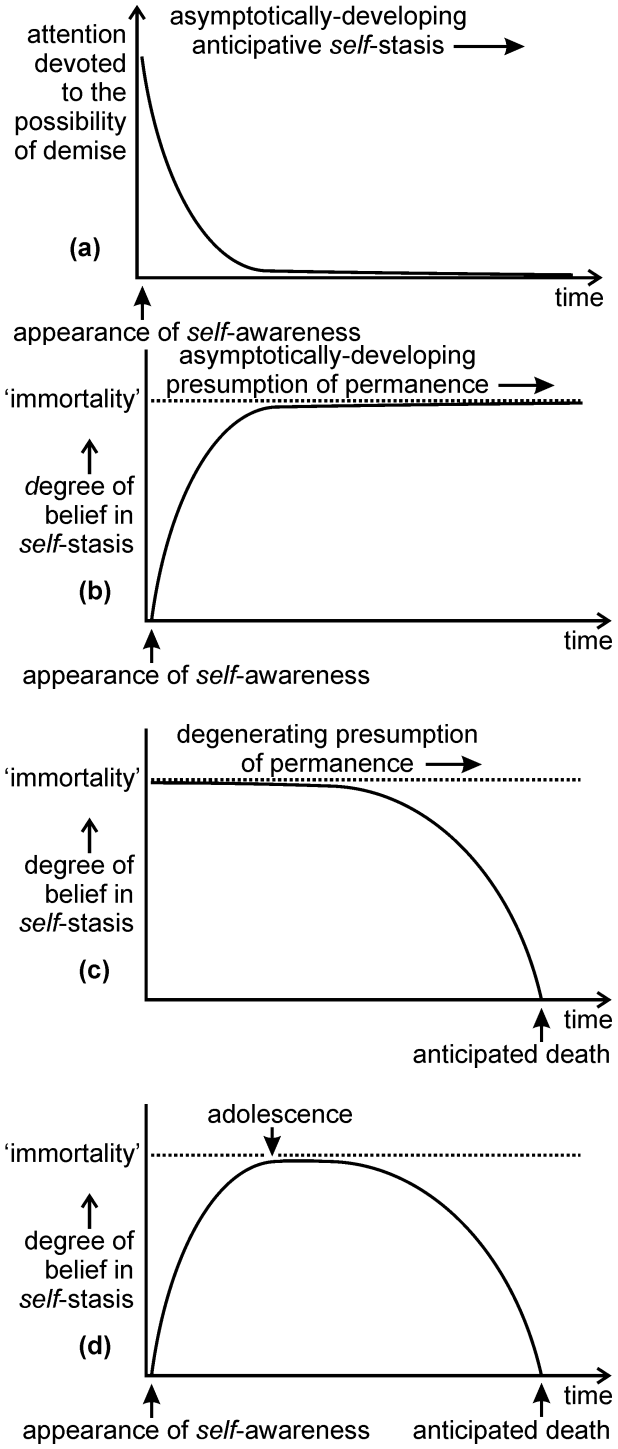

Figure 11: (a) The asymptotic progression of anticipation of *self*-stasis, as an approach to **(b)** The presumption of *self*-permanence, **(c)** Anticipation of the 'brick wall' of death, whose inevitability breaks down the anticipation of im-

mortality, and **(d)** An illustration of the progression from infantile episodic awareness (on the left), through presumed immortality, to acceptance of approaching death (on the right).

If anticipation is to be judged important, then it *must matter* to its executor. In its traditional evocation, evolution is directed by 'the fitness to survive'. But why bother? An *intention* to survive is not enough: an organism must somehow *desire* to survive. But, again, why bother? Our own daily desires are clearly in some way coupled to the desire to survive – either as 'ourselves' or as our offspring – but the primary desire usually remains well concealed behind *hunger*, *thirst*, *satisfaction*, … But what drove *their* survival as our instincts? Ego? A presumption that somehow 'we' are important? Maybe a clue can be found in *stasis-neglect* – or in its absence in specific circumstances. Given effective coupling between short-term and long-term memory we would expect stasis-neglect to be an important factor in judgment of the criticality of particular aspects of our surroundings. However, Thompson and Ogden [10] and Ogden *et al*. [11] have demonstrated that the use of analogy – itself a fundamental component of short-to-long-term coupling and of the transition from episodic to mimetic cognitive processes – is unavailable to neurally-simple animals. In its absence, anticipatory stasis-neglect is unlikely, and such an animal would most probably perpetually feel that it is at the 'sharp end' of Deacon's [2] 'sensory evolution', constantly risking predatory annihilation.

Although anticipation is a valuable tool in guiding our actions, and although its short-term accuracy may be reasonably reliable, the farther we peer into the future the less dependable it becomes. The degree to which we pay attention to deviations from anticipated events or conditions, therefore, reduces with their future distance, and this exacerbates stasis-neglect, and strengthens attention's long-term complementary ecosystem of 'neglected' stasis. If we take the time to reflect on the contents of our attention's ecosystem – on the events and conditions which we systematically ignore – we find that to a great extent their neglect proves to have been justified and, in time, infantile self-questioning evolves into self-confidence and acceptance that the criteria we use in selecting attentive focus are reasonable. Very young children progressively develop the capacity for stasis-neglect, and ultimately acquire a grounding sense of their parents' permanence even when their mother goes out of the room. The implications of stasis-neglect take their most extreme form during adolescence, when the combination of extended freedom and technical independence culminate in an assumption of immortality[54].

[54] … which is most noticeable in the way that young male adults drive cars and motorcycles, and particularly evident in the descriptions given by young soldiers of life in a war zone, where even after friends have been killed by enemy action they often still believe that 'it will never happen to me'.

Figure 11(a) illustrates the normal asymptotic *form* of stasis-neglect – similar to that shown earlier in Figure 5 – but now the 'stasis' we will address is the ecosystemically grounded presumption of permanence: of *self*-stasis. If Figure 11(a) were to comprise the complete story, there would indeed be no reason to bother about survival, but the entirety of our social comprehension informs us not only that all individuals die, but that this systematically occurs at an age of 'three score years and ten'[55]. As we humans move on through adolescence, to maturity, towards the 'brick wall' of death, its inevitability collapses the preceding anticipation of immortality (Figure 11(b)), replacing it by the 'sharp end' of Deacon's [2] 'sensory evolution' and the constant awareness of approaching annihilation (Figure 11(c)).

An important aspect of both *Anticipative Capability* and *IS&W* is the manner in which their application is moderated by our sense of *self* and our relationships with others. But where and what is the 'self'? Metzinger [106] has presented the hypothesis that we are unable to distinguish between the objects of our attention and the internal representations of them which we 'observe'. Consequently, when our attention targets a tool, an object or a situation we effectively transfer ourselves – our 'presence' – to it: when we drive a car, we become the car; when we watch a film, we enter into its action. The most amazing aspect of this transfer of presence is that we can effortlessly skip between different scales of an overall picture.

Metzinger's hypothesis provides a credible model for the independence of *the mind*. As he states:

"*We are systems that are not able to recognize their subsymbolic self-model as a model. For this reason we are permanently operating under the conditions of a 'naïve-realistic misunderstanding': we experience ourselves as being in direct and immediate epistemic contact with ourselves. What we have in the past simply called 'self' is not a non-physical individual, but only the content of an ongoing, dynamical process – the process of transparent self-modeling*" [106, p. 54].

Metzinger does not, however, provide us with any clue as to 'where' we can 'find' the 'self-model', or how it has been generated over the aeons of evolution. He concludes that

"*… the conscious self is an illusion which is no one's illusion*" [106, p. 60].

Although we concur with Metzinger that *the self* is *objectively* illusory, self-consistency would suggest that it is *its own* illusion. First-person awareness depends critically on introspection – that is, on the capacity for *the self* to observe *itself*. The analysis of concepts of 'identity' within a monorational system results in a never-ending inconclusive sequence of different conclusions,

[55] To avoid the necessity of discussing the statistics of death and their modification with the centuries, we have here adopted the figure provided by The (Christian) Bible!

whose record recounts the history of philosophy and of concepts of 'existence' through the ages. Reference to Section 5.4 and Footnote 34 of this paper will indicate that this is unsurprising, and that analytic inconclusivity disappears in a *birational* context. Matsuno's [99] self-consistent view of 'reality' is based on the interpretation of observation as a *mutual measurement*. The recursive inter-correlation of the two sapiences of a birational hierarchy described in Section 6.5 provides an excellent example of 'mutual *self*-measurement' – *from logic, to emotion, to logic, to emotion, to logic…* – and of an evolutionary self-observation which is strongly reminiscent of "*what we should expect an evolutionlike process to feel like*" [2, p.458]. If, as we suggested in Section 6.5, the 'phenomenon' of *self* 'resides' in truncation by stasis neglect of this infinite self-observational sequence, then 'who' in a birational information-processing brain collapses the 'introspective' sequence down to the apparent, if illusory, stability of *the self*? Well – *no one* does! Stasis-neglect suffices. The temporal stability of a natural hierarchy is maintained by the cross-scalar transmission of order, which dominates the transmission of novelty (as it similarly does in a crystal [50, 100]). Consequently, structural stability implies stasis and the initiation of asymptotic anticipative neglect, which effectively truncates the infinitely recursive self-observation and stabilizes an introspective 'Theory of *Self*', or 'Theory of the *Reality* of *Self*'.

Although *the self* is just that – a phenomenological characteristic or property of a *specific* differentiated entity – Ramachandran [87] has suggested that neural mirroring "*dissolve*(s) *the barrier between self and others*." As such it constitutes the birational paradigm *itself*: a system of two mutually-evolvable interrelating aware localizations, whose 'functions', as entity or ecosystem, are contextually interchangeable. *Neural mirroring* consequently provides a useful pictorial vehicle for comparing different correspondences, whether these are entity-to-environment, interpersonal or neurologically internal. An *unknown external environment* can be progressively described internally by an organism through assembly of the stimuli to which it is subjected. Similarly, an *unknown organism* can be progressively described through assembly of the questions it poses of its environment and through its social relations. Both of these processes can be related to *neural mirroring*, as a high-level implementation of Matsuno's [99] 'observation as a mutual measurement'. Ecosystemic containment of a natural hierarchy becomes internalized through the generation and maintenance of its extant scalar levels, creating hyperscalar self-constraint as an indistinguishable reproduction of relevant parts of its ecosystem. Consequently, the creation of an internal transparent *environmental* model is automatically and intimately associated with the creation of an internal transparent *self*-model!

We believe that unification-maintaining hyperscalar survivalist sapient behavior has resulted in long-term evolution of the high-level transparent self-model Metzinger [106] refers to. We suggest that the 'spotlight of consciousness'[56] in humans is focused at any given moment on a single 'location' within

[56] An expression modified from one due to Bernie Baars.

a spatiotemporal hyperscalar 'phase space' which we have individually constructed from the entirety of our individual and social histories, including genetic and epigenetic influences, 'well-known facts' of our socially or individually believed 'reality' (or 'realities'), apparently self- or generally-consistent but personally insufficiently-investigated 'obvious' or 'logical' positions, and otherwise socially- or scientifically-abandoned hypotheses which we employ either consciously or unconsciously (or both!) to fill in inconvenient or excruciatingly obvious omissions from its landscape.

There appears to be a direct equivalence between '*Theory of Self*' (belief in one's own independent reality), '*presence transfer*' (the ability to functionally 'be at' or 'become' a target object or person) and '*Theory of Mind*' (the ability to 'understand' that others have beliefs, desires and intentions that are different from one's own). All of these three are generated from the bi-sapience of neural information-processing, through *neural mirroring* or its equivalent. '*Theory of Self*' is associated with true sapient introspection: it is the result of long-term stabilization of the observations of an organism's internal transparent model by that internal model itself (c.f. "*I think, therefore I am*"). '*Presence transfer*' is effected by viewing the internal model of an external targeted organism, artifact or situation, or the internal model of a fantasized organism, artifact or situation, *through* a transparent internal model of *the self*[57], with the result that *the self* is not distinguished from the target. '*Theory of Mind*' is related to empathy, in that it addresses the indirect mirroring of logic and emotion between *the self* and *an other*, which results in the conclusion of *self-to-other* similarity on the basis of long experience of socially-coupled inter-personal quasi-introspection. Metzinger and Gallese [86] have published the elements of a theory which is aimed at attributing a common *action ontology* to the evolution of theories of Self and Mind through attention to *neural mirroring* in the motor system. As they state:

"*An elementary self-model in terms of body image and visceral feelings plus the existence of a low-level attentional mechanism is quite enough to establish the basic representation of a dynamic subject-object relation. The non-cognitive PMIR (Phenomenal Model of the Intentionality Relation) is thus what builds the bridge into the social dimension.*" [86].

[57] It is clear that whatever logically directed words are chosen to describe the complexity of this situation, they will never be correct! The concept of 'transparency' is itself derived from a linear representation of communication, and is consequently technically at odds with any idea of self-illusion! The authors have chosen to use the word 'through' here instead of 'from' to emphasize the illusory nature of the process's origin.

7 Conclusion: the Mind as an Evolving Anticipative Capability

The argumentation provided by this paper supports the contention that *the mind* is 'nothing other than' an evolving *Anticipative Capability*. As such, although it is tempting to follow Descartes' notions and accept that *the mind* and *the body* are categorically dissimilar, they are *both* pragmatically grounded in the 'mechanics' of scalar fragmentation and its evolutionary reunification. Critically, we have indicated that recognition of an *evolving directivity of evolution itself* provides the link between the apparent randomness of primitive life and human 'free will' as exercised 'by' *the mind*.

Spencer's [21] prescription of 'survival of the fittest', and the many ways in which it has been reformulated as variations of 'survival of the merely adequate', only really takes account of 'static' aspects of organisms or species, in a manner reminiscent of Rosen's view of 'traditional' biology [15]. Although the success of a species' members in a Survival Competition may be readily indicated by a snapshot of its relative population, this says comparatively little about individual capabilities – nor does genetics about phenotypic development in an epigenetic context. Conventional genetically-based models of life do little to unify our schismic preconceptions of societal-individual relationships, and any enhanced reformulation of our individual place in society, of our rights, of our responsibilities, will necessarily depend on the same recourse to scalar properties and *Anticipative Capability* which we have adopted here in our reflections on *the mind*. The central message of scale is that neither perfect isolation nor perfect communion can lead us to peaceful coexistence, and that the ecosystemic message we receive from the living world *can* successfully replace the conventional monorational viewpoint of science and logic.

The achievement of reasonably precise and accurate anticipation in complex multiscalar environments demands internal model-structures which reflect this complex multiscalar character. Although a degree of pre-programming is available through species genetics, the majority of internal-model programming relevant to anticipatory activity is developed through 'live' experience. Many infant animals are capable of lone survival soon after birth, but such is not systematically the case for mammals, and is especially not the case for infant *Homo sapiens*. We have detailed the way in which the 'hard-wired' DNA pool of an offspring's instinctive capabilities can be modified or broadened by the transfer of more abstract *non-DNA-coded* capabilities during the period within which parents and their descendents co-exist. Early species-survival may well have principally depended on random genetic mutation, but this is clearly not the case for later evolved, more complex organisms such as *Homo sapiens*, for whom anticipation delivers wealth, health and some degree of happiness. Traditional 'simplistic' viewpoints maintain that *the mind* and *its activities* can be easily distinguished. In common with Deacon [2], Metzinger [106] and countless others, we do not believe this to be the case, and advocate

that *the mind* is a natural feature of *embodiment*, both of whose origins may be found in the *co-evolutionary* development of physical differentiation and unification.

The *evolution of evolution itself* appears to have progressively followed a path from random mutation towards anticipatory development. The traditional objection to a description of evolution as 'survival of the fittest' is that it only provides a circular definition:

"the fittest *for what*? – the fittest *for survival*!"

The argumentation of this paper supports a modified description which eliminates this circularity, namely:

"the fittest *for what*? – the fittest *for anticipation*!"

However, as internal anticipatory structures are phenotypic and not genotypic properties, this indicates yet another evolution – from *species* dependence towards *individual* dependence through evolution of *the mind*. Our societies are built on complicated and complex interlinked sets of rules, which protect the society through modification of individuals' and groups' actions, and protect the individual through containment of society's authority[58].

To a large extent these rule-sets now remove the necessity for individuals to exercise extreme anticipation in guaranteeing their daily survival, and enable us to direct our mental capabilities to more abstract objectives. It is far easier and safer to drive around in an automobile if we know that all other automobiles will be driving on the left hand side of the road (or, if locally suitable, on the incorrect right hand side!). The net result is to reduce the level of conscious anticipatory effort which is required to promote survival. Through eventual evolution of *the anticipatory mind* from the low-level directivity of primitive organisms our societies have become characterized,

not statically by

'*survival of the adequately fit*'

but dynamically by

'survival of the adequately anticipative'.

References

1. Descartes, R. *Meditations on First Philosophy VI*, http://filepedia.org/meditations-on-first-philosophy/ (translated by J. Veitch), accessed 08/01/2009.

[58] … as do the 10 commandments which figure in The (Christian) Bible.

2. Deacon, T. W. *The Symbolic Species: The Co-Evolution of Language and the Brain*, W.W. Norton & Co.: New York (1997).
3. Diskeeper Corporation, http://www.diskeeper.com, accessed 08/01/2009.
4. Bateson, G. *Steps to an Ecology of Mind*, Ballantine, New York (1972).
5. Turing, A. "Computing machinery and intelligence," *Mind* **59**, 433-460 (1950).
6. Aristotle, *Physics*, http://etext.library.adelaide.edu.au/a/aristotle/physics/ (translated by R. P. Hardie and R. K. Gaye), accessed 08/01/2009.
7. Plato, *Parmenides*, http://classics.mit.edu/Plato/parmenides.html (translated by B. Jowett), accessed 08/01/2009.
8. Cottam, R., Ranson, W. and Vounckx, R. "Life and Simple Systems," *Systems Research and Behavioral Science* **22**, 413-430 (2005).
9. Hoffmeyer, J. and Emmeche, C. "Code-duality and the semiotics of nature." In: M. Anderson and F. Merrell (Eds.), *On Semiotic Modeling*. Mouton de Gruyter, New York, pp. 117-166 (1991).
10. Thompson, R. K. R. and Ogden, D. L. "Why monkeys and pigeons, unlike certain apes, cannot reason analogically." In: K. Holyoak, D. Gentner and B. Koichov (Eds.), *Advances in Analogy Research: Integration of Theory and Data from the Cognitive, Computational, and Neural Sciences*. NBU: Sofia, Bulgaria, pp. 269-273 (1998).
11. Ogden, D. L., Thompson, R. K. R. and Premack, D. "Analogical problem-solving by chimpanzees." In: K. Holyoak, D. Gentner and B. Koichov (Eds.), *Advances in Analogy Research: Integration of Theory and Data from the Cognitive, Computational, and Neural Sciences*. NBU: Sofia, Bulgaria, pp. 38-48 (1998).
12. Weir, A. A. S., Chappel, J. and Kacenik, A. "Shaping of hooks in New Caledonian crows," *Science* **297**, 981 (2002).
13. Carson, R. *Silent Spring*, 40th anniversary edition, Houghton Mifflin Company, New York (2002).http://www.library.uq.edu.au/training/citation/harvard.html
14. "Roll back driver" – Microsoft Hardware Device Manager, Windows XP.
15. Rosen, R. *Life Itself: a comprehensive enquiry into the nature, origin and fabrication of life*. Columbia UP, New York (1991)
16. Cottam, R., Ranson, W. and Vounckx, R. "Emergence: half a quantum jump?" *Acta Polytechnica Scandinavica* **91**, 12-19 (1998).
17. Darwin, C. *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. John Murray, London (1859).
18. Lincoln, T. A. and Joyce, G. F. "Self-sustained replication of an RNA enzyme," *ScienceXpress* **8**, DOI: 10.1126/science.1167856, January 8, 2009.
19. Mendel, G. "Versuche über pflanzen-hybriden," *Verhandlungen des naturforschenden Vereines, Abhandlungen, Brünn* **4**, 3-47 (1866).
20. Lamarck, J-B. (1814). *Zoological philosophy. An exposition with regard to the natural history of animals*. Translated from the original (1809) French edition by Hugh Elliot. Macmillan, London (1814).
21. Spencer, H. *Essays: scientific, political and speculative. Library edition, containing seven essays not before republished, and various other additions, Vol. 1*. Williams and Norgate, London (1891).

22. Pietikainen, S.
http://www.usenet.com/newsgroups/talk.origins/msg08126.html, accessed 08/01/2009.

23. Norman, R.
http://www.usenet.com/newsgroups/talk.origins/msg08242.html, accessed 08/01/2009.

24. Gravely, B. R. "Alternative splicing: increasing diversity in the proteomic world," *Trends in Genetics* **17**, 100-107 (2001).

25. The Human Genome Project.
http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml, accessed 08/01/2009.

26. Mattick, J. S. Quoted by W. W. Gibbs in "The unseen genome: gems among the junk," *Scientific American* **289**, 5, 46-53 (2003).

27. Mattick, J. S. "Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms," *BioEssays* **25**, 930-939 (2003).

28. Hirotsune, S., Yoshida, N., Chen, A., Garrett, L., Sugiyama, F., Takahashi, S., Yagami, K., Wynshaw-Boris, A. and Yoshiki, A. "An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene," *Nature* **423**, 91-96 (2003).

29. Nishida, H., Tomaru, Y., Oho, Y. and Hayashizaki, Y. "Naturally occurring antisense RNA of histone H2a in mouse cultured cell lines," *BMC Genetics* **6**, 1-7 (2005).

30. Lau, N. C. and Bartel, G. P. "Censors of the genome," *Scientific American* **289**, 34-41 (2003).

31. Vitreschak, A. G., Rodionov, D. A., Mironov, A. A. and Gelfand, M. S. "Riboswitches: the oldest mechanism for the regulation of gene expression?" *Trends in Genetics* **20**, 44-50 (2004).

32. Krichevsky, A. M., King, K. S., Donahue, C. P., Khrapko, K. and Kosik, K. S. "A microRNA array reveals extensive regulation of microRNAs during brain development," *RNA* **9**, 1274-1281 (2003).

33. Loots, G. G., Locksley, R. M., Blankespoor, C. M., Wang, Z. E., Miller, W., Rubin, E. M. and Frazer, K. A. "Identification of a coordinate regulator of interleukins 4, 13 and 5 by cross-species sequence comparison," *Science* **288**, 136-140 (2000).

34. Prabhakar, S., Noonan, J. P., Pääbo, S. and Rubin, E. M. "Accelerated evolution of conserved noncoding sequences in humans," *Science* **314**, 768 (2006).

35. Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thåström, A. C. Field, Y., Moore, I. K., Wang, J-P. Z. and Widom, J. "A genomic code for nucleosome positioning," *Nature* **442**, 772-778 (2006).

36. Segal, E. Quoted by N. Wade in "Scientists say they've found a code beyond genetics in DNA," *NY Times*, July 25 (2006).

37. Bateson, W., Saunders, E. R. and Punnett, R. C. "Further experiments on inheritance in sweet peas and stocks: preliminary account," *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character* **77**, 236-238 (1906).

38. Baldwin, J. M. "A new factor in evolution," *American Naturalist* **30**, 441-451 (1896).

39. Morgan, C. L. "On modification and variation," *Science* **4**, 733-740 (1896).

40. Osborn, H. F. "Ontogenic and phylogenic variation," *Science* **4**, 786-789 (1896).

41. Deacon, T. W. "Multilevel selection in a complex adaptive system: the problem of language origins." In: B. H. Weber and D. J. Depew (Eds.), *Evolution and Learning: the Baldwin Effect Reconsidered*, MIT Press, Cambridge, MA (2003).

42. Wiles, J., Watson, J., Tonkes, B. and Deacon, T. W. "Evolving complex integrated behavior by masking and unmasking selection pressures." Presented at the 4th International Conference on Complex Systems, Nashua, NH, 9-14 June (2002).

43. Waddington, C. H. "The epigenotype," *Endeavour* **1**, 18-20 (1942).

44. Waterland, R. A. and Jirtle, R. L. "Transposable elements: targets for early nutritional effects on epigenetic gene regulation," *Molecular and Cellular Biology* **23**, 5293-5300 (2003).

45. Loehle, C. "Social barriers to pathogen transmission in wild animal populations," *Ecology* **76**, 326-335 (1995).

46. Bergman, J. "Darwinism and the Nazi race Holocaust," *Technical Journal* **13**, 101-111 (1999).

47. Darwin, C. *The Descent of Man and Selection in Relation to Sex*, John Murray, London (1871).

48. Paul, A. "Sexual selection and mate choice," *International Journal of Primatology* **23**, 877-904 (2002).

49. Cottam, R., Ranson, W. and Vounckx, R. "Abstract or die: life, artificial life and (v)organisms." In: E.R. Messina and A.M. Meystel (Eds.), *Performance Metrics for Intelligent Systems: Proceedings of PerMIS '03 Workshop*, NIST Special Publication 1014, paper #WeAM1-4. NIST: Gaithersburg, MD, pp. 1-7 (2003).

50. Cottam, R., Ranson, W. and Vounckx, R. "Autocreative hierarchy I: structure - ecosystemic dependence and autonomy," *SEED Journal* **4**, 24-41 (2004).

51. Cottam, R., Ranson, W. and Vounckx, R. "Autocreative Hierarchy II: Dynamics - Self-Organization, Emergence and Level-Changing." In: H. Hexmoor (Ed.), *International Conference on Integration of Knowledge Intensive Multi-Agent Systems*, IEEE: Piscataway, NJ, pp. 766-773 (2003).

52. Antoniou, I. "Extension of the conventional quantum theory and logic for large systems." Presented at the International Conference 'Einstein Meets Magritte', Brussels, Belgium, 29 May – 3 June (1995).

53. Lohman, R. "Structure evolution and incomplete induction." In: R. Manner and B. Manderick (Eds.), *Proceedings of the 2nd Conference on Parallel Problem Solving from Nature*, Elsevier, Amsterdam, The Netherlands, pp. 175-185 (1992).

54. Dubois, D.M. "Review of incursive, hyperincursive and anticipatory systems – foundation of anticipation in electromagnetism." In: D.M. Dubois (Ed.), *Computing Anticipatory Systems: CASYS'98 - 2nd International Conference, AIP Conference Proceedings 465*, American Institute of Physics, Woodbury, NY, pp. 3-30 (1999).

55. Albrecht-Buehler G. "Surface extensions of 3T3 cells towards distant infrared sources," *Journal of Cell Biology* **114**, 493-502 (1991).

56. The Quorum Sensing Site, http://www.nottingham.ac.uk/quorum/, accessed 08/01/2009.

57. Cottam, R.; Ranson, W.; and Vounckx, R. "Consciousness: the precursor to life?" In: C. Wilke, S. Altmeyer and T. Martinetz (Eds.), *Third German Workshop on Artificial Life: Abstracting and Synthesizing the Principles of Living Systems*, Verlag Harri Deutsch, Thun, Germany, pp. 239-248 (1998).

58. Weber, R. "Meaning as being in the implicate order philosophy of David Bohm: a conversation." In: B. J. Hiley and F. D. Peat (Eds.), *Quantum Implications: Essays in Honor of David Bohm*, Routledge and Kegan Paul, London, pp. 440-441 (1987).

59. Schneier, B. *Applied Cryptography: Protocols, Algorithms and Source Code in C*, John Wiley and Sons, New York (1996).

60. Cottam, R., Ranson, W. and Vounckx, R. "Living in hyperscale: internalization as a search for unification." In: J. Wilby, J.K. Allen and C. Loureiro-Koechlin (Eds.), *Proceedings of the 50th Annual Meeting of the International Society for the Systems Sciences*, paper #2006-362, ISSS, Asilomar, CA, pp. 1-22 (2006).

61. Cottam, R., Ranson, W. and Vounckx, R. "Hyperscale puts the sapiens into homo," *New Mathematics and Natural Computation*, in publication (2009).

62. Bennett, C. H., Brassard, G., Crepeau, C., Jozsa, R., Peres, A. and Wootters, W. "Teleporting an unknown quantum state via dual classical and EPR channels," *Physics Review Letters* **70**, 1895-1899 (1993).

63. Feynman R. P. and Hibbs A. R. *Quantum Mechanics and Path Integrals*. McGraw-Hill, New York (1955).

64. Cottam, R., Ranson, W. and Vounckx, R. "Localization and nonlocality in computation." In M. Holcombe and R. Paton (Eds.), *Information Processing in Cells and Tissues*, Plenum Press, London, pp. 197-202 (1998).

65. Cottam, R., Ranson, W. and Vounckx, R. "Diffuse rationality in complex systems" In: Y. Bar-Yam and A. A. Minai (Eds.), *Unifying Themes in Complex Systems, vol. II*. Westview Press, Boulder, CO, pp. 355-362 (2004).

66. Cottam, R., Ranson, W. and Vounckx, R. "Artificial minds?" In: J. Wilby and J.K. Allen (Eds.), *Proceedings of the 45th Annual Meeting of the International Society for the Systems Sciences*, paper #01-114, ISSS, Asilomar, CA, pp. 1-19 (2001).

67. Collier, J. D. "Autonomy in anticipatory systems: significance for functionality, intentionality and meaning." In: D. M. Dubois (Ed.), *Computing Anticipatory Systems: CASYS'98 - 2nd International Conference, AIP Conference Proceedings 465*, American Institute of Physics, Woodbury, New York, pp. 75-81 (1999).

68. Rosen, R. *Essays on Life Itself*. Columbia UP, New York (1998).

69. Gwinn, T. http://www.panmere.com/?cat=3, accessed 08/01/2009.

70. Mikulecky, D. http://www.people.vcu.edu/~mikuleck/, accessed 08/01/2009.

71. Cottam, R., Ranson, W. and Vounckx, R. "Sapient structures for intelligent control." In: R. V. Mayorga and L. I. Perlovsky (Eds.), *Toward Artificial Sapience: Principles and Methods for Wise Systems*, pp. 175-200. Springer, New York (2008).

72. Albus, J. S. "Features of intelligence required by unmanned ground vehicles," available on the NIST publications list as http://www.isd.cme.nist.gov/documents/albus/Features_of_Intelligence.pdf, accessed 08/01/2009.

73. Cottam, R., Ranson, W. and Vounckx, R. "Back to the future: anatomy of a system." In: D. M. Dubois (Ed.), *Computing Anticipatory Systems: CASYS'03 – 3rd International Conference, AIP Conference Proceedings 718*, American Institute of Physics, Woodbury, New York, pp. 160-165 (2004).

74. Cottam, R., Ranson, W. and Vounckx, R. "Cross-scale, richness, cross-assembly, logic 1, logic 2, pianos and builders." Presented at the 2nd International SEE Conference: The Integration of Information Processing, Toronto, Canada, 6-8 October (2001).

75. Gunderson, J. P. and Gunderson, L. F. "Intelligence ≠ autonomy ≠ capability." In: E.R. Messina and A.M. Meystel (Eds.), *Performance Metrics for Intelligent Systems: Proceedings of PerMIS '04 Workshop*, NIST Special Publication 1036, paper #ThAM2-4. NIST: Gaithersburg, MD, pp. 1-7 (2004).

76. Cottam, R., Ranson, W. and Vounckx, R. "A biologically consistent hierarchical framework for self-referencing survivalist computation." In: D. M. Dubois (Ed.), *Computing Anticipatory Systems: CASYS'99 – 3rd International Conference, AIP Conference Proceedings 465*, American Institute of Physics, Woodbury, New York, pp. 252-262 (2000).

77. Zeno, http://philsci-archive.pitt.edu/archive/00001197/02/Zeno_s_Paradoxes_-_A_Timely_Solution.pdf, accessed 08/01/2009.

78. Peirce, C.S. *Collected Papers of Charles Sanders Peirce, 1931-1935*, vols. 1–6, C. Hartshorne and P. Weiss (Eds.), vols. 7–8, A. W. Burks (Ed.), Harvard University Press, Cambridge, MA, 1958.

79. Descartes, R. *Discours de la Méthode*, Editions 10/18: Paris (2002).

80. Fadiga, L., Fogassi, L., Pavesi, G. and Rizzolati, G. (1995) "Motor facilitation during action observation: a magnetic stimulation study," *Journal of Neurophysiology* **73**, 2608-2611 (1995).

81. Kohler, E., Keysers, C., Umiltà, M. A., Fogassi, L., Gallese, V. and Rizzolati, G. "Hearing sounds, understanding actions: action representation in mirror neurons," *Science* **297**, 846-848 (2002).

82. Oberman, L. M., Pineda, J. A. and Ramachandran, V. S. "The human mirror neuron system: a link between action observation and social skills". *Social Cognitive and Affective Neuroscience* **2**, 62-66 (2007).

83. Williams, J. H. G., Whiten, A., Suddendorf, T. and Perrett, D. I. "Imitation, mirror neurons and autism," *Neuroscience and Biobehavioral Reviews* **25**, 287-295 (2001).

84. Molnar-Szakacs, I. and Overy, K. "Music and mirror neurons: from motion to e'motion," *Social Cognitive and Affective Neuroscience* **1**, 235-241 (2006).

85. Gallese, V. and Lakoff, G. "The brain's concepts: the role of the sensory-motor system in conceptual knowledge," *Cognitive Neuropsychology* **22**, 455-479 (2005).

86. Metzinger, T. and Gallese, V. "The emergence of a shared action ontology: building blocks for a theory," *Consciousness and Cognition* **12**, 549-571 (2003).

87. Ramachandran, V. S. "Mirror neurons and the brain in a vat." http://www.edge.org/3rd_culture/ramachandran06/ramachandran06_index.html, accessed 08/01/2009.

88. Rock, A. *The Mind at Night: the New Science of How and Why we Dream*, Basic Books, New York (2004).

89. Antrobus, J.S. and Bertini, M. *The Neuropsychology of Sleep and Dreaming*, Lawrence Erlbaum Associates, Hillsdale, NJ (1992).

90. Haken, H. *The Science of Structure: Synergetics*, Prentice Hall, New York (1984).

91. Cottam, R., Ranson, W. and Vounckx, R. "A diffuse biosemiotic model for cell-to-tissue computational closure," *BioSystems* **55**, 159-171 (2000).

92. Louie, A. H. "A series of unfortunate misprints." http://www.panmere.com/rosen/Louie_LI_typos.pdf, accessed 08/01/2009.

93. Cottam, R., Ranson, W. and Vounckx, R. "Replicating Robert Rosen's (M,R) systems." In: J. Wilby, J.K. Allen and C. Loureiro-Koechlin (Eds.), *Proceedings of*

the 50th Annual Meeting of the International Society for the Systems Sciences, paper #2006-378, ISSS, Asilomar, CA, pp. 1-10 (2006).

94. Cottam, R., Ranson, W. and Vounckx, R. "Re-mapping Robert Rosen's (M,R) systems," *Chemistry and Biodiversity* **4**, 2352-2368 (2007).
95. Edmonds, B. "Implementing Free Will." In D. N. Davis (Ed.), *Visions of Mind - Architectures for Cognition and Effect*, Information Science Publishing, London, pp. 108-124 (2005).
96. LeDoux, J. E. "Brain mechanisms of emotion and emotional learning," *Current Opinion in Neurobiology* **2**, 191-197 (1992).
97. Ward, A. http://www.feralchildren.com/en/autism.php, accessed 08/01/2009.
98. Stern, D. N. *The Interpersonal World of the Child: A View from Psychoanalysis and Developmental Psychology*, Basic Books, New York (1985).
99. Matsuno, K. "The internalist stance: a linguistic practice enclosing dynamics," *Annals of the New York Academy of Sciences* **901**, 322-349 (2000).
100. Cottam, R. and Saunders, G.A. "The elastic constants of GaAs from 2K to 320K," *Journal of Physics C: Solid State Physics* **6**, 2015-2118 (1973).
101. Pearce, J. M. S. "Corpus callosum," *European Neurology* **57**, 249-250 (2007).
102. Akelaitis, A. J. "Psychobiological studies following section of the corpus callosum: a preliminary report," *American Journal of Psychiatry* **97**, 1147-1157 (1941).
103. Akelaitis, A. J., Risteen, W. A., Herren, R. Y. and van Wagenen, W. P. "Studies on the corpus callosum. III. A contribution to the study of dyspraxia in epileptics following partial and complete section of the corpus callosum," *Archives of Neurology: Psychiatry* **47**, 971-1008 (1942).
104. Sperry, R. W., Gazzaniga, M. S. and Bogen, J. E. "Interhemispheric relationships: the neocortical commisures: syndromes of hemisphere disconnection." In: P. J. Vinken and G. W. Bruyn (Eds.), *Handbook of Clinical Neurology, vol 4*, North-Holland, Amsterdam, pp. 273-290 (1969).
105. Pribram, K. H. "Proposal for a quantum physical basis for selective learning." Presented at the 4th International Conference on Emergence, Complexity, Hierarchy and Order, Odense, Denmark, 31 July – 4 August (2001).
106. Metzinger, T. "The subjectivity of subjective experience: a representationalist analysis of the first-person perspective," Networks 3-4, 33-64 (2004).

The Challenges for Implementable Theories of Mind

Pentti O. A. Haikonen

Department of Philosophy, University of Illinois at Springfield

Abstract
Implementable theories of mind would be of great value to the designers of artificial minds. Existing philosophical theories of mind tend to be loose and metaphorical and therefore do not provide very much guidance to a mind engineer. Unfortunately a complete implementable theory of mind does not yet exist even though there are several attempts toward that direction. The development of an implementable theory of mind faces several major challenges. Among these are the mind-body problem, the identification of the processes of mind, the problem of meaning and understanding, emotions, qualia and consciousness. These issues have been addressed via high-level algorithmic approach and low-level system approach and the combination of these, but each approach has proven to have its own challenges.

1 Introduction

Cognitive robots need brains and minds. Human brain has some 10^{14} synapses that are supposed to store memorized information. If one synapse were to store one bit then the brain's maximum memory capacity would be around 100 Terabits. On the other hand 32 Gigabyte (=2,56*10^{11} bits) miniature memory cards are now available and Terabyte memory cards are just around the corner. Biological synapses are not digital memory locations and their architectural organization is different from random access memories, but nevertheless the lesson is that semiconductor industry is now beginning to be able to produce devices with the circuit element density and complexity comparable to those of the brain. The brain is the site of the mind; does the aforesaid lead to the conclusion that artificial minds are just around the corner, too? The answer is a definite yes, provided that we are able to locate the correct corner. The correct corner is, of course, the implementable theory of mind. This, unfortunately, is not yet available in a concise, complete and tried engineering form even though several attempts towards this already exist. (e.g. Anderson et al 2004, Duch 2005, Haikonen 2003, 2007). Somehow it seems easier to explain the workings of the brain than to devise an engineering theory of mind that would allow the creation of a thinking machine. The brain operates independent of the correctness of the explanation, but a thinking machine will not work if the theory is not right.

What is a mind? What should a mind do? What kind of an information processing system can be called a mind? Should a mind be aware of itself, be self-

conscious? What does it mean when something has a mind of its own? A theorist must look into these questions while looking for an implementable theory of mind. These are also issues that the philosophers of mind have treated over centuries. During these musings philosophers have stumbled on the mind-body problem; the apparent immateriality of mind and consequently, the apparent impossibility of interaction between the immaterial mind and the material body. It follows from the definition of material and immaterial substances that this problem is unsolvable, therefore implementable theories of mind cannot be dualistic ones in the sense of Descartes.

Whose mind is it? An artifact may behave as if it had a mind of its own, yet it may only be executing a collection of preprogrammed commands. In this case the artifact's operation reflects only the mind of the designer, not any of its own. Clearly no real mind has been designed or created. –A well is constructed for the water. However, a successful well digger does not supply the water, he only excavates a suitable hole for the water to seep in. In an analog way, a successful designer of mind should only design machinery that supports the mind and let the contents and caprices of the mind accumulate in the course of operation. In the following the constitutive aspects of an implementable theory of mind are examined.

2 What kind of a theory?

What kind of a theory would an implementable theory of mind be? Philosophical theories of mind tend to be abstract and metaphorical and consequently they are not very helpful for designers of artificial minds. Engineers are able to design systems as soon as the specifications for the system to be designed are given. A metaphor is not a proper specification, an algorithmic description of a desired function is. Thus, at first sight, it would seem that an implementable theory of mind should be algorithmic.

An algorithm is a sequence of instructions, which will lead to the desired outcome when executed properly. In a computer the instructions refer to the set of available operations such as the memory storage or recall, arithmetic operation, shifting a bit string, etc. A sequence of instructions that does not lead to a definite outcome should not be considered as an algorithm. Sometimes algorithms are seen as deterministic processes. However, this is not always the case as an algorithm may involve probabilistic and random operations, e.g. the utilization of randomly generated numbers.

The human mind appears to be non-deterministic; the mind is supposed to have "free will". Consequently the inaccurate idea that algorithms are necessarily deterministic may lead to the conclusion that the human mind must be non-algorithmic. For instance, Penrose has proposed that the mind would rely on non-algorithmic quantum mechanic processes (Penrose 1989). However, the operating temperature of the brain does not readily support quantum computing and the apparent freedom of will must have another explanation.

The operation of any system that obeys natural laws can, in principle, be simulated by algorithms; the accuracy of the simulation is another issue. The brain is such a system and the basic operation of individual neurons and syn-

apses can be simulated with some accuracy. However, the real time computer simulation of a neural system with the complexity of the human brain and some 10^{14} synapses remains a really hard challenge.

High-level symbolic theories of mind are algorithmic and computational. These theories describe syntactic interactions between abstract entities, symbols, and in this way avoid the need to model and compute the operation of low-level units such as neurons and synapses. An early example of this approach is the computational theory of mind (CTM), proposed by Putnam (1961) and further developed by Fodor (1975). Newell and Simon (1975) had a similar idea. According to their Physical Symbol System Hypothesis a physical symbol system has the necessary and sufficient means for general human level intelligent action. Newell and Simon believed to have empirical evidence for this even though they admitted that the main evidence would be the absence of competing hypotheses, i.e. their proof was a proof by ignorance. Putnam and Fodor had a similar line of argument; they argued that mind is necessarily computational because symbolic computation is (as they claimed) the only known method to achieve results that otherwise can only be achieved via thinking; "it is the only game in town". However, so far nothing close to an artificial mind has materialized from these theories.

High-level symbolic theories provide algorithms that describe how further symbols are to be determined on the basis of given symbols. This computation is syntactic and as such does not require the grounding of meaning of these symbols, these do not have to refer to something. However, in practical applications, such as robots, the grounding of meaning is necessary. Robots are situated in and interact with the real word and consequently the mind of the robot must deal with real world entities. This leads to the practical problem: how the abstract symbols are to be derived from the information provided by the robot's sensors. This is a pattern recognition problem; the presence of an object is to be deduced from patterns of sensory signals. This is also a classification problem. Symbols stand for discrete well-classified entities that can be ordered into ontologies. This would work if it were possible to classify every entity in the world univocally. However, this is hardly the case, classes are artificial and arbitrary. Consequently, every object may be a member of not one but numerous classes (Clancey 1989).

The phenomenal aspects of mind such as the feel of pain, pleasure and perceptual qualia pose also a problem to symbolic theories, because these phenomena are supposed to take place at a sub-symbolic level.

High-level symbolic theories of mind can be formulated as computer programs and can be run on an ordinary computer.

Low-level sub-symbolic theories describe system reactions and interactions between low level signals in neural systems and architectures. The equivalents of higher level symbols may exist and may consist of a number of low level signals. Higher level symbols of this kind have fine structure and consequently modified symbols can depict modified entities. Absolute object recognition and classification is not necessary, an object may be seen in different roles depending on the context. Only the interactions between low-level sig-

nals are defined in algorithmic ways and are built in the neural architecture. Higher level cognitive functions arise from these via adaptation and learning. No implicit or preprogrammed algorithms for high-level operations are provided. The phenomenal aspects of the operation, if there will be any, are expected to be related to the dynamics of the system reactions that arise in the architecture. True realization of this approach calls for specific hardware that is able to support dynamic system reactions.

Low level sub-symbolic theories of mind can also be formulated as computer programs, which can be run on an ordinary computer. However, these executions should be seen only as simulations of the proposed neural hardware. The simulation of very large number of synapses usually calls for some simplifications and shortcuts in order to keep the processing time reasonable. Therefore these simulations do not necessarily produce all aspects of the theory and one should be critical and realistic when attributing phenomenal aspects to these simulations.

Which approach, the high level or low level, symbolic or sub-symbolic, would be the preferred one? Would a hybrid symbolic/sub-symbolic approach be able to combine the strengths of both approaches while avoiding their shortcomings? The brain is not a symbolic computer, but a biological neural network, which operates with sub-symbolic signals. Yet it manages to handle symbolic thought, too. Therefore, there must be a way in which a sub-symbolic system bridges naturally the gap between sub-symbolic and symbolic representations. For instance, Kelley has proposed that no gap actually exists, the sub-symbolic and symbolic representations are the ends of an intellectual continuum (Kelley 2003). In the same sense, Haikonen has proposed a way in which a neural system can utilize sub-symbolic representations as higher level symbols (Haikonen 2007).

Which aspects should an implementable theory of mind cover? Cognitive psychology has described many processes of mind and these can be used as a starting point. A successful theory should also explain meaning, qualia and consciousness in implementable terms.

An implementable theory of mind would be an engineering theory, which is described by commonly accepted engineering terms; mathematics, operational diagrams, circuit diagrams, system architectures and specifications. On the other hand, the aspects to be described belong to the realm of cognitive sciences. Here the interdisciplinary nature of this undertaking will be an interesting challenge and consequently engineers will have to study a bit of cognitive psychology and brain theories. Thereafter the engineering cycle of <identification of requirements – specification – design – test – revision> will hopefully meet this challenge.

3 The Processes of Mind

All animals that can execute motor responses have also more or less complicated nervous systems. One fundamental function of these nervous systems is the generation of motor response commands. In order to respond to something a nervous system must acquire information about that something.

Therefore some kinds of sensors that detect external and internal conditions are also necessary; the nervous system must be perceptive. In this kind of a system a motor response can be a reaction that is triggered by a sensory percept. Useful action may result, but sometimes blind reactive responses may be harmful or even fatal. A more complicated system may remedy this shortcoming by evaluating the fitness of the intended action with the help of experience; memorized instances of similar cases and the good/bad value of their outcomes. This calls for the ability to evoke memory-based imagery and to imagine itself executing the act. This, in turn, is related to the ability of "thinking about itself". If these capabilities were accepted as prerequisites for a mind then the minimum functions of a mind could be readily identified and they would be: Perception, reaction, deliberation and reflection. This conclusion has been reached and shared by Nilsson, Sloman and others including the author (Nilsson 1998, Sloman 2000a, 2000b). Thus the elementary functions of mind are seen as those of a controller and planner.

The above list of basic functions offers a good starting point, but a more detailed evaluation of the functions and processes is necessary for an implementable theory of mind. Cognitive psychology identifies the following processes of cognition: Perception, prediction, attention, learning, memory, understanding, reasoning, imagination, introspection, general intelligence, emotions, volition (See e.g. Aschraft 1998, Nairne 1997, Haikonen 2003). This list must be augmented with the additional functions and processes of pleasure, displeasure, pain, good/bad criteria and match/mismatch detection. Additionally, special and important hallmarks of human mind are the use of natural language and the flow of inner speech. However, these listings of cognitive functions should be mainly considered as kinds of check-lists; the listed functions and processes are not necessarily autonomous and independent of each other and some of them may be only loose descriptions of phenomena created by completely different processes. Nevertheless, the challenge for the potential developer of artificial mind theories becomes now visible; instead of actually clarifying the essential issues these lists highlight the wide spectrum of functions and processes to be quantified. Things are complicated further by the fact that these items relate to the functional layer of mind; the content layer is another story. Yet it is the content that determines what we are; our behavior, motives, values and culture. These are the subject of behavioral, social and cultural studies and go beyond the basic theory of mind.

4 Mind, Meaning and Understanding

Our thoughts are intentional; they are about something, they refer to something and have meaning that we understand. Likewise, a robot with an artificial mind should understand and have meaningful thoughts. Folk psychology has it that reasoning cannot take place without understanding and the utilization of meaning. However, it is known that mathematical and logical reasoning operates without meanings; no semantics, only syntactic rules. One plus one is two no matter what is being counted, be it apples or animals. It is exactly this abstraction property of mathematics and logic that make them so powerful. A computer works well without any grounding of meaning. Accordingly the computational theory of mind proposes that understanding can be effected via syntactic computation. This view has been criticized and op-

posed by e.g. Searle (1980, 1984, 1997). On the other hand the opponents of Searle have argued that syntax will somehow convey semantics if executed properly.

The question about semantics and syntax is a complicated one. In many cases it would seem that syntax would indeed suffice, but then there are cases that are not so clear. Consider the following examples:

- A candy bar has two sections. How many sections remain if one section is cut away? (two minus one is one).
- A candy bar has two ends. How many ends remain if one end is cut away? (two minus one is two).
- A triangle-shaped cookie has three corners. How many corners remain if one corner is cut away? (three minus one is four).

Mathematics may be context-free, but its application may not be. Simple arithmetic seems to fail in two examples here and correct answering seems to call for the visualization of the problems; the evocation of topological meaning. In general terms, in this example the "meaning" of an entity would seem to involve potential connections to a number of other mental concepts and physical world objects and "understanding" would seem to involve the proper activation of the relevant connections amongst all the possible connections and the consequent evocation of the relevant concepts.

An implementable theory of mind must address the problem of meaning and understanding properly. This requirement is especially apparent in the context of robotics. A cognitive robot must be able to understand what it is doing and why. Robots must also understand the commands given by their masters and they must be able to communicate their intentions to their masters. A robot cannot obey the command "go to the kitchen and bring me a soda can" if it does not understand the meanings of the words and the structure of the sentence, how these relate to the world and to the executable actions of the robot. But even this is not always sufficient. For instance, the master of the robot may give a verbal command: "Robot, please" or the master may simply snap his fingers. What is the robot supposed to do? This depends on the situation and context; perhaps the robot is expected to serve drinks to guests or escort somebody out. The robot must also understand the implicit conditions and limitations of each situation; while executing given commands the robot must not cause any collateral harm and damage.

5 Mind and Qualia

Human consciousness is characterized by qualia, the "phenomenal feel and quality" of every percept. Qualia are the way in which sensory information manifests itself in mind. Therefore, to be phenomenally conscious is to have qualia-based perception of the environment and self.

Qualia depict qualities of the sensed entities. The sensory faculties of vision and audition generate qualia that are related to the properties of the entities in the visual and auditory scene. It is known that visual and auditory stimuli are transformed by the eye and ear into neural signals that project into the depths

of brain; yet the resulting qualia that depict visual objects and sounds seem to reside outside. In this way the individual comes to experience its existence as a center point in the world. This illusion does not readily take place in digital signal processing and therefore calls for a special explanation.

Qualia are subjective; there is no known way in which one's subjective experience, own feel of qualia can be transmitted to another person. However, the similarity of our biological built allows us to assume also similar feel of qualia. Thus we may assume with some confidence that a given real world quality such as that of a sound, taste or color will evoke same kind of qualia in different persons. But even here exceptions exist. For instance, a person with normal color vision has no way of knowing how a color blind person experiences the colors of red, green or brown. A color blind person may report no difference between these colors, but which would be the actual percept quale? Would it be the same as normal person's perception of red or perhaps green or brown? Or would it be something completely different?

Qualia are often associated with good/bad property; in fact they as themselves may feel pleasant or unpleasant. Music capitalizes on this property of qualia, the pleasantness of certain sounds, chords, rhythms and melodies. Without qualia music would be all but pointless. Thus, it seems that artificial minds that do not have qualia would not enjoy music in the same way as most humans do, if not at all, as the feel of enjoyment itself is based on qualia.

Computational theories of mind do not consider any feel of qualia as a necessary part of the cognitive process. In fact, it would be quite difficult to maintain that the execution of a computer program would involve any kind of subjective feel in a computer. Why would this feel be necessary anyway? Digital signal processing methods are quite able to handle qualities of the world. They can acquire and quantify information about physical qualities and represent these in numeric form. Powerful numeric algorithms for transformations, filtering, pattern recognition, motion detection and other signal processing tasks are available and can solve many of the related problems without any considerations of qualia. However, if necessary, computational qualia can be defined and represented as numeric values of variables: "if the variable p has the value ten, then the system is in great pain". But then, obviously this line of execution is an example of naïve anthropomorphism and should be recognized as such.

On the other hand, low-level sub-symbolic theories do not exclude the possibility of subjective qualia. For instance it has been proposed that the subjective feel of pain and pleasure would be related to system reactions in a system consisting of associative neuron networks (Haikonen 2003). Further research is called for also along this avenue.

At this moment the actual nature of qualia is still some kind of a mystery and a major challenge to any worthwhile theory of mind.

6 Mind and Emotions

Human mind is also characterized by the spectrum of emotions that can be triggered by various conditions. Emotions have been seen as non-rational states of mind that should not have any part in rational thinking. However, in recent years research has revealed that emotions do have an important role in cognition (LeDoux 1996, Damasio 2000, 2003). Percepts are seen to have emotional significance, which guides attention and modulates learning. Emotional significance is also seen to be an important factor in judgment and decision-making. Emotions seem to have motivational effects, too. Emotions have some connection to qualia; to be in an emotional state feels like something. In which way should emotions be incorporated in an implementable theory of mind? Could emotions be useful for a robot? Some attempts towards this direction already exist (e.g. Dodd and Gutierrez 2005, Haikonen 2003, 2007, Lee-Johnson and Carnegie 2006, Shirakura, Suzuki and Takeno 2006)

7 Mind, Consciousness and Self

Are mind theories also theories of consciousness? In nature minds and consciousness seem to go together, all beings that seem to have minds seem to be conscious, too. The content of consciousness is also mind's content at any moment even though mind is seen to involve also sub-conscious components. Usually mind and consciousness are attributed to an autonomous actor who is aware of itself, its mind and existence. A proper theory of mind should address also the problems of consciousness and self-consciousness.

The philosophy of consciousness divides the problem of consciousness into two parts, namely the so-called easy and hard problems (Chalmers 1995). The easy problem is related to the explanation of the cognitive functions that consciousness is supposed to execute or are otherwise associated with consciousness. The hard problem is related to the phenomenal aspect; consciousness as subjective qualia-based perception of the environment and self, the "feel". A developer of conscious machines may wish to define the focus of his pursuit along this demarcation. Machines that are supposed to execute the easy problem, but not the hard problem may be called "functionally conscious". Machines that execute also the hard problem may be called "phenomenally conscious".

The concept of "functional consciousness" is not without problems. This concept could be justified if consciousness actually executed a certain function. Consequently, a machine could be said to possess functional consciousness if it executed the same or a similar function. Baars (1997) proposes a number of functions for consciousness: Prioritization, access to unconscious resources (this is trivial tautology!), decision making and executive control, recruiting and controlling actions, error detection, understanding, and others. Given these functions there are two possibilities:

1. These functions are executed because the system is conscious, i.e. "consciousness executes these",

2. The system is conscious because these functions are executed by the system. In this case the style of execution may make the difference between a conscious and a non-conscious system.

Cognitive functions can be executed without consciousness and therefore are not a strong indication of any functionality of consciousness. On the other hand, decision making has been seen as a proof of the proposition that consciousness has a functional executive role. However, Libet's experiments and other studies (Libet 1993, Wegner 2003) seem to show that consciousness does not have decision-making power and decisions are made sub-consciously. Thus it may be possible that consciousness does not execute any function, instead it may be only an inner appearance in the system created by a special way of execution of the supposed functions of consciousness (Haikonen 2007). This leads to the following conclusion: If consciousness were only an inner qualia-based appearance with no function then no functional consciousness in the previous sense could exist. A system that would reproduce only the outer appearances of a naturally conscious system would not create the equivalent of the subjective qualia-based inner appearance of consciousness. Consequently no proper emulation or simulation of consciousness would take place. "Functionally conscious machines" would be functional but not conscious; the label would promise too much.

8 Mind and Inner Speech

Human mind is characterized by inner speech. In folk psychology inner speech is often seen as thinking and the main content of the human mind and is understood as a main difference between man, animals and machines.

The running of a computer program does not involve inner speech. Consequently inner speech has been largely ignored by AI researchers. However, cognitive psychology and neuroscience has seen inner speech as a key component of consciousness (e.g. Morin & Everett 1990, Morin 1993, 1999, 2003, 2005, Siegrist 1995, Schneider 2002). Recently also some machine cognition researchers have recognized inner speech as a relevant component of human-like cognition and consciousness (Clowes 2006, Haikonen 2003, 2006, 2007, Steels 2003a, 2003b). The relevance of inner speech to consciousness seems deceivingly obvious; how else could we know what we think if we did not hear our inner speech? This observation may easily lead to the conclusion that language and inner speech were necessary conditions for consciousness. However, this is not necessarily the case; there are also other forms of conscious thinking such as visual and kinesthetic imagination.

Humans explain their situation to themselves via the silent inner speech. Morin (2005) sees this self-talk as a device that can reproduce and extend social interactions leading to self-awareness. During social interactions people may receive comments about themselves, the way they are and behave. Inner speech may repeat these comments as such or as first-person transformations. This may lead to enhanced awareness of the commented features and to modified self-image.

Human mind is able introspect itself, it is self-aware. Duval and Wicklund (1972) define self-awareness as the state of being the object of one's own attention. This would include the paying of attention to one's own mental content such as percepts, thoughts, emotions, sensations, etc. Morin (1990, 2005) has seen inner speech as important means for introspection and processing of information about the self and the creation of self-awareness.

Inner speech utilizes a natural language. Natural language understanding and generation is a notoriously difficult discipline that, unfortunately, a mind theorist is not able to avoid. Existing machines do not "think" in a natural language and existing linguistic theories do not really help there. It may be possible that bold new approaches to linguistics will be called for.

9 Conclusions

An implementable theory of mind would be a theory that is expressed in engineering terms which allow the simulation of the processes of mind or the design of hardware systems that support the said processes. The contents and processes of mind should be meaningful, that is, mind objects should refer to real world entities. Thus perceptive processes are called for; these processes would execute symbol grounding. Cognitive psychology has identified several cognitive functions. It is obvious that an implementable theory of mind should be compatible with these.

Human mind operates with qualia. The act of perceiving the world through qualia seems to be the very essence of human consciousness. Artificial minds without qualia may be called functionally conscious, but not without problems; functional consciousness may be a valid concept only if consciousness actually executed some function and that function could be emulated.

Human mind utilizes inner speech. This inner speech is one hallmark of human consciousness that animals most probably do not share. Simple minds without inner speech can be envisioned, but a theory of mind may not be complete if it does not include the phenomenon of inner speech and allow communication via a natural language.

An implementable theory of mind should address the workings of the functional layer. Except for simple reactions and reflexes the behavior of the system with a mind would be determined by the mind's content; the accumulated experience, emotional states, motives and good/bad values. This process would be most interesting to observe in an artifact, yet it would belong to the realm of behavioral psychology and would be beyond the basic theory of mind.

The solving of the technicalities of mind would have important implications to information technology and also our own philosophical view about ourselves. The spectrum of unsolved issues provides great opportunities to creative researchers.

References

Anderson J. R., Bothell D., Byrne M., Douglass S., Lebiere Ch., Qin Y. (2004). An Integrated Theory of the Mind. *Psychological Review*, 2004, Vol. 111, No 4, 1036 - 1060

Ashcraft M. H. (1998). *Fundamentals of Cognition*. New York: Addison Wesley Longman Inc.

Baars B. J. (1997). *In the Theater of Consciousness*. New York: Oxford University Press

Chalmers D. J. (1995). Facing up to the Problem of Consciousness. *Journal of Consciousness Studies*, Vol 2, No 3, 1995, pp. 200 – 219

Clancey W. J. (1989). The Frame of Reference Problem in Cognitive Modeling. *Proceedings of 11th Annual Conference of the Cognitive Science Society*, Ann Arbor. Lawrence Erlbaum Associates; 1989. pp. 107–114

Clowes R. (2006). The Problem of Inner Speech and its Relation to the Organization of Conscious Experience: a Self-Regulation Model. In T. Kovacs and J. Marshall (Eds.) *Proceedings of the AISB06 Symposium*. The Society for the study of Artificial Intelligence and the simulation of behaviour, UK. Vol. 1. pp. 117 – 126

Damasio A. (2000). *The Feeling of What Happens*. London: Vintage

Damasio A. (2003). *Looking for Spinoza*. USA: Harcourt Inc.

Dodd W., Gutierrez R. (2005). The Role of Episodic Memory and Emotion in a Cognitive Robot. *Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication* (RO-MAN), Nashville, Tennessee 13 – 15 August 2005; 692–697

Duch W. (2005). Brain-Inspired Conscious Computing Architecture. In *The Journal of Mind and Behavior* Vol. 26 (1-2) 2005, pp. 1 - 22

Duval S., Wicklund, R. A. (1972). *A theory of objective self awareness*. New York: Academic Press

Fodor J. (1975). *The Language of Thought*. New York: Thomas Crowell

Haikonen P. O. (2003). *The Cognitive Approach to Conscious Machines*. UK: Imprint Academic

Haikonen P. O. (2006). Towards Streams of Consciousness; Implementing Inner Speech. In T. Kovacs and J. Marshall (Eds.) *Proceedings of the AISB06 Symposium*. The Society for the study of Artificial Intelligence and the simulation of behaviour, UK. Vol. 1. pp. 144 – 149.

Haikonen P. O. (2007). *Robot Brains, Circuits and Systems for Conscious Machines*. UK: Wiley & Sons

Kelley T. D. (2003). Symbolic and Sub-symbolic Representations in Computational Models of Human Cognition. *Theory & Psychology* 2003;13(6):847–860

LeDoux J. (1996). *The Emotional Brain*. New York: Simon & Schuster

Lee-Johnson C. P., Carnegie D. A. (2006). Towards a Computational Model of Affect for the Modulation of Mobile Robot Control Parameters. *3rd International Conference on Autonomous Robots and Agents* (ICARA 2006) 12 – 14 December 2006, Palmerston North, New Zealand

Libet B. (1993). The neural time factor in conscious and unconscious events. *Experimental and theoretical studies of consciousness*, Wiley, Chichester (Ciba Foundation Symposium 174), 1993, pp. 123 – 146

Mckenna T. M. (1994). The Role of Interdisciplinary Research Involving neuroscience in the Development of Intelligent Systems. In: Honavar V, Uhr L, editors. *Artificial Intelligence and Neural Networks: Steps toward Principled Integration*. USA: Academic Press; pp. 75–92

Morin A. (1993). Self-talk and self-awareness: On the nature of the relation. *The Journal of Mind and Behavior*, 14: 223-234

Morin A. (1999). On a relation between self-awareness and inner speech: Additional evidence from brain studies. In *Dynamical Psychology: An Interdisciplinary Journal of Complex Mental Processes*. Retrieved from http://cogprints.org/2557/ on 14.12.2005

Morin A. (2003). Let's Face It. A review of The Face in the Mirror: The Search for the Origins of Consciousness by Julian Paul Keenan with Gordon C. Gallup Jr. and Dean Falk. *Evolutionary Psychology*, 1:161-171

Morin A. (2005). Possible links between self-awareness and inner speech: Theoretical background, underlying mechanisms and empirical evidence. In *Journal of Consciousness Studies*. Volume 12, No. 4-5, April-May 2005

Morin A., Everett, J. (1990). Inner speech as a mediator of self-awareness, self-consciousness, and self-knowledge: an hypothesis. In *New Ideas in Psychol*. Vol 8. (1990) No. 3, pp. 337 - 356

Nairne J. S. (1997). *The Adaptive Mind.* USA: Brooks/Cole Publishing Company

Newell A, Simon H.A. (1975). Computer Science as Empirical Inquiry: Symbols and Search. 1975 ACM Turing Award Lecture. *Communications of the ACM*. March 1976, Vol. 19 No. 3. pp. 113 – 126

Nilsson N. J. (1998). *Artificial Intelligence: A new Synthesis*. San Francisco: Morgan Kaufmann

Penrose R. (1989). *The Emperor's New Mind*. Oxford University Press.

Putnam H. (1961). Brains and Behavior. Originally read as part of the program of the American Association for the Advancement of Science, Section L (History and Philosophy of Science), December 27, 1961

Schneider J. F. (2002). Relations among self-talk, self-consciousness, and self-knowledge. *Psychological Reports*, 91: 807-812

Searle J. R. (1980). Minds, Brains, Programs. *The Behavioral and Brain Sciences*, number 3, 1980, pp. 417 - 427, Cambridge University Press

Searle J. R. (1984). *Minds, Brains & Science*. London England: Penguin Books Ltd

Searle J. R. (1997). *The Mystery of Consciousness*. London: Granta Books;

Shirakura Y., Suzuki T., Takeno J. (2006). A Conscious Robot with Emotions. *3rd International Conference on Autonomous Robots and Agents* (ICARA 2006) 12 – 14 December 2006, Palmerston North, New Zealand

Siegrist M. (1995). Inner speech as a cognitive process mediating self-consciousness and inhibiting self-deception. *Psychological Reports,* 76, pp. 259-265

Sloman A. (2000a). From intelligent organisms to intelligent social systems: how evolution of meta-management supports social/cultural advances. *Proceedings of the AISB'00 symposium on how to design a functioning mind*. UK: University of Birmingham; pp. 130 – 133

Sloman A. (2000b). Introduction: Models of Models of Mind. *Proceedings of the AISB'00 Symposium on How to design a functioning mind*. UK: University of Birmingham; pp. 1 – 9

Steels L. (2003a). Language Re-Entrance and the "Inner Voice". In O. Holland (Ed.), *Machine Consciousness*, pp. 173 – 185, UK: Imprint Academic

Steels L. (2003b). Evolving grounded communication for robots, In *Trends in Cognitive Science*, 7(7), July 2003, pp. 308 – 312

Wegner D. M. (2003). The mind's best trick: how we experience conscious will. *TRENDS in Cognitive Sciences*, Vol. 7 No 2 February 2003 pp 65 – 69

Interview:

Questions for a Journal of Mind Theory

Konrad Talmont-Kaminski and Ricardo Sanz

Institute of Philosophy, Marie Curie-Sklodowska University
Autnomous Systems Laboratory, Uniersidad Politécnica de Madrid

1 Context & Mission

In the past formal and abstract models have attempted to shed light on the topics of the mind and brain, robotics and artificial intelligence. This has created a vast proliferation of information, which currently lacks a single dominant model for understanding mental processes. This journal's aim is to consolidate and explore these formal and abstract tools for modeling cognitive phenomena, and to create a more cohesive and concrete formal approach to understanding the mind/brain, striving for precision and creating clarity in this topic of interest.

This is a battery of questions on the very possibility of JMT posed by JMT managing editor Jaime Gomez and answered from two apparently very different perspectives: philosophy and engineering.

2 Questions

1) First off, for the purpose of putting things in perspective, There seems to be some skepticism about the usefulness of formal approaches. Is formal logic the best mode for thinking about mental processes? Are the grounds of validity of the laws of logic to be found in language, in conceptual structures, in the nature of representation, in the world, or where?

RS:

Formal logic is an abstract framework and as such, the grounds of its validity are to be found in its own structure. The programs of Frege and Hilbert established this thread and the axiomatizations of Russell & Whitehead or Peano reflected this into logics and set theory; a program that was partially broken by Gödel results. The question is whence the formal can bear any strong rela-

tion to the real. My belief is that the answer is yes. The reason for believing this is a question of plain evidence: laws of physics seem to be in strong correlation with reality (cf. the bewilderment shown by Wigner). And it is not just a question of approximation –that we can always approximate any data set with an arbitrarily complex function– it is a question of how simple while precise the laws are in the formal side and in the real side.

KT:

Given that I am supposed to present the limitations of formal methods I should probably begin by making it clear that I do not think that any serious inquiry into cognition can go far without the use of such tools. Formalization is a vital element of the approach used by scientists to understand the world, allowing us to obtain a precise grasp of natural phenomena, as well as revealing to us when we lack such understanding. Any attempt to think about the mind without recourse to formal tools would be unlikely to get far beyond a collection of insights. Having said that, I think it fair to say that for much of the previous century assessments of the value of formalization have been overly hubristic. The depressing longevity of GOFAI seems to me one aspect of this phenomenon. In short, my position is not so much skeptical as pluralist – formal tools are necessary but far from sufficient. And they carry with them numerous problems.

The utility of formal tools for investigating mental processes will look very different depending upon the context in which one places those processes. One can see them as the imperfect implementation of reasoning strategies, the structure of which is investigated by logic and other formal approaches. One can also see them as the evolved means certain organisms use to direct their interactions with their environment. While the two views – the logical and the biological – are not necessarily contradictory they do start with very different kinds of assumptions. The biological view forces a bottom-up perspective in which mental capabilities are the result of evolutionary and developmental processes that began with nothing more than simple chemical replication. On this view cognitive limitations are not an unfortunate detail to be abstracted away from but are at the heart of how we've managed to incrementally transcend those limitations by making efficient use of the limited resources we did have access to. Also, biological solutions are not optimal nor are they universal. Instead, they are at best adequate to current needs in the particular environment they are usually applied. If one uses logic as a model of what mental processes are meant to be like one ends up with highly unrealistic assumptions, such as a deductively closed set of beliefs, that must be continuously fought against if one is to arrive at something plausible (Brown 1990, Hooker 1995, Bickhard 2009). The point isn't that all formal models must embody those assumptions. They do not. However, to avoid having to identify them as problematic one by one, it is far preferable to start from the point of view that mental processes are to be understood as biological phenomena.

It might be argued that while the logical view does a poor job of description, it is primarily meant to be normative – it talks not of what human mental processes are like but of what we should be trying to make them like. Yet this move does not buy the freedom from human foibles one might desire. On the

one hand, it is generally accepted that ought does imply can. This means that while the details of current limitations may be no longer as relevant, the overall unavoidability of limitations is not. On the other hand, putting the view in normative terms raises the question of the relevance of the norms. Why should people try to have true beliefs, for example? The fall back claim that this is simply what it means to be rational solves nothing unless one is comfortable with the conclusion that the choice to be rational is arbitrary. Any substantive naturalist answer, however, will have to come back to the human predicament of needing to make our way in a world whose capacity to affect us exceeds our understanding of it.

The reservations I have raised may seem to be merely grounds for a view of reasoning that acknowledges the relevance of both the biological grounding and the logical structure. The two are not on a par, however. The problems with the logical view of human cognition can be ultimately traced back, I would argue, to Hume's old problem: a problem that I see as necessitating the naturalist (rather than scepticist) response that there can be no universal solution to the problem of how to come to grips with our world. While the problem has come to carry the name of the problem of induction it might be much better titled the problem *for* induction, given that it was always clear, pace Popper (1959), that deduction had nothing useful to say on the topic. What is worse, as Couvalis (2004) points out, the problem also affects our justification for using deductive arguments since our use of these must presume our ability to use them is reliable, and evidence for this must be inductive. So long as induction was conceived of as a logic, no useful response could be given to the problem. In the end, the trick has turned out to be not to seek a solution by proposing ever more complex logics but to learn to live with the problem – "The Humean predicament is the human predicament", to quote Quine (1969: 72). Without an overall framework to work within we are left muddling along in the best of biological fashion. In our efforts we are free to make use of whatever tools we can access and, undoubtedly, formal tools are among the most useful. However, they are only made use of within the broader biological context.

In talking about this way of seeing cognition one is obliged to bring up Herbert Simon. Through his collaboration with Alan Newell, Simon can be deemed to be one of the father's of GOFAI (Newell, Simon 1976). Throughout his life he made brilliant use of numerous formal methods to model aspects of human mental processes. At the same time, however, his bounded rationality approach brings together the biologically-informed points I have been seeking to make (Simon 1983). Something of the significance of the view of rationality he provides us with can be seen in the ensuing disagreement between two groups of researchers who adopted his concept of heuristics: Kahneman and Tversky (Tversky, Kahneman 1974) on one hand and Gigerenzer and his colleagues on the other (Gigerenzer, Todd, ABC Research Group 1999). Not surprisingly, I stand with Gigerenzer in claiming that Kahneman and Tversky failed to appreciate Simon's overall position when they used statistics as the standard to which they compared human cognitive heuristics instead of examining how effective the two are when dealing with real-life data. In effect, I see Gigerenzer's work as a very good example of just what can be done using formal tools within an overall biological view of human mental processes.

Where does this leave the laws of logic? Clearly, I can not hold that they are the laws of reasoning. Hume put that notion to rest, I think: even if many did not notice and insisted on exhuming it. Yet, I have no wish to claim laws of logic to be a human construct. Having said that I must own up to a certain degree of fascinated ignorance as to what their actual nature is. One insight I do find convincing in this context is Peirce's (1905) definition of what is real as that which has properties that are independent of what we think of that thing. Thus, the Earth's equator is real in this sense, as its length is roughly 40 thousand kilometers independently of what we think about it.

Are embodied and situated approaches more relevant than the use of formal tools for the modeling of biological phenomena, in particular mental processes?

RS:

I don't really understand what "embodied" and "situated" mean. In a very precise sense all real systems are embodied and situated. All they have a body and are placed somewhere. For some authors, "embodied" does not mean just "having a body" but having an operation driven by a "mind" that is scattered through all the body. I cannot but agree with this spread mind idea in the understanding that minds are informational-control processes that are distributed through all the body. The problem is then not the question of "embodiment" but the very possibility of existence of "non-embodied" minds. There is no way of having a real system that is purely abstract. Abstractions are necessarily reified if they exist. Therefore, the embodied mind vs formal mind is a false dichotomy. AI systems based on inference engines are as embodied as any robot in a strong ontological sense. Formal tools are used to think about systems and then mapped into embodied and situated realizations.

A different consideration places the distinctions embodied-non embodied and situated-non-situated in the side of the thinker, scientist or engineer and not in the target system itself. This means that there is a way of thinking and building robots that may be labeled as "embodied and situated robotics". The same can be said for the analysis of alive systems or for the theorizing. The problem here is very simple indeed. What can we say of a model of a system that puts the mind in the brain and not through all the body? What we can say about this kind of model will depend on the system being modeled: in this system the information and control processes happen in the brain, the model may be good; if there are information and control processes beyond the brain, the model is certainly bad.

So, there is no such thing as "embodied modeling"; there are just good models and bad models, and what the "embodied and situated" approach has discovered is a blatant aspect of systems: the dynamical phenomena -especially the one driven by information- can happen in all subsystems. This is nothing new but common understanding in all science and engineering and a central topic of control theory: controllers -minds- must necessarily take into account the dynamics of the body to properly control it. Thinking that a controller can move a robot arm to perform any task in the absence of bodily considerations

is not just "non-embodied robotics" is simply bad engineering. Something that lurks here is the ignorance of a simple fact: given a concrete system, not all behaviors are dynamically possible and hence a working mind cannot actually decide upon the proper actions in the absence of the knowledge of the dynamics. This can be read, as "minds cannot be separated from bodies" as embodiers do or can be simply read as "controllers must take dynamics into account". Nevertheless, this is not a new insight, but common trade in cybernetics and control engineering.

So the question of embodiment is just whether or not bodily dynamics are taken into account when acting. And the question of situatedness just whether or not environmental dynamics is taken into account when acting -the analysis being similar for that of embodiment.

This is so for both the artificial and the natural. The case of the biological phenomenon is a particular case of this more general phenomenon of bodily/world dynamics being of relevance for bodily/world behavior. The "embodied and situated" thinking about biological phenomena or robot construction is hence just trying to avoid the naïve approaches of the illiterate.

KT:

I think the answer to this question is already to be found within it. The two things are not to be thought of on the same level. In the one case we are dealing with tools, tools that no modern scientific inquiry can do without. In the other, we are dealing with a particular approach towards the problem of how to understand mental processes. Any such approach will consist of a set of views as to what is relevant to the phenomenon and how we should go about trying to understand it. The strength of the embodied/situated approach is that it pays due attention to both blades of Simon's metaphorical scissors: one blade being human cognitive limitations, the other being the structure of the environment. All too often, formal models of cognition pay scarcely any attention to either. At the same time, I think that the embodied/situated approach offers a way to understand how symbols can come to have reference without the need for some external reference-maker (Bickhard 2004) – a problem that purely formal approaches can never really hope to deal with.

The "two cultures" conflict that C.P. Snow pointed out looks far from being resolved, indeed "the cultural panorama" seems to be more and more atomized within each of Snow's cultures. The practitioners of science and engineering in the cognitive sciences seem to diverge and question the methods used and achieved by each other rather than reaching any joint consensus.

If the sciences seek to understand the physical world and engineering seeks to build better systems, is it justifiable to build artificial systems designed to carry out tasks that are already easily accomplished by human beings? (We are asking here whether building a humanoid robot that pours a cup of tea properly or walks straight sheds light on the motor sensor mechanisms that humans follow to carry out such tasks).

RS:

The case of the two cultures is a case of misunderstanding of what the cultures are. They are not separated by educational reasons, but because they are not commensurable they have very different purposes. In a rough analysis, the people in both worlds -the scientists and the humanists- try to make their living in a particular niche. In a sharper analysis the "scientific" culture has as single objective to make living easier (this is then decomposed in sub-objectives like understanding how the world works or moving water to our homes). The "humanities" culture has as objective the personal promotion of each author in a cultural context (in a sense it is mostly show business).

In another reading, humanities could be understood as the engineering of cultural assets of experiential value; however the lack of solid theories have driven them to create self-perpetuating myths like the many arts, religions, or political regimes. The arguable attempt of the humanities to build an understanding of the human -a scientific endeavor indeed- is devastated by the lack of objective decision making processes between theories. The proper way of understanding man is cybernetics (in McCulloch words, "understanding human understanding") but the humanities tend, in general, to neglect the role that scientific knowledge about humans has to play in their business.

This is a difficult gap to fill because the problem of scientific theorizing about human thoughts and phenomenologies is daunting. Scientists are not willing to risk or waste their careers in such a minimal hope task and humanists do have the interest but lack the competences for the necessary work.

The question of whether it is justifiable to build artificial systems designed to realize tasks that are already easily accomplished by human beings can be answered in this Snowean gap context. The question of whether it makes sense to build a machine to better understand humans has a simple answer: yes. Our mathematical incompetence to solve -in formal- some human systems problems makes necessary the construction and experimentation to explore -in physical- the enormous range of design alternatives.

The theories of "the human" that we may have are of three kinds:

- Rigorous mathematical theories -as those of physics- that we cannot solve analytically but in their simplest forms far from the complexities of a full-fledged human mind.

- Literary theories from the humanities lacking the necessary intersubjectivity and positive character.

- Executable models, that are reifications -usually in simplified form- of mathematical theories to be used in the performance of experiments.

Obviously the best to have are the rigorous mathematical theories -for the purposes of science, not for the self-promotional purposes of humanities- because they would be universally predictive. However, the possibility of analytically solving billions of simultaneous Izhikevich neuron equations is well

beyond reach. On the other side, the executable models only do a particular prediction -of no universal value- but at least give us something of objective value.

But the value of the executable models can only be such if their results are feed-back into the very theory that is the original backdrop of the executable model. Only in this way the construction of humanoid robots will prove of any scientific value. However, the lack of rigorous specifications of this theoretical backdrop and the model-associated simplifying assumptions convert most of the work on humanoid robotics in just a media show trying to make profit on human empathy for humans. These "researchers" are much worse than the humanists working in their self-promotion because they pretend to be doing real science –placing them between the bullshitters and plain liars of Frankfurt.

KT:

I think I must to a degree question the basic assumption behind this question. While it would be foolish to deny on-going atomization within the sciences, it is vital to point out the degree to which a cohesive approach is being developed that offers the potential for bringing together the various approaches and, ultimately, the "two cultures". I should perhaps begin, however, by stating that I think that it would be a very bad idea for a philosopher to claim that there is no point to some scientific project. Firstly, because philosophy is always going to be wide open to the *tu quoque* response. Secondly, and more importantly, because philosophers have tended to be very poor judges in matters of this sort – I am reminded of the apocryphal philosopher's objection, "It may work in practice, but does it work in theory?"

It is standard among humanists to fear that were science to get its hands on the subject matter of the humanities, it would reduce them to nothing more than atoms and chemical reactions. The fear is, I think, profoundly misplaced though understandable. A writer tries to grasp the whole complexity of a phenomenon in hand and, having failed to do so, feels that he has failed completely. A scientist begins by trying to represent some of the most significant aspects of the phenomenon and only having done that asks what else she has left out of the picture. Thus, the scientist's initial understanding of some phenomenon is bound to look simplistic to the humanist. However, unlike the product of the humanist's pen, the scientist's achievement is only meant to be one step along the road toward a fuller comprehension of the phenomenon in question. More profoundly, science itself changes to accommodate the requirements placed upon it by the phenomena it examines. Today's science is not what it used to be. And that's a good thing. The tools, methods and concepts available to the scientists of one hundred years ago were insufficient to examine the kinds of complex biological phenomena that some of the most exciting of today's science looks at. The humanist does not have to worry that the subtleties of the human condition will be crushed by science – it is the science that will ultimately have to grow subtle enough to do them justice.

Given that scientific methods must accommodate the differing subject matters they are applied to, the various scientific disciplines must grow apart in terms of the way their practitioners work. Yet, this does not necessarily entail an atomization of the resultant world view. The first reason can be seen in the

move toward interdisciplinary approaches that focus upon particular problems and are willing to draw upon methods taken from a variety of different disciplines. I would argue, however, that much more profound is the way evolutionary theory has come to take on an organizing role across a plethora of disciplines (Wilson 2007). This is particularly significant when it comes to the question of the relation between the sciences and the humanities as it is precisely disciplines such as history, sociology, economics, psychology and anthropology which all deal with humanity that are set to undergo the biggest changes. Far from becoming more atomized, these disciplines are becoming linked together by the understanding that they examine different aspects of biological phenomena that are the result of the working out of evolutionary processes. The resultant world-view has underpinned much of the recent science writing that, in direct reference to Snow, has been called 'the third culture' by John Brockman (1995).

In light of such a co-operative approach, work on artificial systems that carry out tasks humans are capable of, offers a plethora of possible benefits. While assuming that humans always achieve those goals in the same ways as the artificial systems would entail ignoring one of the blades of Simon's scissors, such systems do potentially provide an empirical proof of philosophical ideas concerning human capabilities. Understandably, the benefits for philosophers are of the most direct relevance to me, yet the story is much the same for other disciplines. Thus, Rodney Brooks' work on robots has fed back into theories of biological locomotion (Brooks 2002). The process of exchange of ideas is not likely to be straight forward and differences must emerge sooner or later but the exchange of ideas has already proved itself to be very useful.

The hard problem of consciousness or, how we explain consciousness in terms of its neurological basis is a highly controversial topic to which philosophers and scientists have divergent approaches and answers. Philosophers like Ned Block argue that the claim that a phenomenal property is a neural property seems just as mysterious—maybe even more mysterious--than the claim that a phenomenal property has a certain neural basis. Do you think the hard problem of consciousness is a problem, philosophical dilemma or scientific challenge at all?

RS:

The problem of consciousness is no more and no less than a scientific problem. There are some observed regularities and we still lack a positive universal law that captures all of them. This problem is said to be hard because of the apparent difference between first and third person experiences. But there is no such thing as a third person experience. The experience of dropping a bottle of wine from the top of the tower of Pisa is the same type of first – person experience as when drinking the wine. The only issue at stake in first/third person science is abstract repeatability of experiences in controlled settings. With abstract repeatability, I mean the experience described in a level of abstraction that gets rid of unnecessary details. In the case of the drop, we can abstract from the concrete tint of the sunlight, the position of the earth in the orbit or the concrete number and nationalities of the other tourists in the tower. If we describe the experience at a certain abstraction level -e.g. the

number of milliseconds a clock ticked- we can expect to obtain some laws (obviously if the world behaves in such a way).

The separation of what is relevant and what is not to achieve the required level of abstraction is hence the cornerstone of shareable experiences -i.e. the very nature of science and engineering. This is a problem for consciousness research because the human brain is very complex and not easily accessible. Time will come where a deep understanding of brain structure will be ready to be used in the systematic analysis of massive data coming from real time brain observation. Then we will be able to separate wheat from chaff and to establish rigorous correlations ---i.e. scientific laws--- between stimuli and qualia. The laws of redness (?) will come. As will come laws of love and self-hood. There is no mystery here, just ignorance and limited experimental capability.

KT:

If the 'hard problem' is only a philosophical problem then it is not much of a problem at all. But I don't think that it need be just that. Consciousness is a real phenomenon and, as such, it ought to be investigable by scientific means. An essential step is to clarify what the 'hard problem' might possibly be – something that Chalmers' various formulations do not actually achieve. A useful approach, it seems to me, is to break up the problem into Tinbergen's (1963) four questions as they would apply to consciousness:

1. *What, if any, is the function of consciousness?* The by-product explanation seems highly implausible. It seems much more likely that consciousness is a necessary element of any sufficiently advanced representation of the environment and a system's potential for action within it.
2. *How did consciousness evolve?* Investigations of non-human animals that are closely related to us – primarily chimpanzees – have led to significant insights into this problem though we are still far from a satisfactory answer. One important point that appears to be becoming clear is that consciousness is not all or nothing, with various animals exhibiting different elements of consciousness.
3. *What is the mechanism underlying consciousness?* Phenomenology has traditionally been an area of philosophy very much opposed to scientific investigation. That, however, is not longer true with phenomenologists such as Shaun Gallagher (2005) working closely with neuroscientists to answer questions such as this one.
4. *How does consciousness develop in children?* A subject thoroughly investigated by developmental psychology.

It might be argued that the 'hard problem' is not the same as any of the questions I have listed. I would charge, though, that however the problem is formulated, the answers are going to be found by pursuing these four questions. Any lingering sense of mystery will mostly be due to the inherent dualism of human folk psychology. If this means there is no 'hard problem': well, the four questions I have individuated are hard enough to have occupied scores of scientists for decades.

Another problem often referred to in the philosophy of mind literature seems to be the problem of access consciousness or – How can we find out whether there is conscious experience without the cognitive accessibility necessary for reporting that conscious experience, since any evidence would have to derive from reports that themselves derive from that cognitive access?
Do you think the problem of access consciousness is a problem, a philosophical dilemma or scientific challenge at all?

RS:

This argument is permeated with two fallacies: the homunculus fallacy and the first person fallacy. This last refers to the false argumentation about the intrinsic difference between first person experience and third person experience. In the same sense that we can observe and measure someone digesting we will be able to observe and measure someone feeling. The second one comes from thinking that there is a part of our brain/mind that is "me" and the rest is something that I own and/or use. Obviously, language cannot tell us about all what is going in the brain (or the body as a whole). But when we use language to "talk with a person" what we are doing is indeed to "talk with a fragment of the person" but identifying that fragment with the person itself is a mistake. This means that in general, reports on consciousness are necessary fragmentary because the information available to the reporting mechanism is partial.

With the development of a general scientific theory of consciousness and the advances in experimental resources (see previous question) we will be in a situation of scientifically being able to tell -in third person lingo- when someone is having a particular experience. The path to follow will be similar to the present capability of saying when someone is suffering epilepsy or heart attack. Experimental signal will tell us if some phenomenon is happening and in what degree. No need for verbal reports will be then necessary.

KT:

There seem to me to be two ways to understand this problem. The first is as the traditional philosophical problem of other minds (Hyslop 2005). In this form the problem is presented as the question of how to avoid the skeptical conclusion that we cannot know anything about other minds. It is important, however, to appreciate the true weight of such globally skeptical arguments. Hume set out more than one but had the good sense to recognize that they actually lacked the wherewithal to undermine normal practices – even if we could not find any counter to such skeptical worries people would and, probably, should keep on as they do. Being global such skeptical objections do not leave any meaningful choices so that one may as well assume that they are incorrect for some unknown reason and carry on regardless. What can be said about such skeptical arguments is that they can be understood to be indicators of where we are yet to properly understand what it is that we do. Assuming that we can have knowledge of other minds, the fact that we cannot show

why the skeptical conclusion is incorrect shows that we do not understand something about how we obtain that knowledge.

A much more constructive way of seeing the problem seems to be as the question of how it is that people do form beliefs about the mental states of others. In this context two approaches have recently been popular; the first being that people possess a folk theory of mind which allows them to conclude what mental states other people have (Ravenscroft 2004), the second being that people simply simulate what they would feel and think in the other person's situation (Gordon 2004). In this form the question is obviously open to empirical investigation and offers a clear way into solving the problem – a property skeptical worries typically lack.

In The Structure of Scientific Revolutions, T.S. Kuhn argued that science does not progress via a linear accumulation of new knowledge, but undergoes periodic revolutions or paradigm shifts. Kuhn distinguished three main stages in the development of science. The first, pre-science, which lacks a central paradigm, is followed by normal science, during which scientists attempt to explain observed facts within the paradigm. The failure of results to conform to the paradigm is seen not as refuting the paradigm, but as due to researchers' shortcomings. However, as anomalous results continue to be produced, science reaches a crisis. At this stage, revolutionary science leads to a new paradigm, which works some of the old-good results along with the anomalous results into a new framework.

Is the Kuhnian paradigm an inappropriate metaphor for the working of the human mind en soi même? Do you think that Kuhn's account of the development of scientific paradigms provides significant insights into the current state of the cognitive sciences? Which phase of the three are we in now?

RS:

The search for a theory of mind is indeed the very elucidation of the nature of science: correlation between knowledge and reality.

The prescience phase could be equated to Pinker's blank slate but I don't think our minds start from scratch. Or genetic material takes us directly into a "normal-science phase" of the mind that corresponds with the normal operation of a brain when using established knowledge and finely tuning it to the concrete environment of the agent.
Brain revolutions happen continuously in the mind/brain driven by mismatches between the real and the expected. They are true revolutions in the catastrophic sense of Thom and take the form of non-linear attractors as Freeman has showed us.

Concerning the application of Kuhn model to cognitive science, I think that all people agree on the common paradigm given by theoretical neuroscience. In this sense, we are in the phase of normal science but we have a big problem with the character of the anomalous results. For some researchers there are

plenty of anomalous results or even topics not addressed by the theory -e.g. the qualia issue- that the established paradigm does not address -cf. Chalmers "hard problem".

For other researchers -including myself- we are still lacking some pieces in the global picture of the theory but there are not such anomalous results. What we have is a critical incompetence in applying and deriving predictions from the theory when addressing problems of real scale. There are no anomalous results because there are no complete predictions concerning experiments with real systems. Projects like Blue Brain try to test this hypothesis by applying high performance computation to the simulation of big portions of the brain.

KT:

Kuhn's work had a revolutionary effect upon philosophy of science. It played a big role in breaking down a lot of the traditional assumptions and distinctions. Without Kuhn ground-breaking research, the kind of biologically inspired approach that I, among many others, am pursuing would not be likely to find fertile ground to grow. Having said that, the Kuhnian paradigm has not weathered well. The interaction between philosophers, sociologists and historians of science that Kuhn made possible has shown his views to be grossly simplistic – a prime example of such work being provided by the research carried out by Nancy Nersessian (2002). As such, paradigms and normal science may remain useful metaphors to use but we must be wary of putting too much theoretical weight upon them. This means that any evaluation of the current state of cognitive sciences in Kuhnian terms must be taken with a grain of salt. Certainly, there is much that is reminiscent of what Kuhn would say about pre-science. However, this is not for the lack of a paradigm but for their surplus. Many people in the cognitive sciences spend their working days doing what looks very much like normal science – solving small problems within the framework of an overarching research paradigm. Yet, the paradigms they work within often turn out to be very different from those used by others whose work is none-the-less of great relevance to their own. The reason, I think, is that cognitive sciences are a prime example of science that is problem-focussed rather than discipline-focussed. This makes them a meeting place for methodologies and theoretical assumptions that come from fields as wildly disparate as engineering, biology and philosophy. That the result has been as progressive as I think it has is just one more piece of evidence for the conclusion that Kuhn's views did not do justice to science.

Undeniably, the Galilean distinction between primary and secondary properties led to a great advance in science because it permitted scientists to work on physical phenomena while avoiding scholastic disquisitions or the distractions of issues that were perceived by the church authorities as eminently human and therefore divine. Do you think this Galilean distinction between quantitative properties and qualitatives ones is still valid? How close we are to explaining the qualia in a quantitative manner?

RS:

I don't think this distinction is any longer valid or useful. It is clear that in the past it helped focus on certain aspects of the nature that were at the same time easier and more politically correct. However, this is not the case now that we focus on very complex properties -like consciousness- and there are no religious issues at stake.

All the properties -whether primary or secondary- are measured in a process of interaction between an observed object and a measurement device. In a sense, the simpler the interaction process the less effect can be attributed to the measurement device and hence the measurement is closer to being a measurement of an intrinsic property of the object (with the necessary provisions for Heisenberg's uncertainty). In this very sense, qualia are abstractions -higher level measurements- derived from interactions between the sensed object and the sensing agent -a grown-up device, indeed.

The explanation of human qualia in a quantitative manner is certainly coming; but is necessary to perform two previous steps:

The formulation of a theoretical model of qualia of universal character -i.e. not chauvinistically anthropomorphic or animalmorphic. This is in due course in the theoretical consciousness community and will coalesce in some years.

The development of detailed measurement devices of neural activity of higher spatial and temporal resolution able to observe concrete individual neuronal assemblies in vivo. This is a very difficult problem and may not be solved in many years.

However, this second step may not be necessary if the theory of qualia is solid enough as to give precise accounts of all extant phenomena as to be accepted as a satisfactory explanation by the scientific community. This may be necessary for dealing with the irreducible believers on the special nature of consciousness -mysterians-, that may only be convinced after the prediction and confirmation of ad-hoc suitable experimental tests.

KT:

The point that this discussion really seems to lead to is the perhaps surprising degree to which the world has turned out to be describable using mathematics. The realization that it could be, together with an awareness of the primacy of experiment, are two factors that are commonly seen as the intellectual underpinnings of the scientific revolution. In that context Galileo is, of course, a good example to think back to. Faced with this development, numerous times people have sought to cordon off certain areas of experience as, allegedly, beyond the reach of scientific methods – those related to aspects of subjective experience being most commonly deemed beyond the reach of science. However, as I have already pointed out, scientific methods are a moving target, with science having proved itself adept at altering to fit novel problems. In effect, distinctions such as those between primary and secondary properties, have lost much of their philosophical significance – the difference between

them being spelled out in terms of differences in the methodological tools necessary to investigate them. Despite what Steven Jay Gould (1997) thought, science does not have a circumscribed magisterium, its field of competence is instead constantly growing thanks to developments in methodology. This means that anyone who would wish to talk about things 'beyond the ken of science' will find themselves in ever more constrained circumstances. As such, I do not think that there are any fundamental philosophical reasons for thinking that qualia can not be investigated scientifically. Indeed, I would argue that research into qualia is already an everyday occurrence and has been for years. One example of this is, I think, the research into the quite striking effect placebos have upon perceived pain. That these studies might not meet the standards some would hold them up to has more to do with often unrealistic philosophical expectations regarding the nature of knowledge than with the validity of the information we are gathering. In particular, something that has struck me on numerous occasions is how often good science can be done even when the quality of the arguments used – the philosopher's acid test of rationality – is less than sterling. Of course, this does not mean that there are not examples of research going seriously astray for the want of a syllogism.

Aristotle claimed that definitions assumed the existence of some primitive concepts that could not be defined as otherwise we would never be capable of defining anything. Do you find this approach "usable" in the current scientific paradigm? How does this claim relate to our discussion here?

RS:

The end of the apparent definition infinite regression may be a set of closed laws that bound magnitudes and have predictive power. This may be read as a self-sustaining network of definitions (as is the case of physics: $f = m \cdot a$). My impression, however, is that the coming definitions in mind theory will be in terms of extant physics and information theory in a totally reductionistic sense.

KT:

The threat of infinite regress is a traditional philosophical bogeyman that has reappeared in numerous guises. Another of its guises is the idea that there must have been a first mover who put the universe into motion but who was not put into motion by anything else. In every case the infinite regress only threatens because of some failure to understand the matter in hand. In the case of the first mover argument, the solution is to grasp the basics of the big bang. In the case of the idea that there must be primitive concepts, the best solution is to understand language not as a collection of inter-defined concepts but as a tool used by living organisms to coordinate their activities (Millikan 1984 and Gärdenfors 1995). We learn our first words by learning to associate them with conspicuous elements of our environment, aided by a grasp of ostention that develops very early on. We do not require definitions to be able to use language but must merely be able to obtain a pragmatically adequate level of co-ordination with our interlocutors. Indeed, definitions come somewhat late and are often post hoc and clearly inadequate to actual use. And this is even true of scientific and philosophical practice. Aristotle's problem that

meaning must ultimately come from the outside of a set of inter-definitions foreshadowed the problem that all purely formal accounts of language have today. The solution is not to start by looking within language but to look at the interactions between the language-users and their environment.

Real epistemological understanding requires that attention be paid not only to the propositions known or believed, but also to knowing subjects and their interactions with the world and each other. All serious empirical inquirers – historians, literary scholars, journalists, artists, etc., as well as scientists – use something like the hypothetico-deductive method. How does someone's seeing and hearing contribute to the warrant of a claim when key terms are learned by association with these observable circumstances?

RS:

The theory of mind we are looking for will represent the convergence and resolution of ontology and epistemology into one and single theory. The theory of mind will indeed explain and predict how a knowing subject interacts with the world in a meaningful sense. They key here will be the provision of a theory of mind that is indeed a theory of science: how it is possible for knowledge of something to be correlated with the reality of this very something.

The way on how this epistemological-ontological consilience is going to happen can be captured in a simple vision:

Nature is organized and what actually happens rigorously follows laws. There are no surprises or miracles. The question of what the exact nature of these laws is or whether they are probabilistic or not- is irrelevant four our very theoretical purpose. The only requirements for a theory of mind is that they are predictive -to be used as anticipatory tools- and that they are knowable -i.e. can be captured in an information-control infrastructure. In this sense we can trust what someone has learnt -by building associations among observable circumstances- if we are able to discount from the learned laws the concrete, particularity-laden distortions coming from the individual processes of perception and action.

KT:

Obviously, I whole-heartedly agree that one can't talk about cognition and related topics without a rich account of cognizing beings, their interactions and their relationship to their environment – Simon's scissors once more. Conveniently, we happen to have a plethora of such beings available for us to study. I do not mean merely us, humans. I mean all living beings that in some way alter their behavior to adjust it to the requirements of the environment, and that includes pretty much all life including very many single-celled organisms (Campbell 1974, Haack 2007). But what can looking at paramecia teach us about warrant, one could object. Given that we have rejected the possibility of grounding our understanding of cognition in logic or some other formal system, we are left with looking at the previously mentioned four questions that Tinbergen delineated – What are the function, evolutionary history, mechanism and developmental history of cognition? One important

insight that I think we have already gained from this line of research is that cognition is like other evolved systems in that it consists in the addition of new systems that build upon and utilize existing systems. Because of this, the attempt to assemble high-level cognition without first constructing the 'ground floor' was fundamentally misguided (Ziemke, Lindblom 2006).
At the same time, looking for the justification of beliefs in some Platonic heaven is going to be as fruitless as in any other heaven. Ultimately, warrant will have to be paid out in terms of our interactions with our environment. This means that the dualism of causes and reasons has to be broken down. At one end this is being achieved by neuroscientists who are filling in the blanks in the neural pathways that take us from sensory input to object recognition. At the other end is the work being done by naturalistically inclined philosophers who are reconstructing the relations between the normative and the descriptive.

How can works of imaginative literature or art convey truths they do not state? Could incorporating this non-formal more abstract trajectory possibly be useful? How does the precision sought by a logician differs from that sought by a novelist or poet?

RS:

At the end, the problem of conveying truths is a problem of conveying a particular abstract form or structure. The vehicle can be directly abstract and tightly correlated with the aspects and complexities of the truth at hand (cf. Wigner comments of the effectiveness of mathematics in physics). But the vehicle can also be less abstract, more concrete and experiential and still convey the form that constitutes the truth to be transmitted.

The discovery of truth in arts will hence try to get rid of the details of the medium and even the concrete message -remember MacLuhan's analysis- and focus on the abstractions reified in the message. This implies a voyage from the minute details of the physicality into the transcendental forms of the hierarchical abstraction. The main difference between the logician and the artist is not that they try to convey different truths, but that they have a different strategy for the packing of it. Logicians strive for the truth as it is; artists want also the truth but they enjoy more the process of unpacking it from the media.

KT:

I remember how disappointed I was to learn a number of years ago that work in analytical philosophy on truth in fiction was concerned with such questions as how it is that it is true to say that Sherlock lived in Baker Street. There is very little in the tradition that has sought to cast light on the somewhat weightier question of how it is that Conrad's *Lord Jim* contains more truths than all the Mills and Boone novels. Yet, steps towards understanding questions such as this are now being made by writers such as Susan Haack (2008). As with other problems in philosophy, I expect that work by biologists looking at literature (Gottschall, Wilson 2005) will come to play a significant role here, too.

Personally, I am convinced of the epistemic value of art. I am also convinced that, thanks to its open-ended nature, science will be able to make use of it – assuming human society maintains the ability to carry out scientific work. However, I find that the kind of science that could achieve this lies beyond what I can currently imagine.

I fear that at times my answers to some of these questions have veered close to poetry, its epistemic value notwithstanding. Certainly, I have made little effort to argue for most of the claims I have made as that would have called for much fuller responses than I have given. I can only hope that those who wish to find proper statements of such positions will be able to find them, including arguments in support of them, in the works I have referred to. Most definitely, very little of what I have claimed in my answers is original to me. The basic conclusion of what I have been aiming at can be stated as the view that while a theory of mind should, for methodological reasons, be formalized it will necessarily be biological in its content.

References

Bickhard, M. (2004). Process and Emergence: Normative Function and Representation. Axiomathes 14: 135-169

Bickhard, M. (2009). The Biological Foundations of Cognitive Science. New Ideas in Psychology 27: 75–84

Brockman, J. (1995). The Third Culture: Beyond the Scientific Revolution New York: Simon & Schuster

Brooks, R. (2002). Flesh and Machines New York: Pantheon

Brown, H.(1990). Rationality London: Routledge

Campbell, D. (1974). Evolutionary Epistemology. P. A. Schilpp, ed. The Philosophy of Karl R. Popper LaSalle, IL: Open Court

Couvalis, G. (2004). In induction epistemologically prior to deduction?. Ratio 17.1: 28-44

Frankfurt, H. G. (2005). On bullshit. Princeton University Press, Princeton, NJ

Freeman, W. J. (2000). Neurodynamics: An Exploration in Mesoscopic Brain Dynamics. Springer, 1 edition

Gallagher, S. (2005) How the Body Shapes the Mind Oxford: Oxford University Press

Gärdenfors, P. (1995). Language and the evolution of cognition. LUCS 41

Gigerenzer, G., Todd P. and the ABC Research Group, eds. 1999 Simple Heuristics that Make Us Smart New York: Oxford University Press

Gottschall, J. and Wilson, D.S. (2005). The Literary Animal Evanston, IL: Northwestern University Press

Gordon, R. (2004). Folk Psychology as Mental Simulation. Stanford Encyclopedia of Philosophy http://plato.stanford.edu/entries/folkpsych-simulation/

Gould, S. J. (1997). Nonoverlapping Magisteria. Natural History 106: 16-22

Haack, S. (2007). Defending Science – Within Reason Amherst: Prometheus

Haack, S. (2008). Putting Philosophy to Work Amherst: Prometheus

Hooker, C. (1995) Reason, Regulation and Realism Albany, NY: SUNY Press

Hyslop, A. (2005). Other Minds. Stanford Encyclopedia of Philosophy http://plato.stanford.edu/entries/other-minds/

Izhikevich, E. M. (2007). Dynamical systems in neuroscience: the geometry of excitability and bursting. MIT Press, Cambridge, Mass.

McCulloch, W. S. (1965). Embodiments of Mind. MIT Press

McLuhan, M. (1964). Understanding media: the extensions of man. Gingko Press

Millikan, R. G. (1984). Language, Thought, and Other Biological Categories Cambridge, MA: MIT Press

Nersessian, N. (2002). Kuhn, conceptual change, and cognitive science. T. Nichols, ed. Thomas Kuhn Cambridge, UK: Cambridge University Press

Newell, A. and Simon, H. (1976). Computer science as empirical inquiry. Communications of the ACM 19.3: 113-126

Peirce, C. S. (1905). What Pragmatism Is. Collected Papers 5.430-432.

Pinker, S. (2002). The blank slate: the modern denial of human nature. Viking, New York

Popper, K. (1959). The Logic of Scientific Discovery. London: Hutchinson

Quine, W van O. (1969). Epistemology Naturalised. Ontological Relativity and Other Essays New York: Columbia University Press

Ravenscroft, I. "Folk Psychology as a Theory" Stanford Encyclopedia of Psychology http://plato.stanford.edu/entries/folkpsych-theory/

Simon, H. (1983). Reason in Human Affairs Stanford: Stanford University Press

Thom. R. (1975). Structural Stability and Morphogenesis: An Outline of a General Theory of Models. W. A. Benjamin

Tinbergen, Niko (1963). On Aims and Methods in Ethology. Zeitschrift für Tierpsychologie 20: 410-433

Tversky, Amos and Daniel Kahneman (1974). Judgement under uncertainty: Heuristics and biases. Science 185.4157: 1124-1131.

Wigner, E. P. (1960). The unreasonable effectiveness of mathematics in the natural sciences. *Communications in Pure and Applied Mathematics*, 13(1),.

Wilson, D. S. (2007). Evolution for Everyone New York: Dell

Ziemke, T. and Jessica L. (2006). Some methodological issues in android science. Interaction Studies 7.4: 339-342

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, nº 1, 2008

Vindication of a Rigorous Cognitive Science ix
Ricardo Sanz and Jaime Gómez

Feature:
Toward a Computational Theory of Mind 1
James Albus

Regular Articles:
The Mind as an Evolving Anticipative Capability 37
Ron Cottam, Willy Ranson and Roger Vounckx

The Challenges for Implementable Theories of Mind 93
Pentti O. A. Haikonen

Questionnaire:
Questions for a Journal of Mind Theory 105
Konrad Talmont-Kaminski and Ricardo Sanz

ISBN 9788461330201-8

Autonomuos Systems Laboratory
Universidad Politécnica de Madrid