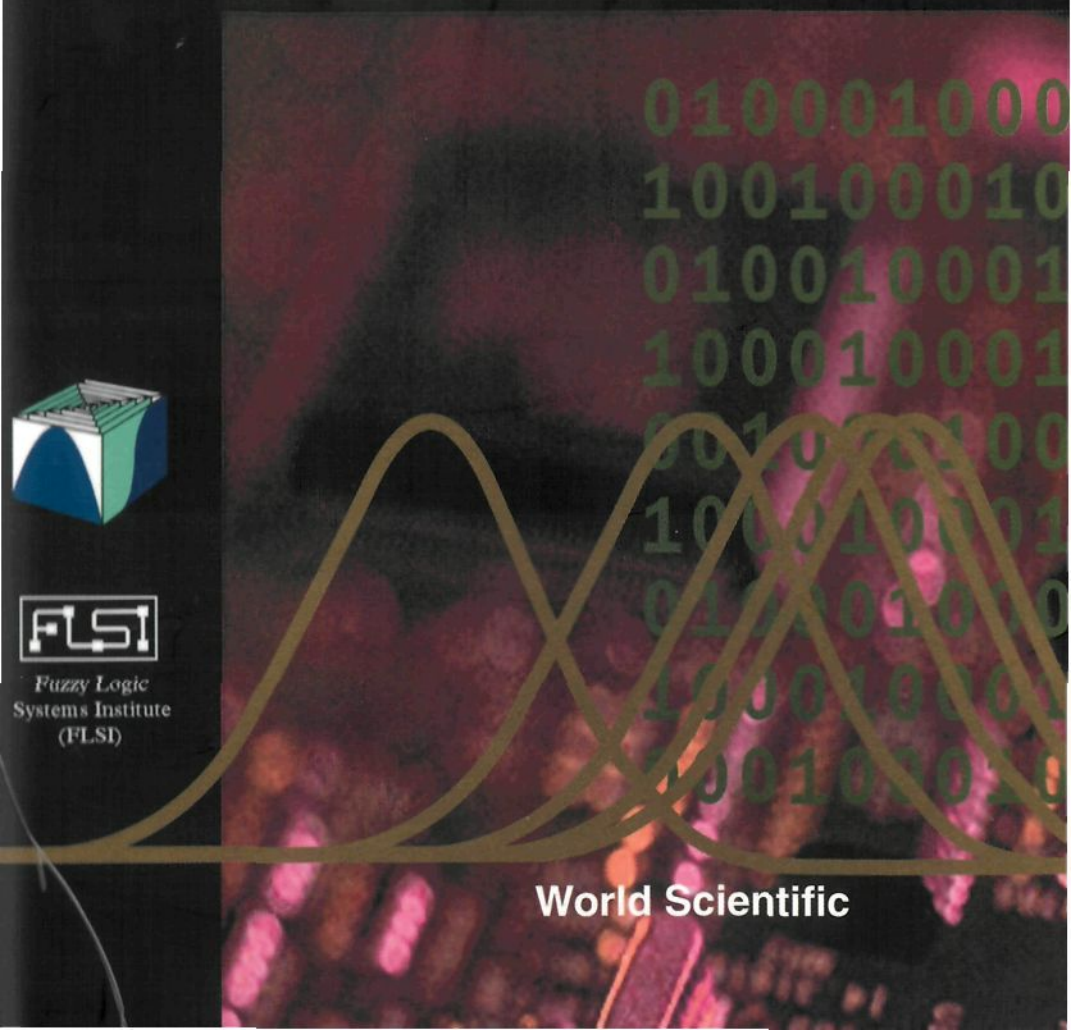
Soft Computing Series — Volume 3

What Should be Computed to Understand and Model Brain Function?

From Robotics, Soft Computing, Biology and Neuroscience to Cognitive Philosophy

Editor: Tadashi Kitamura



Fuzzy Logic Systems Institute (FLSI)

World Scientific

What Should be Computed to Understand and Model Brain Function?

From Robotics, Soft Computing, Biology and Neuroscience to Cognitive Philosophy

Fuzzy Logic Systems Institute (FLSI) Soft Computing Series

Series Editor: Takeshi Yamakawa *(Fuzzy Logic Systems Institute, Japan)*

Published

Vol. 1: Advanced Signal Processing Technology by Soft Computing
edited by Charles Hsu (Trident Systems Inc., USA)

Vol. 2: Pattern Recognition in Soft Computing Paradigm
edited by Nikhil R. Pal (Indian Statistical Institute, Calcutta)

Forthcoming

Vol. 4: Practical Applications of Soft Computing in Engineering
edited by Sung-Bae Cho (Yonsei University, Korea)

Vol. 5: A New Paradigm of Knowledge Engineering by Soft Computing
edited by Liya Ding (National University of Singapore)

FLSI Soft Computing Series — Volume 3

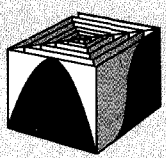
What Should be Computed to Understand and Model Brain Function?

From Robotics, Soft Computing, Biology and Neuroscience to Cognitive Philosophy

Editor

Tadashi Kitamura

Kyushu Institute of Technology, Japan



FLSI

Fuzzy Logic Systems Institute (FLSI)

World Scientific

Singapore • New Jersey • London • Hong Kong

Published by

World Scientific Publishing Co. Pte. Ltd.

P O Box 128, Farrer Road, Singapore 912805

USA office: Suite 1B, 1060 Main Street, River Edge, NJ 07661

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

WHAT SHOULD BE COMPUTED TO UNDERSTAND AND MODEL BRAIN FUNCTION?
— From Robotics, Soft Computing, Biology and Neuroscience to Cognitive Philosophy
FLSI Soft Computing Series — Volume 3

ISBN 981-02-4518-1

Printed in Singapore by Uto-Print

Series Editor's Preface

The IIZUKA conference originated from the Workshop on Fuzzy Systems Application in 1988 at a small city, which is located in the center of Fukuoka prefecture in the most southern island, Kyushu, of Japan, and was very famous for coal mining until forty years ago. Iizuka city is now renewed to be a science research park. The first IIZUKA conference was held in 1990 and from then onward this conference has been held every two years. The series of these conferences played important role in the modern artificial intelligence. The workshop in 1988 proposed the fusion of fuzzy concept and neuroscience and by this proposal the research on neuro-fuzzy systems and fuzzy neural systems has been encouraged to produce significant results. The conference in 1990 was dedicated to the special topics, chaos, and nonlinear dynamical systems came into the interests of researchers in the field of fuzzy systems. The fusion of fuzzy, neural and chaotic systems was familiar to the conference participants in 1992. This new paradigm of information processing including genetic algorithms and fractals is spread over to the world as "Soft Computing".

Fuzzy Logic Systems Institute (FLSI) was established, under the supervision of Ministry of Education, Science and Sports (MOMBUSHOU) and International Trade and Industry (MITI), in 1989 for the purpose of proposing brand-new technologies, collaborating with companies and universities, giving university students education of soft computing, etc.

FLSI is the major organization promoting so called IIZUKA Conference, so that this series of books edited from IIZUKA Conference is named as FLSI Soft Computing Series.

The Soft Computing Series covers a variety of topics in Soft Computing and will propose the emergence of a post-digital intelligent systems.

Takeshi Yamakawa, Ph.D.
Chairman, IIZUKA 2000
Chairman, Fuzzy Logic Systems Institute

This page is intentionally left blank

Volume Editor's Preface

Artificial Intelligence (AI) is one of the established methods to build computational intelligence as a model of brain function. It is obvious, however, that AI cannot solve a problem its closed knowledge base does not cover. This means that it is not possible for AI to control behavior in an unstructured environment because any behavioral knowledge about such an environment cannot be embedded in the knowledge base in advance. That is, traditional AI, such as a production system, can never ascertain the final solution of a Frame Problem, a problem in which subsidiary problems whose number explosively increases must be solved in order to solve the initial problem. A human also may encounter a Frame Problem when he/she lacks default knowledge necessary to solve a given problem. But in that case, the human intelligence may stop tackling the problem halfway and ask others for help. A human body reacts, gets tired of the problem, jumps out of its context and ignores it. This difference between the human intelligence and AI can be recognized as due to human embodiment while AI itself has neither a body nor the knowledge a body acts with. A human body is ready to react to the environment whether or not the process of the intelligent stops, and conversely, human intelligence works when behavior is stopped. A robot also has a body, but even a robot that has a body to act with can have no embodiment unless the interaction between body and intelligence is embedded as such.

In contrast to AI, Subsumption Architecture (SSA), a behavior-based architecture for a mobile robot proposed by Brooks at MIT, employs neither high-level, centrally goal-oriented, nor symbolic algorithms but embeds several fixed reactive behavior modules loosely connected to each other. This architecture brings into focus what is embodied while ignoring what is behind the embodiment. This architecture expects the emergence of meaningful behavior from simultaneous execution of these behavior modules. This architecture has been called a non-Cartesian machine in the sense that it has no central program like a human ego or consciousness to control behavior.

There have been a variety of models of brain function, with concepts between the two extremes of AI and SSA, such as soft AI and artificial neural networks. In order to embed the relationship between intelligence and behavior into a machine, investigations crossing academic borders seem

necessary to understand and model brain function. The first transcendence of borders necessary is the technical one needed to make intelligent machines such as a humanoid robot, an animal-like behavior architecture, an interpreter of fiction, and an evolving learning machine. This technical erosion is conducted into areas such as biology, ethology, neuroscience and psychology as well as robotics and soft computing. These fields obviously have different approaches to analysis of brain function, robotic architecture implementation and computer simulation of neural networks. These efforts at modeling brain function improve our comprehension of brain function.

In the first overstepping of cross-disciplinary boundaries, we may find clues to the difference between models of brain function for behavior and those without behavior. A computer science such as soft computing can build an evolving, intelligent system, but it is impervious to behavioral environments because soft formalism ignores the body, as traditional AI does. Robots may be successful in understanding the meaning of their environments from their behavior because they have bodies to act with in the environment. The neuroscience of the cerebellum, on the other hand, can provide fundamental suggestions for the design of robot motion control.

The second erosion of cross-disciplinary boundaries will cut across scientific areas such as biology, cognitive science and philosophy into comprehensive, less technical, and more abstract aspects of brain function. These aspects enable us to know in what direction and how far an intelligent machine will go. This second academic infringement may be able to suggest: (1) whether or not context can be programmed, (2) what the intelligent machine lacks in approximating a human and/or animal as a whole living being, and (3) what the difference between the whole and part in brain function is.

The structure of this volume is as follows. Chapters 1 to 6 treat the first type of academic erosions of cross-disciplinary boundaries for building intelligent machines: acquisition of a primitive language by an emotional robot (Ogata), animal-like behavior design (Kitamura), a model of the reader's wish inspired by a literary text (Tokosumi), a lesson from neuroscience for a model of behavior (Dufossé) and an evolving software machine of fuzzy neural networks (Kasabov). Chapter 6 gives a general theory of design of

intelligent algorithms on the basis of an overall analysis of artificial and natural intelligences (Teodorescu). Chapters 7 to 11 cover the second type of erosion. Chapter 7 explains the necessity of paradox inherent in the observation of evolution of behavior (Gunji) and Chapter 8 provides experiments that demonstrate such a paradox (Kitabayashi). Chapter 9 argues that meaning exists only in brains, and that dynamical description of the brain can lead to semantic machines (Freeman). Chapter 10 describes a general definition of life with consciousness from the viewpoint of cognitive science (Bickhard) and Chapter 11 discusses the limits of functionalism and connectionism from a philosophical view of intentionality (Basti). Two guests wrote the last two chapters, and the other nine authors were speakers at the Conference Iizuka'98.

The editor of this volume has had valuable help from my colleagues in editing this volume. I must thank in particular Dr. Ken'ichi Asami and Mr. Toru Tokuyama for their enthusiastic cooperation in completing the camera-ready version of all the papers in this volume in spite of several changes of file styles. The Editor is also very thankful to Nissan Science Foundation and Electro-Mechanic Technology Advancing Foundation for Supporting the Sessions the Editor organized for this volume at the Conference Iizuka'98.

Iizuka
Tadashi Kitamura
Volume Editor
November 1, 1999

This page is intentionally left blank

Contents

Series Editor's Preface v

Volume Editor's Preface vii

Chapter 1 Consideration of Emotion Model and Primitive Language of Robots 1
Tetsuya Ogata and Shigeki Sugano

Chapter 2 An Architecture for Animal-like Behavior Selection 23
Tadashi Kitamura

Chapter 3 A Computational Literary Theory: The Ultimate Products of the Brain/Mind Machine 43
Akifumi Tokosumi

Chapter 4 Cooperation between Neural Networks within the Brain 53
Michel Dufossé, Author Kaladjian, and Halim Djennane

Chapter 5 Brain-like Functions in Evolving Connectionist Systems for On-line, Knowledge-Based Learning 77
Nikola Kasabov

Chapter 6 Interrelationships, Communication, Semiotics, and Artificial Consciousness 115
Horia-Nicolai L. Teodorescu

Chapter 7 Time Emerges from Incomplete Clock, Based on Internal Measurement 149
Yukio-Pegio Gunji, Hideki Higashi, and Yasuhiro Takachi

Chapter 8 The Logical Jump in Shell Changing in Hermit Crab and Tool Experiment in the Ants 183
Nobuhide Kitabayashi, Yoshiyuki Kusunoki, and Yukio-Pegio Gunji

Chapter 9 The Neurobiology of Semantics: How Can Machines be Designed to Have Meanings? 207
Walter J. Freeman

Chapter 10 The Emergence of Contentful Experience 217
Mark H. Bickhard

Chapter 11 Intentionality and Foundations of Logic: A New Approach to Neurocomputation 239
Gianfranco Basti

About the Authors 289

Keyword Index 303

Chapter 1

Consideration of Emotion Model and Primitive Language of Robots

Tetsuya Ogata, and Shigeki Sugano
Waseda University

Abstract

This study discusses the communication between autonomous robots and humans through the development of a robot that has an emotion model. The model refers to the internal secretion system of humans and it has four kinds of the hormone parameters to use to adjust various internal conditions such as motor output, cooling fan output and sensor gain. As the result of the experiments, the hormone parameters enabled the robot to adjust its conditions like homeostasis in humans and generate the primitive emotional behaviors. In this paper, human's mental images and language are given consideration as a method for emotional expression. The hypothesis model for the acquisition of the internal expressions of robots and the experimental results using a real autonomous robot are described.

Keywords : emotion, internal secretion system, self-preservation, intelligence, affordance, behavior, animal language, judgment, desire, communication, autonomous robot, wamoeba, fuzzy, neural network, self-organizing, behavior based AI

1.1 Introduction

In recent years, interactive simulation games have become very popular. Human beings enjoy communication with machinery, which is not controlled by operators but behaves autonomously in ways, which the observers cannot anticipate. These machines will be successful in the future.

Robot hardware and virtual agents have been developed to date using the above considerations. Bates developed a virtual agent that mimics the behavior of living things using animation [1]. SONY proposed "Robot Entertainment,"and developed a pet-robot has with 4 legs [2]. Moreover, there is some research which has applied robot behavior to human-machine interface. F.Hara proposed "Active Human Interface (AHI)," and made a robot, which makes the facial expressions to indicate the conditions at the machinery [3].

In human communications, the intentions of behaviors are usually exchanged by language. Much research on language communication between humans and machinery (computer and/or robot) can be done from the early AI works. It is extremely powerful as a tool for a human machine communications.

However, most machines in this research have only the "Model based intelligence", given to them by the designer a priori. In this frame, the following faults can be pointed out.

1) Humans who communicate with the machine become tired, because the patterns of speech communication of the robot are limited.
2) It is difficult to establish communication in particular situations, which the designer did not predict.

The other hand, R. Brooks proposed the Subsumption Architecture (SA) [4]. In this architecture, the parallel controls are emphasized, and it has already been proven to be successful for mobile robot control. However, because SA has no explicit expression of behaviors and the planning etc., it is difficult not only to generate the behavior process, but also to explain the purpose of the behavior (intention) to the observer. Therefore, in recent years, much attention has been given to the some research of on robot behavior learning and behavior emergence based on the concept of an "Embodiment." [5][6]

Also, it seems that the model (symbol) which the robot acquired by these techniques can be applied to not only behavior generation but also communications with the coexisting human. It might be important to consider the interpretation of the model of the robot for human robot communications.

We aim to realize robot autonomous intelligence for the human cooperation by proposing the "emotion model" of robots, which refers to the human internal secretion system. [7] To date, we have developed the autonomous robot, WAMOEBA-2 (Waseda Amoeba, Waseda Artificial Mind On Emotion BAse). [12]

This paper describes the algorithm, which generates the internal expressions, which are the origin of the language of WAMOEBA-2.

1.2 Acquisition Algorithm of World Model of Robots

1.2.1 *Affordance Theory*

The researches, which consider the methods for generation of the internal model in robots often, discuss the theory of the ecological optics of J.J.Gibson. [8] Gibson insists that the sense organs of the living system can extract the

meanings of environmental information directly. It can be said that "the meaning of the environmental information" would be the evaluation for "self-preservation." [9] Gibson called such environmental information "affordance," which supports the agent's behaviors. The agent extracts the affordance through physical interactions with the environment. As a result, the agent comes to acquire the world model, that is, the relationships between senses and behaviors. Moreover, the agent comes to have a "common recognition" with other agents existing under the same environment, i.e. "emergence of communications." The affordance theory places too much value on the information gained from the senses, however, it can essentially avoid the "symbol grounding problem" as Harnad has discussed [10].

1.2.2 *Self-Preservation Evaluation Function*

In the case of a robotic system, the affordance would be defined by the task of the robot. The model-based intelligence, which is designed by humans, is not always enough to execute the task. R. Pfiefer noted that human designers often make the internal expression of robots too complex. (Over Design Problem) [11] In order to avoid this problem, there is some research, which tries to make robots acquire an original internal expression by using learning algorithms. [5][6] In these researches, the robot divides the sensor / motor space by statistical methods etc. and squeezes them by the reinforcement signals according to the tasks. Then, the acquired model of the robot usually depends on the form of the reinforcement signals given by the human designer.

We aim to achieve emotional behavior in robots, and have developed a robot which has one evaluation criteria of "self-preservation" which is most primitive creature instinct for the emergence of intelligence. [9] Because self-preservation is an abstract concept, it is difficult for robots to be introduced to it as concrete task. However we proposed a fuzzy set membership function named "evaluation function of self-preservation" to evaluate the self-preservation of robots from the point of view of the durability and safety of the robot. [7]

This function is one kind of fuzzy set membership function, which converts sensor input into the evaluation values of durability (breakdown rate) of robot hardware between -1 to 1. The shapes of these functions are chosen depending on the basic hardware specs. For example, the evaluation function of the voltage of the battery is shown in Fig. 1.1. Because the standard voltage of the battery indicates the best condition for the robot, the output of the self-

preservation evaluation function is set 0. If the voltage decreases, the evaluation function outputs the minus value, and if the voltage increases, the function outputs plus value. Using this function, all sensor information can be interpreted as reinforced signals from the viewpoint of self-preservation (hardware durability).

It can be said that the sensor space transformed by the evaluation functions would connect the behaviors of the robot directly. Based on the above consideration, we considered the human-robot communication as follows.

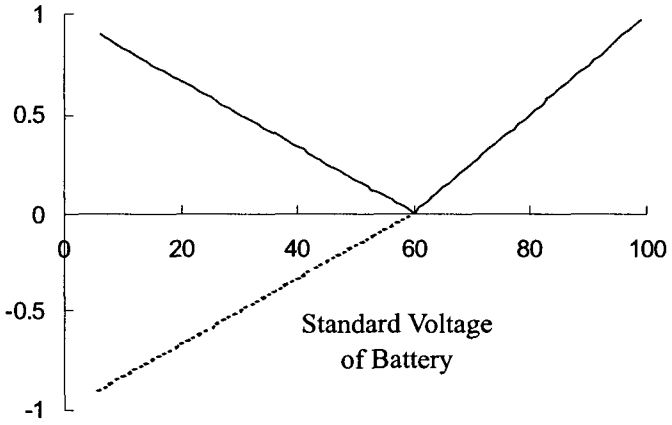


Fig. 1.1 Evaluation Function for Self-Preservation (Ex. Voltage of Battery).

1.3 Primitive Language

Human "language" originates from the logic and the theories represented by Chomsky's theory [12] etc. However, we treated the language of the robot as a simple tool for communication, which can express the internal model, acquired by a learning algorithm. The language need not apply to strict language logic.

In psychology, there are some communications, which are called "Primitive Language." For example, some animal can speak simple words, and communicate with humans. [13] Koko is an adult female gorilla who has learned some 2000 words of American Sign Language. And then there's Kanzi. Kanzi is a bonobo, or pygmy chimp, from equatorial Africa whose language abilities are

said to compare favorably with those of a two-year-old human child. Further, Pepperberg trained an African parrot named Alex to use some forty English words. Moreover, it is thought that "Holophrastic Speech" of infants is similar to these expressions.

There are some opinions that these simple expressions have no strict grammar or rules, and that they are not languages. However, in the above reports, animals and infants developed rich communications involving emotional information. If these primitive expressions were applied to robot systems, they would become effective interface tools. The properties of the primitive language are shown as follows.

1) There are no particles (lack of grammatical system), and most of them consist of "noun" "verb" and "adjective."
2) Most of the expressions mean "desire."

From the above features, it is thought that the robot can have primitive language using the clustering techniques of the sensor space.

1.4 Autonomous Robot: WAMOEBA-2

It is effective for the robot to have many kinds of sensors not only to generate behaviors, but also to communicate with humans. However, if the structure and the evaluation criteria of the robot were extremely different from those of humans, it would be difficult to establish the communication between the robot and humans.

Based on the above considerations, we have developed the autonomous robot named WAMOEBA-2 to realize the emotional communication with humans. WAMOEBA-2, shown in Fig. 1.2 and Fig. 1.3 is a completely independent robot, which contains its own batteries and control systems in its body. [14] The dimensions are 983 (L) x 862 (W) x 1093 (H) [mm], and the weight is around 100 [kg]. The characteristics of the robot hardware are 1)-communication functions, 2)-model of the internal secretion system, and 3)-behavior generation algorithm.

1.4.1 *Communication Functions of WAMOEBA-2*

From the perspective of being "human friendly," the arrangement of the sensors and the motors refers to the morphologies of the creatures represented by human beings. WAMOEBA-2 has two arms for emotion expressions using gestures. In addition, it has two LCDs on the head and the chest to indicate internal

conditions. It is important for WAMOEBA-2 to detect visual, sound, and tactile information, because human beings can generate this information directly. WAMOEBA-2, therefore, has various sensors, e.g. four ultrasonic range sensors, two color CCD cameras, two microphones (right and left), two touch panels (the head and the chest), and eight tactile sensors on the vehicle.

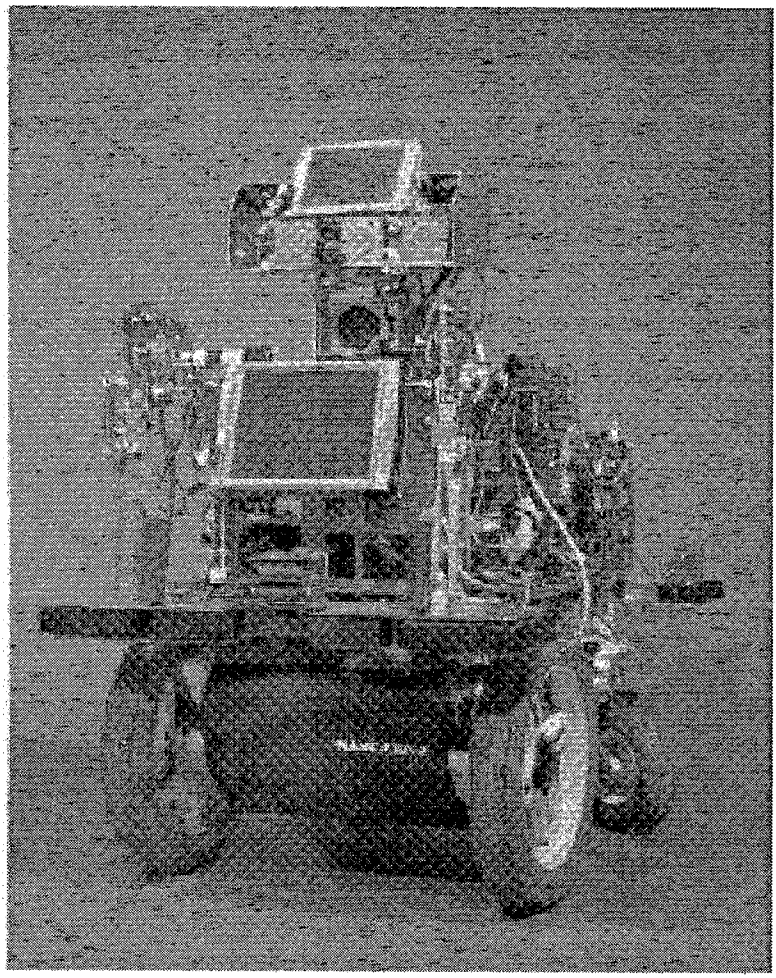


Fig. 1.2 Autonomous Robot: WAMOEBA-2 (Photograph).

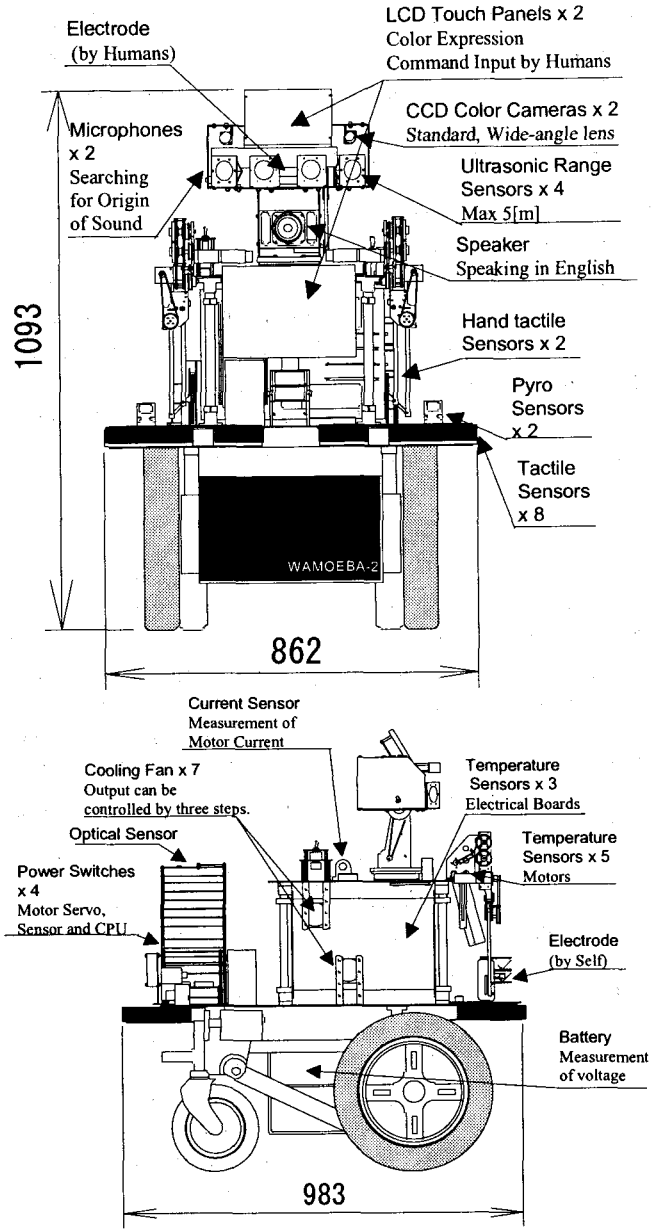


Fig. 1.3 Configuration of WAMOEBA-2 Functions.

The motor chair is adapted to the vehicle part, so that WAMOEBA-2 can acquire a wide activity area, so it does not have to stay indoors. WAMOEBA-2 can drive for about 30 [min.] using the battery in the motor chair.

1.4.2 *Model of the Internal Secretion System*

The original characteristic of WAMOEBA-2 is internal mechanism architecture for modeling the internal secretion system of humans. The internal secretion system controls the entire state of the living organism using hormones. It is thought that, for robot hardware, these organisms correspond to the control mechanisms of electric power consumption and circuit temperature, etc. Table 1.1 shows the results of the consideration of correspondences between human's autonomic nervous system and the hardware mechanisms. Based on this assumption, we constructed the original hardware architecture in WAMOEBA-2 described as follows.

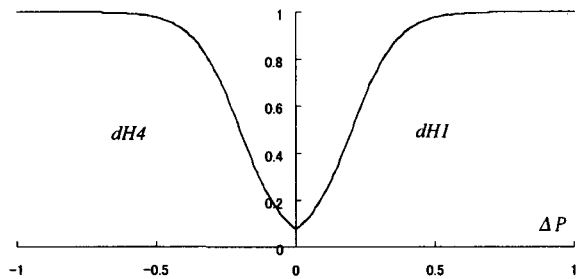
Influence of Autonomic Nervous System in Human	Mechanical System		
	Influenced Part	Input Information	Adjustor
Heart Beat Sugar Density in Blood	Actuator Battery	Torque Sensor Voltage, Current Sensor	Actuator Output
Gastrointestinal Activity	Battery	Battery Load (Fruid Level Sensor)	Charging Current
Sweat, Cowlick Musclar tiredness	CPU, Electiric Curcuit, Actuator	Temperature Sensor	Cooling Fan
Arousal	Program Cycle Speed	Data Processing Times	Occupaton Memory
Pupillary Light Reflex	Camera	Optical Sensor	Squeezing
Excretion	Structural Part Electric Curcuit	Rust and/or Dirt	-
Self-Restoration of Organization	Wiring	Test for continuity (Tester, Voltmetor)	-
	Structural member	Deformations (Strain Gage etc.)	-
	Sensors	Referrence to Input Information	-

Table 1.1 The Correspondence of the Internal Secretion System and Hardware.

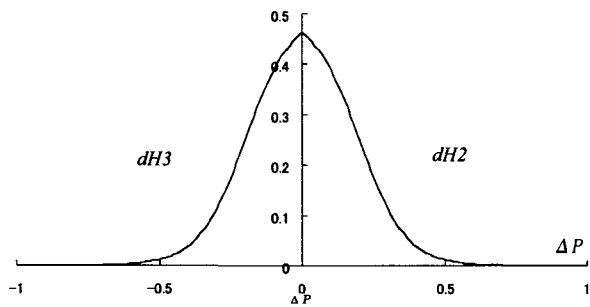
WAMOEBA-2 receives the battery voltage and the driving current. Moreover, using temperature sensor IC, it can acquire eight positions of temperature, which are the motors (the head, the neck, the shoulder, the elbow, and the motor chair) and the circuits (the image processing board and A/D boards etc.) It can control the output of the cooling fans, and the switch the power supply of each motor on or off by itself.

In order to control the internal hardwares, WAMOEBA-2 calculates the output of the hormone parameters using total value P of all self-preservation evaluation functions Ei described in section 1.2.2 as follows;

$$\frac{dH_i}{dt} = \alpha_i \cdot \left(P - P^{th}\right) + \beta \cdot \sigma_i\left(\frac{dP}{dt}\right) + \gamma \cdot \left(H_i - H^{th}\right) \qquad (1)$$



The differential of the value of the evaluation functions



The differential of the value of the evaluation functions

Fig. 1.4 Four-Hormone Parameter in WAMOEBA-2.

where , α_i, β and γ are coefficients that correspond to the potential, the change quantity, and the stabilization. α_i represents if the hormone output is continuous. P^{th} and H^{th} represent the standard values about P and H. $\sigma(x)$ is the sigmoid function, which suppresses dP/dt within the range of 0-1. There are four kinds of the hormone parameters [*H1* to *H4*] corresponding to four conditions: if the evaluation value P is positive or negative (mood) and if P changes dynamically or not (arousal), using four sigmoid functions shown in Fig. 1.4.

These hormone parameters affect many hardware conditions such as sensor gains, the motor outputs, the temperatures of the circuits, and energy consumption in parallel. The affects of each hormone are decided referring to the physiology [15] shown in Table 1.2. Table 1.3 shows examples of the correspondences between the morphologies of the emotional expressions caused by the hormone parameters. However, these are not fixed but are changed by the mixture condition of the four hormone parameters.

		H1	H2	H3	H4
Actuator Output		Up	Down	Down	Up
Cooling Fan Output		Down	Up	Up	Down
CCD Camera Viewing Angle		Decrease	Increase	Increase	Decrease
Ultrasonic Sensors Sensing Area		Narrow	Wide	Wide	Narrow
Sound	Volume	Up	Down	Down	Up
	Speed	Up	Down	Down	Up
	Loudness	Down	Down	Up	Up
LCD Color		Red	Blue	Yellow	

Table 1.2 Affects of the Hormone Parameters of WAMOEBA-2.

Radical Unpleasantness	cause	Bumper switches, Ultra-sonic range sensors (radical approach)
	expression condition	Decrease of the viewing angle, Increase of the motor output, Red color expression on the head LCD and Low voice
Unpleasantness	cause	Temperature of the motors and the electrical circuits, Ultra-sonic range sensors
	expression condition	Increase of the viewing angle, Decrease of the motor output, Blue color expresson on the head LCD
Pleasantness	cause	Charge
	expression condition	Decrease of the viewing angle, Decrease of the motor output, Yellow color expression on the head LCD and High voice

Table 1.3 Outline of the Affects of a Hormone Model.

1.4.3 *Motor Agent*

Next, the methodology by which WAMOEBA-2 generates its behavior for emotional communication should be discussed. We considered some algorithms of robot behavior as follows.

A conventional model-based robot behaves based on the environmental model given a priori. It requires accurate sensor input, an optimal environment, and a large amount of calculation. R.Brooks proposed a "behavior-based approach."

However, there is a limitation on the variety of behavior, because there are only the combinations of each behavior module, which are fixed a priori. In communication, humans can easily forecast robot behavior through the experiments, and would then become tired. It is an extremely difficult problem to design a behavior module for communication with humans. We thought that the behavior should be described not at the level of the "task" but at "motor activity" in order to generate the diversity of the behavior. R.Brooks has developed a humanoid robot, "Cog," which moves its arms based on oscillators in the motors. [16]

We proposed a "motor agent" as the behavior generation mechanism of WAMOEBA-2. In the motor agent algorithm, each motor acquires all sensor information and other motor drive conditions through the network in the robot hardware. Based on this information, each motor decides its actions autonomously. Motion command M_i of motor i is calculated as follows:

$$a_i = \sum_p w_{ip}^s S_p + \sum_{j \neq i} w_{ij}^m M_j \tag{2}$$

Here, the input value of sensor p is defined as S_p, the output of motor j is M_j, and the activity of motor i is a_i. The commands for motor i are generated using the absolute value and the positive and negative values of ai. In this architecture, the morphology of the behaviors depends on weight value w, in which descriptions are not explicit. The initial value of w depends on the physical arrangement of the motors and the sensor; i.e., w is a large value when the distance between the sensors and the motors is small. In this stage, a designer who observes the behaviors of WAMOEBA-2 adjusts w. Table 1.4 shows some connections, which have a large w value between the sensors and the motors.

Motor Part	Sensor Input
Head (2 DOF)	Vision (Moving Area, Color Area), Sound, Ultra-sonic Range sensors, Bumper switches
Shoulder (x 2)	Degree of Head Motor, Hand touch sensors
Elbow (x 2)	Degree of Shoulder Motor, Hand touch sensors
Vehicle (2 DOF)	Degree of Neck Motor, Vision, Sound, Ultra-sonic Range sensors, Bumper switches, Hand touch sensors

Table 1.4 Outline of "Motor Agent" in WAMOEBA-2.

Based on only implicit expressions, the "motor agent," WAMOEBA-2 generates the behavior using the whole body, e.g. imitation of the movement area, the sound origin, and avoidance behavior, etc. Although the w value is fixed a

priori, the hormone parameters affect all motor activity: a_i. That is, hormone parameters *H1* and *H4* (which cause the exciting conditions in WAMOEBA-2) increase the ai, and, on the other hand, *H2* and *H3* (which cause the calm conditions) decrease a_i. As the results, the forms of the behavior change as follows: for example, when WAMOEBA-2 is quiet due to the low battery etc., it imitates moving object by the head only. In the exciting conditions however, it follows the object using the arms and/or the vehicle. The behaviors based on motor agent are expected to surprise observers. Fig. 1.5 shows the structure of the hardware and software of WAMOEBA-2.

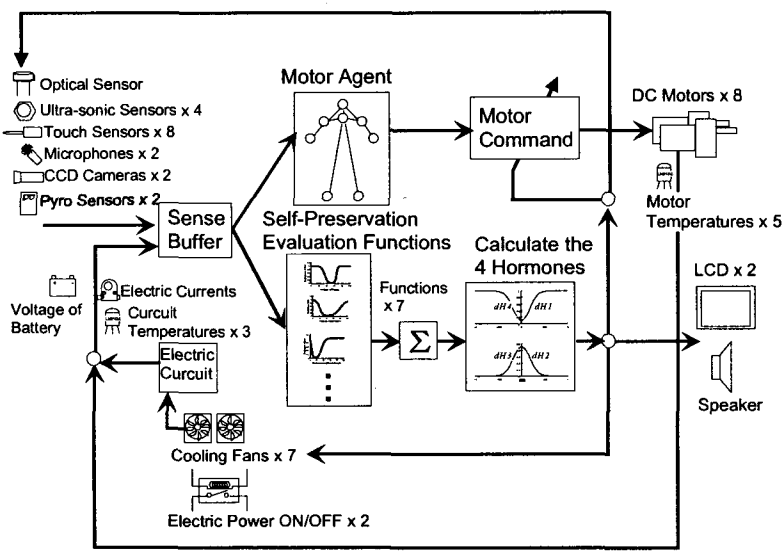


Fig. 1.5 System Structure of WAMOEBA-2.

1.5 Communication of WAMOEBA-2

Humans can communicate with WAMOEBA-2 by hand waving, clapping, calling, touching the tactile sensors etc. WAMOEBA-2 makes various reactions such as approaching, escaping, making sounds, eye tracking, and arm stretching. The motor agent generates these actions. In addition, WAMOEBA-2 changes its motion speed, volume/speed/loudness of sounds, and color output on LCD

by hormone parameters.

Most conventional emotion models have a limited ability to communicate with humans. Usually, a human being observes and judges the expressions of the emotion model, and the recognition rate is the evaluation of the model. In communication between WAMOEBA-2 and humans, there is no scenario like this. The psychological impressions at humans change dynamically according to the behavior of the robot and/or the humans. The characteristics of WAMOEBA-2 communication are as follows.

1) Adaptability in real world: Since WAMOEBA-2 is an independent and behavior based robot, it is not necessary to standardize its environment. Moreover, there is no limitation for humans in the standing position and/or motion, etc.
2) Diversity of the ways to communicate: Human beings can communicate without special interface tools. Moreover, neither "words" nor "gestures" etc. for communication are specified, and preliminary knowledge is not needed.
3) Development of communication: Communication is developed according to the behavior of humans and WAMOEBA-2 in real-time. There is no "story" and/or "scenario" set beforehand by a designer.

It is believed that the "freedom degree" mentioned above (where humans are not restrained in communication with robots) is an important element in order to realize robot-human emotional communication.

1.6 Model Acquisition Algorithm of WAMOEBA-2

WAMOEBA-2 converts the sensor information using the self-preservation evaluation functions, and categorizes them by the Kohonen self-organizing map (SOM) [17] showed in Fig. 1.6. At first, the output of each self-preservation function is input to the sensory layer (26 neurons) in each sensory modality. Further the sensory layer is Hopfield NN, and the relationships between all of the sensors can be memorized in this layer. Using SOM algorithm, the data patterns in the sensory layer are mapped into the cognitive layer (100 neurons).

The sensory layer is divided into four parts: the internal area, the visual area, the auditory area, and the somatic area. For example, the internal inputs are the energy consumption and the voltage of battery etc. and the visual inputs are the camera information (moving area, color area etc.) the brightness sensor output, the range sensor output etc.

The information acquired in the neurons of the cognitive layer involves not

only external information, but also the internal conditions and the behaviors. Moreover, since the sense patterns consist of the sensor data only, each cognitive layer neuron can be interpreted to the "symbol" of WAMOEBA-2.

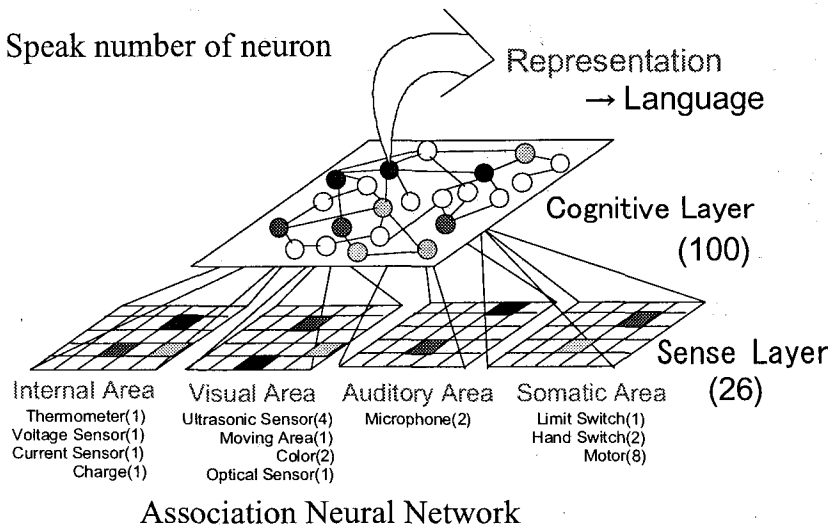


Fig. 1.6 Kohonen Neural Network in WAMOEBA-2.

To investigate the morphology of the symbol of WAMOEBA-2, we performed experiments. The experimental environment was set in an indoor room with a size of 6.0 (L) x 7.4(W) [m]. There are two types of experiments, as follows.

a) Environment A: There are no people and WAMOEBA-2 moves alone.
b) Environment B: There are three persons who give the stimulus on WAMOEBA-2.

Fig. 1.7 shows the progress of the self-organizing of Kohonen NN in WAMOEBA-2. "Numbers of neurons" is the numbers of necessary neurons out of a hundred in the cognitive layer for WAMOEBA-2 to recognize 80% of sensory patterns in each experiment period (3 min.).

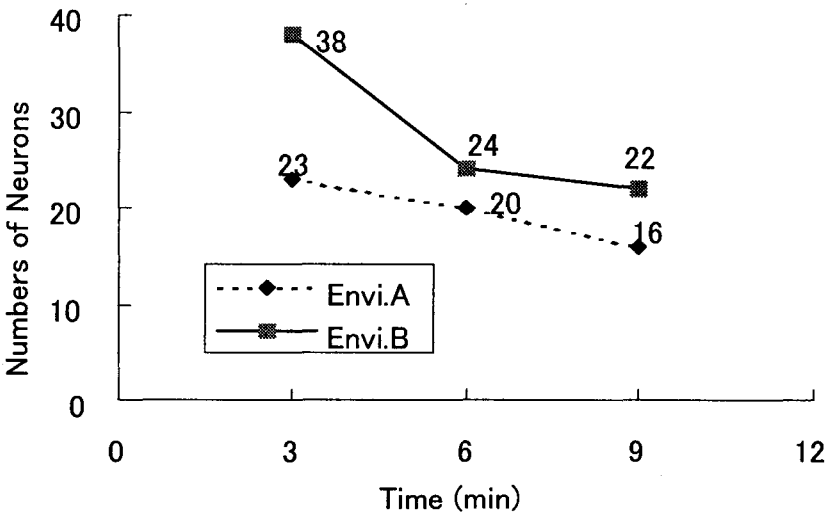


Fig. 1.7 Self-Organizing of Kohonen NN in WAMOEBA-2.

It can be confirmed that the number of neurons settles as the self-organizing proceeds. It is thought that the complexity of each environment reflects the number of neurons.

The acquired symbols could be classified into two types. One type is fixed patterns in the early stage of the experiment, and the other type is changing patterns according to the environmental change. These symbols can be applied to not only human-machine communication, but also to the behavior planning of the robot. [18] The concrete examples of the clustering sensor patterns are shown as follows.

a) Charge + Ultrasonic sensor (center): A human charges WAMOEBA-2 from the front.
b) Ultrasonic sensor (left) + Hand switch (left) + Left Rotation: A human touches the left-hand switch and WAMOEBA-2 evaded it.
c) Moving area + Move forward + Low voltage of the battery: A human waves a hand in front of the WAMOEBA-2, and the robot chases it.

The symbols shown in this paper have similar characteristics of the "subsymbol" which was proposed by connectionism in the 1980's. For example, K. Nakano proposed the model where two robots acquired concepts such as "catch" and "foreign enemy" from the combination of the "micro features"

such as "four legs" and "hair" etc. [19]

However, in the case of WAMOEBA-2, the micro features, which compose the symbol, are not designed by human's subjectivity. They depend on the sensors and the motor outputs equipped on WAMOEBA-2.

1.7 Diversification of Expression

From the view of human-machine interface, the expressions generated by the proposed algorithm are simple reports of the conditions of the robot hardware. In order to give variety to the speech of WAMOEBA-2, we considered different kinds of animal expressions. Concretely, the following two expressions were introduced.

(a) Judgment
(b) Desire

The concrete algorithms are explained in detail as follows.

1.7.1 *Judgment*

From external sense information which shows external objects such as color and sound, etc., the internal conditions such as charge and consumption of electric power, etc. are associated using the Hopfield NN. This is an utterance, which expresses the influences which external objects have on WAMOEBA-2. It can be interpreted as the judgment of the meaning of the external object by WAMOEBA-2. The algorithm is shown in Fig. 1.8.

1) In SOM sense layer, the parts involved in internal information (the voltage of the battery, the motor current, and the temperature of the circuits and the motors) are cleared.
2) Internal information is associated from external sensor information (the audiovisual information and motor drive commands, etc.) by using the Hopfield NN. Then, all internal sense information connecting to the external information is associated. The entire sense pattern is a vague expression, which does not exist in the real world.
3) The present entire sense pattern is recognized by SOM, and the recognition neuron is detected. This recognition neuron indicates the sensor pattern, which exists in the reality world, and is the nearest the associated sense pattern.

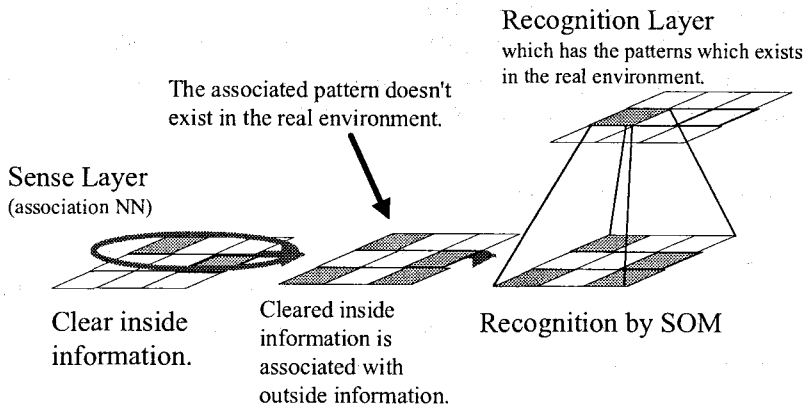


Fig. 1.8 Algorithm of Judgment of WAMOEBA-2.

1.7.2 *Desire*

From the internal conditions which have negative output of the self-preservation evaluation function, the external sensor patterns (for example, the combination of the moving area and the ultrasonic range sensors which detects human beings) are associated. It can be interpreted that the external sensor pattern expresses the desire object of WAMOEBA-2. The algorithm is shown in Fig. 1.9.

1) The external information part in the sense layer of the SOM is cleared.
2) The neuron values of the internal information parts are reversed of plus / minus, and the external information is associated by using the Hopfield NN. Each sensor value is converted within the range of -1 to 1 by the self-preservation evaluation function. For example, if the voltage of the battery is low, the external information, which relates to the high voltage of the battery, can be extracted by the combination of the reversing operations and the association.
3) The present entire sense pattern is recognized by SOM, and the recognition neuron is detected.

In the experiments, we set WAMOEBA-2 to utter the neuron numbers of the cognition layer of SOM. The conditions of the utterance are set as follows.

i) When the excitement degree of the recognition neuron exceeds the threshold.

ii) When the total value of the self-preservation evaluation function exceeds the threshold.

WAMOEBA-2 diversified and stabilized the uttered neuron numbers through interaction with the environment. It was confirmed that several testees were able to understand the utterance expressing the demand ("Charge" and "Get out of the way" etc.) of WAMOEBA-2 through the interaction.

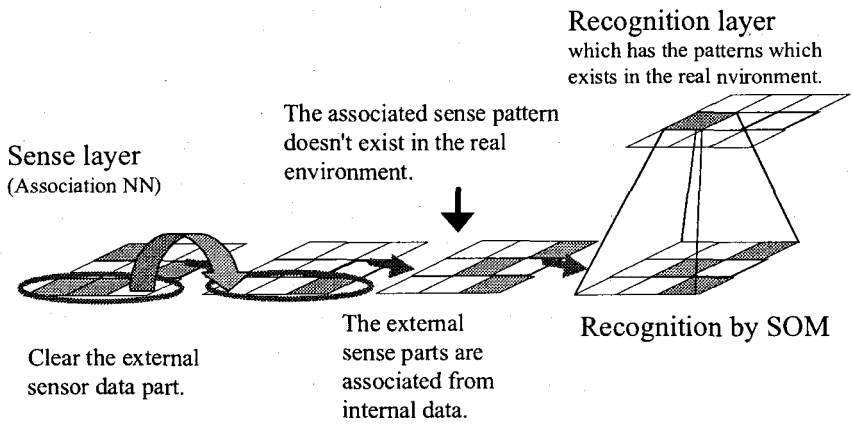


Fig. 1.9 Algorithm of Desire Generation of WAMOEBA-2.

1.8 Evaluation Experiment

Experiments concerning "classical conditioning" were performed to confirm the function of introduced algorithm in WAMOEBA-2.

One testee made WAMOEBA-2 acquire the internal model in the SOM through physical interaction with the robot for about 10-min. In this experiment, when the testee made WAMOEBA-2 to charge, he gave the input on the left microphone of WAMOEBA-2 without fail.

After the experiment, the sound was input to the left microphone, WAMOEBA-2 utters the neuron number 33. The sense pattern of the neuron 33 is as follows.

1) The neck turns to the left side.
2) The center of the ultrasonic range sensor detects an object.
3) Charge

This sensor pattern shows a situation where human makes WAMOEBA-2 charge and can be interpreted as a judgment result of the sound.

Furthermore, when the voltage of the battery decreases, WAMOEBA-2 expresses the neuron number 33. This can be interpreted as a desire.

1.9 Conclusion and Further Perspectives

This study examined the emotional communication between an autonomous robot and human beings. The autonomous robot WAMOEBA-2 has been developed by implementing the emotion model by referring to the internal secretion system, and the motor agent, which generates various behaviors, based on implicit descriptions in the network. WAMOEBA-2 acquires the internal information such as the voltage of the battery, the consumption current, and the circuit temperature etc. Moreover, WAMOEBA-2 has four kinds of hormone parameters, which control the internal conditions by using the cooling fans and electrical switches. These parameters are calculated from the original algorithm; "the evaluation function for self-preservation." The parameters influence various parts, such as motor outputs and sensor gains etc. As the results of these functions, WAMOEBA-2 can adjust the internal conditions such as the motor current and the temperature of the circuit.

This paper focused on "language" as an expression technique of the internal condition of robots, and proposed the speech expression method of the robot WAMOEBA-2, referring to the infant's "holophrastic speech" in psychology.

If robots come to understand the environment and select their behaviors based on original evaluation criteria, how humans interpret the internal expressions which robots acquire by the learning techniques is important. The communication between autonomous robots and human beings discussed here is a new concept, which is different from conventional robot-human communication. However, if autonomous robots will be used in homes and hospitals, etc. then the communication discussed here will be indispensable. This kind of the communication has many problems, e.g. methods to maintain human-friendliness and the human empathy to robots etc.

In the future, through more experiments with WAMOEBA-2, the expression acquired in the recognition layer should be investigated. Furthermore, it is interesting to consider the methodology, which translates the internal expression to the natural language.

References

[1] Bates, J., The Role of Emotion in Belivable Agents, Communications of the ACM, pp.122-125, (1994).

[2] Fujita, M. and Kageyama, K., Robot Entertainment, in Proc. of the 6th Sony Research Forum, pp.234-239, (1996).

[3] Hara, F. and Kobayashi, H., Computer graphics for expressing robot artificial emotions, in Proc. of IEEE Int. Workshop on Robot and Human Communication, pp.155-160, (1992).

[4] Brooks, R.A., Robust layered control system for a mobile robot, IEEE Journal of Robotics and Automation, RA-2, pp.14-23, (1986).

[5] Asada, M., Noda, S., and Hosoda, K., Non-Physical Intervention in Robot Learning Based on LfE Method, In Proc. of Machine Learning Conference Workshop on Learning from Examples vs. Programming by Demonstration, pp.25-31, (1995).

[6] Tani, J., Model-Based Learning for Mobile Robot Navigation from the Dynamical System Perspective, IEEE Trans. System, Man and Cybernetics Part B, Special issue on robot learning, vol.26, no.3, (1996).

[7] Sugano, S. and Ogata, T., Emergence of Mind in Robots for Human Interface - Research Methodology and Robot Model-, in Proc. of IEEE Int. Conf. on Robotics and Automation, pp.1191-1198, (1996).

[8] Gibson, J.J., The Ecological Approach to Visual Perception, Houghton Mifflin, (1979).

[9] Kato, I., Homini-Robotism, in Proc of Int. Conf. on Advanced Robotics (ICAR '91), pp.1-5, (1991).

[10]Harnad, S., The symbol grounding problem, Physica D, vol.42, pp.335-346, (1990).

[11]Pfeifer, R., Cheap Designs, exploiting the dynamics of the system-environment interaction, Proc. of the Bielefeld conference on Prerational Intelligence, (1991).

[12]Chomsky, N., Generative grammar. Its basis, development and prospects. SELL special issue, Kyoto University of Foreign Studies, (1987).

[13]Griffin, R., Animal Thinking, Cambridge, Mass: Harvard University Press, (1984).

[14]Ogata, T. and Sugano, S., Mechanical System for Autonomic Nervous System in Robots, IEEE/ASME Int. Conf. on Advanced Intelligent Mechatronics (AIM '97), Paper No.113, (1997).

[15]Nieuwenhuys, R., Voogd, J., and Chr.Huijzen, The Human Central Nervous System-A Synosis and Atlas, Springer-Verlag, (1988).

[16]Brooks, R.A., Breazeal, C., Irie, R., Kemp, C.C., Marjanovic, M., Scassellati, B. and Williamson, M.M., Alternative Essenses of Intelligence, American Association for Artificial Intelligence (AAAI), (1998).

[17]Kohonen, T., Self-Organization and Associative Memory, Springer, Berlin, London, (1988).

[18]Ogata, T., Hayashi, K., Kitagishi, I. and, Sugano, S., Generation of Behavior Automaton on Neural Network, in Proc. of IEEE-RSJ Int. Conf. on Intelligent Robot and Systems (IROS'97), pp.608-613, (1997).

[19]Nakano, K., Isotani, R., and Ohmori, T., Self-Organizing System Obtaining Ability of Communication -Primitive Model for Language Generation-, in the Transactions of the Institute of Electronics, Information and Communication Engineers, Vol.J70-A, No.5, pp.806-815, (1987) (Japanese).

Chapter 2

An Architecture for Animal-like Behavior Selection

Tadashi Kitamura
Kyushu Institute of Technology

Abstract

CBA is a six-layered architecture of consciousness linked to behavior such as reflex action, detour, and ambush. Two emotion-valued criteria are given for behavior selection. While a level of behavior is chosen to maximize the consciousness intensity, an action at the level chosen is selected to increase the pleasure. CBA is efficient for behavior selection because performing a complex task elevates the performer's level of consciousness. Inhibition of behavior triggers an elevation of the level of consciousness and behavior. The model design of detour and ambush was tested using two small mobile robots that had a limited temporal and spatial information of their environments. Emotion-valued criteria for behavior selection explain the meaning of behavior obstruction.

Keywords : robotics, behavior, consciousness, CBA, emotion, behavior selection, representation, symbol, embodiment, AI, soft AI

2.1 Introduction

The design of artificially intelligent systems including the traditional artificial intelligence (AI) and soft AI, is based on the assumption that human intelligence is structurally closed and autonomous. This assumption may be true, except when intelligence is interrupted by an environmental input that changes the context for solving the problem the body has encountered. Such a situation occurs when an organ senses an obstruction, and an action is stopped by it. This premise, at the same time, may seem to be supported by the fact that behavior is relatively independent of the intelligence, and the organs supporting the behavior are automatic. But in our daily life, the autonomy of intelligence is almost always interrupted, i.e, we are almost always exposed to interaction and cooperation between behavior and intelligence in order to discover a logical scheme to solve a behavioral problem. We often even find that intelligence

indirectly helps solve a physical organ's crisis in order for the whole body to survive, such as searching for food and calling an ambulance.

On the other hand, the traditional AI and soft AI do not take any action to solve a problem their closed knowledge system does not cover because the knowledge system has no knowledge about such an action. It is widely recognized that the traditional AI and soft AI, such as production system, never comes to the final solution of a problem if it is a Frame Problem. A Frame Problem is one which is composed of multiple interlocking problems, and one in which a subsidiary problem must be solved in order to solve the initial problem. A human also encounters this sort of problem when he/she lacks default knowledge of a given problem. But in that case, the human intelligence stops trying to solve the problem and jumps out of the context of the Frame Problem by taking an action such as asking what he/she should do, or even by ignoring the problem. This difference between the human intelligence and AI, traditional or soft, can be recognized as due to the human embodiment while AI itself does not have a body to take an action with. This means that the body is ready to act when the intelligence process stops and vice versa. But even a robot having a body to act with may have no embodiment if the interaction between body and intelligence is embedded as such.

In contrast to AI and soft AI, Subsumption Architecture (SSA), a behavior-based architecture for a mobile robot proposed by Brooks [5], employs neither high-level, centrally goal-oriented, nor symbolic algorithms but embeds several fixed reactive behavior modules almost independent of each other in a robot. The idea of this architecture brings into focus what is embodied while ignoring what is behind the embodiment. This architecture expects the emergence of meaningful behavior from simultaneous execution of these modules. This architecture has been called a non-Cartesian machine in the sense that it has no central program like a human ego or consciousness to control the robot. This architecture succeeds in generating a simple goal-oriented performance of robots to some extent [8], but it has not yet been reported to play a game of chess.

In order to embed into a robot the interaction between intelligence and behavior, a consciousness seems necessary to mediate between the two extremities; a logic system, rigid or soft, without behavior and behavior without representation. It may not be possible to give a precise, quantitative definition of "consciousness". However, consciousness is what subjectively activates behavior and is objectively visible through behavior. What is a more important aspect of consciousness is that it is aroused to recognize the meaning of the obstruction when behavior is obstructed. While a human recognizes the mean-

ing of behavior obstruction as the cause of obstruction, its meaning to a lower level animal is given as an emotional value, positive or negative. Thus consciousness is linked to behavior in the sense that consciousness embodies the meaning of obstructed behavior memorized in the body. This idea is included in the concept of consciousness by Tran. He proposed a conceptual model of the hierarchical relationship between consciousness and behavior based on a currently accepted theory of animal evolution and human development [1].

We designed a software architecture, Consciousness-based Architecture (CBA) with an evolutionary hierarchy, based on Tran's model, to link animal-like reactive behaviors with symbolic behaviors. The feasibility of the architecture was tested by computer simulation of behaviors including sleep, reflex action, approach, and detour [2]. Since this work, we have designed behavior selection criteria based on the environmental meaning as two-valued, positive and negative, emotions. With these criteria integrated in CBA, experiments using two robots loading this architecture successfully demonstrated ambush to capture a prey [3].

In this study, the behavior selection criteria are redesigned to take into consideration the interdependency of external and internal perceptions so that behavior can be more flexibly and efficiently selected. This study shows the results of experiments of ambush and detour with a limited use of temporal and spatial information of the environment and discusses comparison to BBA and the limit of the proposed architecture.

2.2 Architechture(CBA)

From a Husserlian phenomenological viewpoint, becoming conscious of an object is a feedback process in which meaning is given to the object. Consciousness directed toward an unidentified object is a process in which the meaning of the object is retrieved from the subject's memory based on observation of it. In the meantime, the meaning of an object becomes more certain as the object identification proceeds further. Based on the phenomenological analysis, Tran, a Vietnamese philosopher, proposed that the level of consciousness directed toward an object is the one that gives meaning to the object and is also the source behavior conducted toward the object, when the behavior is obstructed. Thus he linked consciousness to behavior development in his conceptual model of the hierarchical relationship between mental process and behavior as shown in Table 1. In this hierarchy, the level of consciousness activated selects and produces an action at the immediately higher level than the level

of inhibited behavior. He assumed that the mental process of an animal has evolved in the phylogeny from single-celled animals to humans, just as human consciousness develops in its ontogeny. Comparison studies on the encephalization of animals [7] and ethological studies of behavior development [6] support this assumption.

We designed Consciousness-based Architecture (CBA) to implement Tran's conceptual model. CBA has a six-level hierarchical structure of relationships between behavior and consciousness in animals lower than the apes, or infants younger than two years,. The hierarchy of CBA, as shown in Fig. 2.1, is efficient because a complex task is processed with elevation of the level: Inhibition of behavior elevates the level of consciousness and behavior.

Level	Phylogeny	Ontogeny (age)	Consciousness Field	Behavior
8	man	4 years	conception	linguistic actions
7	man/ape	2 years	symbolic representation	production of tools
6	ape	18 months	symbolic images	use of tools
5	monkey	1 year	temporal/spatial relationship of objects	use of media, mates' motion and geography
4	quadruped mammal	9 months	stable emotion to object	detour, search, pursuit, manipulation of limbs
3	fish	5 months	temporary emotion to present object	capture, approach, attack, evade, escape
2	earthworm	1 months	valued sensation of pleasure and displeasure	orientation and position of body and limbs
1	sea anemone, jellyfish	0	memoryless sensation	reflex action, displacement, feeding
0	any	Any	basic consciousness of awakening, dream	basic reaction of survival, sleep

Table 2.1 Relationship between Consciousness Field and Behavior[1].

[1] The first column denotes the level, and the second the phylogeny where typical examples are shown. The third column shows the ontogeny where typical ages are shown when the consciousness and behavior of the level first appears, the fourth the consciousness field, and the last column, typical actions the consciousness at the corresponding level triggers. For simplification, animals on the boundaries are ignored, and the infants' ages given in the table are average ones at which the corresponding consciousness and behavior first appears.

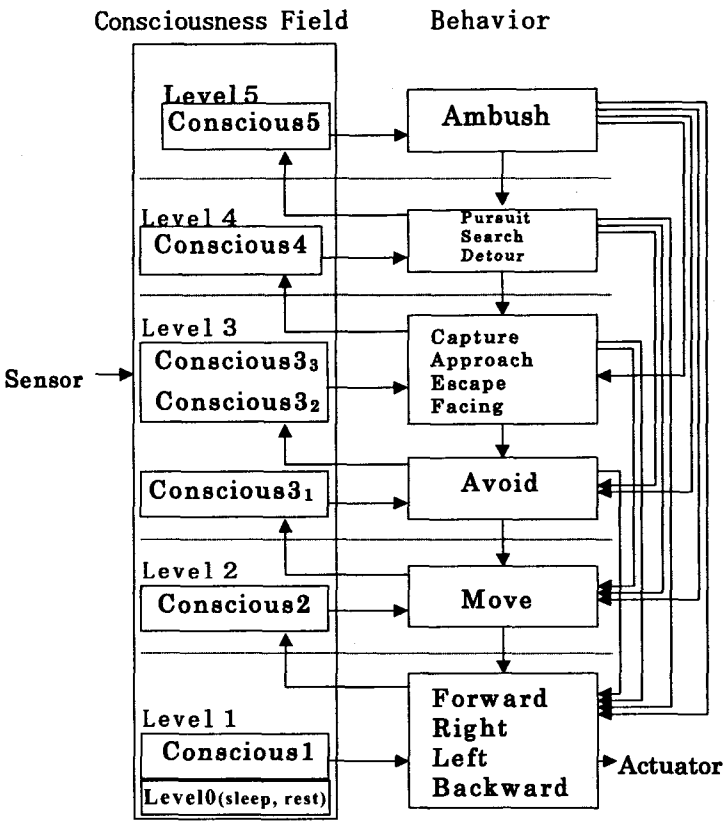


Fig. 2.1 Hierarchy of the CBA[2].

The major functions of CBA for behavior selection as mentioned in the next section are: (1) to make the animal at level p conscious of an obstruction at level p+1, and avoid it at level p+1 if an action at level p is inhibited; (2) to activate any level of consciousness and any level of behavior lower than current

[2] Inputs of the sensors come into all the level of the consciousness field, i.e., the perception by sensor inputs is part of consciousness. Each arrow from behavior to consciousness indicates that the consciousness level thus indicated is chosen and activated through the central function for behavior selection. A real-time image of such activation of consciousness is shown on the computer screen, and is used as a behavior design tool. An arrow from the consciousness field selects and activates behavior at the level the arrow indicates.

level p of behavior based on a perception-dependent level selection criterion; (3) to select an action at an activated level by an action selection criterion. Perception by sensors is linked to all levels of the consciousness, but the level at which a perception appears depends on the perception. The memory of consciousness and behavior is distributed into all levels of the consciousness field, and is used according to behavior function.

The behavior function B_{ij} for jth action at level i is defined as $\{B_{1j}\}$={Move Forward, Right Turn, Left Turn, Move Backward}, $\{B_{2j}\}$={Move}, $\{B_{3j}\}$={Avoid, Face, Escape, Approach, Capture}, $\{B_{4j}\}$={Search, Pursue, Detour}, $\{B_{5j}\}$={Ambush}. Each B_{ij} is an appropriate combination of lower behaviors.

Although an off-line learning algorithm may obtain an appropriate combination of behaviors, in this study all the combinations of behavior are determined and fixed for simplicity. A real-time image of the activation of consciousness and behavior is shown on the computer screen, and used as an interface tool for behavior design. A user can input virtual obstacles and objects on the screen of the consciousness field. Details of this interface is described elsewhere [3].

According to Tran's conceptual model of consciousness and behavior, we also assume that the mental process exists in any animal from single-celled animals to humans, that is, it has consciousness. The stipulation is not trivial but raises the issue of how the underdeveloped mental process an animal or infant compares to that of the adult human. But it is obviously not easy to solve this problem experimentally. Before answering this question, the word *consciousness*, while still applicable to any animal in our context, must be extended in meaning. The word "*consciousness*" is used in a broad sense to mean awareness of an object externally and/or being internally capable of responding to it, where external response means behavior and the internal one consciousness of emotions, representations, and symbols. Thus all living beings from a single cell to a human can be said to have "*consciousness*".

2.3 Criteria for Behavior Selection

The flow of the whole processes of the central function for behavior selection is shown in Fig. 2.2. Change of behavior depends on the two criteria of emotion as defined below, $C_k(t)$, the grade of consciousness for level selection and $I_i(t)$ representing the degree of pleasure for action selection at a selected level. Suppose that the animal is conducting a behavior B_{pr} at level p. There are three

main loops for behavior selection depending on the criteria. The first, major one is from A through B and C to A where "behavior inhibition" occurs and then the level of the consciousness goes up to p+1 level. The second loop is from A through B and E to A where the animal stays at the same level p, and the third one from A through B and C to A where the level goes down.

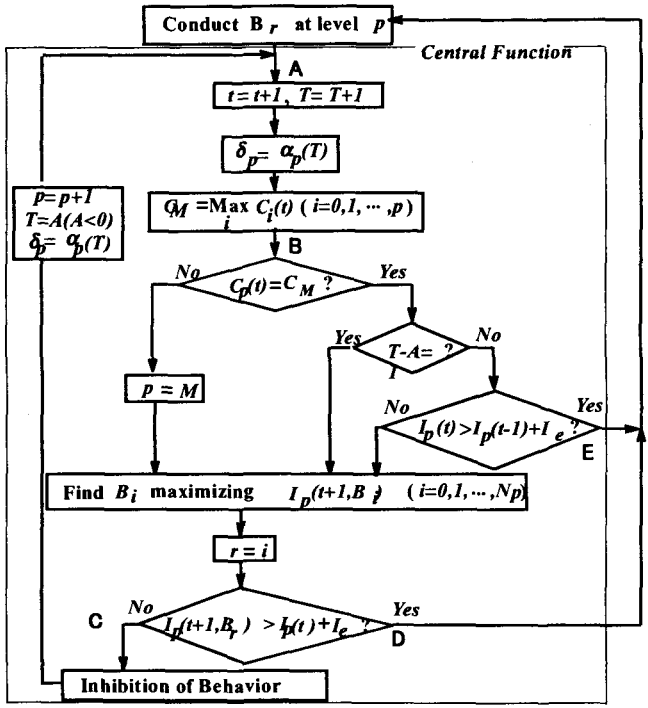


Fig. 2.2 Flow Chart of Behavior Selection[3].

If there is a level M maximizing the criterion $C_i(t)$ for i ($0<i<p$) at a time t, then the animal decides to perform an action B_{Mq} in the behavior set $\{B_{Mj}(j=1,..,N_M)\}$ such that $I_M(t+1,B_{Mq})>I_M(t)+I_e$ for a fixed threshold I_e. But if

[3] There are three main loops for behavior selection depending on the criteria. The first, major one is from A through B and C to A where "behavior inhibition" occurs and then the level of the consciousness goes up to p+1 level. Short-time prediction of the pleasure value is achieved in the discrimination of C or D. The second loop is from A through B and E to A where the animal stays at the same level p, and the third one from A through B and C to A where the level goes down.

$I_M(t+1,B_{Mj}) \leq I_M(t)+I_e$ for any j (j=1,..,N_M), then the last behavior B_{pr} is considered to be "inhibited", and the level of consciousness elevates from p to p+1. In the case of behavior inhibition, α_{p+1}, absorption in behavior, in the following equation (2) plays an important role in securing the continuous selection of level p+1 after the behavior inhibition at level p until a different level than p+1 is selected.

$$C_i(t)=\begin{cases} \sum_{j=1}^{N_E} |\beta_{ij}| + \sum_{j=1}^{N_I} |\gamma_{ij}|, & (i \leq p) \quad (1) \\ \delta_{p+1}(t) + \sum_{j=1}^{N_E} |\beta_{p+1j}| + \sum_{j=1}^{N_I} |\gamma_{p+1j}|, & (i = p+1) \quad (2) \end{cases}$$

where

$$\text{for } \delta_{p+1}(t) = \begin{cases} \alpha_{p+1}(t) & \text{only when the level remains at p+1 after changing from p to p+1} \\ 0, \text{else}, \end{cases}$$

and

$$I_i(t) = \sum_{j=1}^{N_E} \beta_{ij} + \sum_{j=1}^{N_I} \gamma_{ij} \quad (3)$$

where $\alpha_{p+1}(t)$ ($1>\alpha_{p+1}>0$) is absorption, explained later, in a selected behavior at level p+1. β_{ij} and γ_{ij} are the intensity of the external stimulus at level i and that of jth internal perception such as hunger, respectively, and β_{ij} depends on the position of the object: the position is measured by the four categories of distance of zero ($0 \leq d<20$), near ($20 \leq d<100$), medium ($100 \leq d<400$), and far ($400 \leq d$) for d in mm. β_{ij} and γ_{ij} are normalized between -1 and 1, positive for a positive emotion and negative for a negative emotion. N_E and N_I are the numbers of external objects and internal perceptions respectively. The time functions, α_i, β_{ij} and γ_{ij} should be determined depending on the modeled animal, but are assumed to be explicitly time-invariant for simplicity in the present study.

The animal continues the same action B_{pr} at level p if $C_p(t)$ is maximal in $C_j(t)$ for $0<j<p+1$ and $I_p(t) \geq I_p(t-1)$. But if $I_p(t)<I_p(t-1)$ for the same action B_{pr}, then the animal chooses B_{p_i} such that $I_p(t+1,B_{pi})>I_p(t)$. If there is no such behavior, then the behavior is inhibited, and the animal seeks a behavior at level p+1. If there is no choice of behavior at level p to increases $I_p(t)$, then the animal, whose highest level of behavior is p, would remain at p. In addition, if the maximal grade of consciousness $C_M(t)<C_0$ for a fixed positive real number C_0, the animal does not move from level 0, which is resting or sleeping.

Absorption α_p represents the intensity of the consciousness that an animal tries to continue to conduct a purposive behavior at level p. In this study, it is assumed that absorption occurs only while the animal stays at level p immediately after behavior at level p-1 is inhibited. The criterion $I_i(t)$, by which an action is selected from a set of actions at level i, represents a summation of emotional values, i.e., grade of comfort, which is defined by removing the symbols of absolute value from $C_i(t)$ without the absorption term. The physical meaning of the maximal $C_i(t)$ at each moment, i.e., a summation of all the absolute grades of perception, represents the activity of the consciousness at level i. $I_i(t)$ is the grade of pleasure for survival at level i. Psycho-physiological interactions between internal and external perceptions can be taken into consideration in the relationships between β_{ij} and γ_{ij}, e.g., hunger producing a chill, and sensation of fullness producing drowsiness.

2.4 Behavior Design

Actions at each level have been carefully tested using the interface tool and small Swiss mobile robots, Kheperas, hooked up to a SunSparc10 host computer. The robots have eight infrared proximity sensors that can sense an obstacle within 2 cm. The robot also has vision at the top, which detects real-time distance to a striped object up to 40cm ahead. Distance is measured by built-in encoders with an error of less than 2%. Behavior level design was achieved with level 1 first, level 2 was piled on top of it, and so on up through the whole hierarchy from 1 to 5. Then $C_i(t)$ and $I_i(t)$ (i=0 to 5) were appropriately tuned and fixed to realize the desired behaviors for the experiment [3].

Level 1: Any external stimulus results in a reflex action, either attraction with $\beta_{1j}>0$ or repulsion with $\beta_{1j}<0$, and a simultaneous perception, which appears in the consciousness field without memory. Four actions are defined: turning right or left, and going forward or backward. What the robot likes is defined as a black object and dislikes is defined as a striped one. This level of consciousness

has no memory.

Level 2: The consciousness field at this level values a stimulus as either pleasure or displeasure, depending on the memory of stimuli. If a jth stimulus which causes a reflex action at level 1 goes over the threshold at both levels 1 and 2, i.e., $|\beta_{1j}|=|\beta_{2j}|=1$, then the emotional value and position of the stimulus are memorized. This memory, therefore, works as a coordinate system to help ensure survival. At this level, there is only one behavior named 'move', which orients and moves the body toward a goal.

Level 3: Consciousness at this level is memoryless emotion. Thus, if the robot approaching a favorite object is blocked by an obstacle placed in its way, thus losing sight of the object, it can no longer search for the object. Only the two opposite values, positive and negative are used for simplicity, although the emotion at this level should be partitioned into more than two values, such as desire, pleasure, comfort and hatred, anger, pain, fear depending on the object. Four actions are defined at this level: avoidance, facing, approach, and escape. We simulated the hesitation of a hungry animal at the highest level 3 with an object of prey and a hated object located in the same place. Suppose that the animal feels $C_3(t)$ and $I_3(t)$ with increasing hunger $\gamma_{31}(t)$ and absorption $\alpha_3=0.4$. $\beta_{31}(d)$ and $\beta_{32}(d)$ are tuned so that $\beta_{31}(d) + \beta_{32}(d)$ is a single-peaked function of the distance d in $0<d<20$. The robot moves toward the objects for $d>10$ and then backward for $d<10$ because $I_3(t)>I_3(t-1)$ is true due to the single-peakedness of I_3. This behavior looks like hesitation, but the robot finally arrives at the prey at c as its increasing hunger becomes dominant in $I_3(t)$. Details of the behavior like hesitation are shown elsewhere [3]

Level 4: For an action, such as search, pursuit, and detour, an animal must have a stable and lasting emotion for the object. If a prey disappears behind an obstacle, stability of desire enables the animal to continue to search and pursue. Detour and pursuit were successfully tested. At t=0, the animal began moving toward the object at level 3. At t=1 an obstacle was placed in front of the robot to hide the object. Level 3 behavior was not an option so that $I_3(t+1)<I_3(t)$, movement toward the object at level 3 was inhibited, and the robot changed to level 4. At t=2, the robot then detoured the obstacle to pursue the object in its memory so that $I_4(t+1,\text{Detour})>I_4(t)$. An experiment of detour using two robots is shown in the next section.

Level 5: Only one choice of behavior, ambush, is assumed. It sequentially achieves the three behaviors: (i) finding the closest obstacle for use as a hiding place, (ii) moving to the hiding place, and (iii) moving around behind the hiding place to watch the prey. Two experiments using the whole hierarchy of the

levels of 1 to 5 are shown in the next section. This level has the same long-term memory as level 4.

2.5 Experiments

Two experiments were carried out using two Khepera robots, Himiko and Takeru. In these experiments, Parameters of each animal in the behavior selection criteria are determined so that: (1) Himiko likes Takeru, but dislikes theTakeru's striped nest, and Takeru hates Himiko, (2) CBA of levels 0 to 5 are installed in Himiko, and levels 0 to 3 in Takeru, (3) Himiko's moving speed, 24 mm/s. is slower than Takeru's, 40 mm/s, (4) both Himiko and Takeru are hungry, i.e., the time-decreasing function $\gamma(\tau)$ for both is appropriately defined.

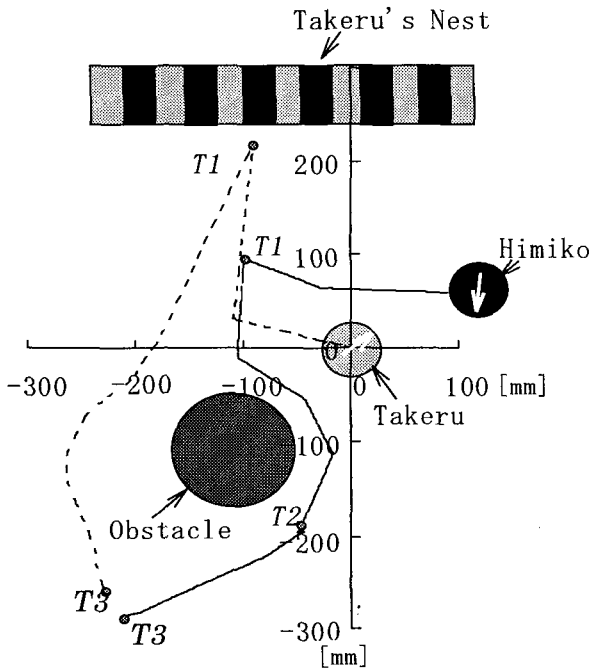


Fig. 2.3 Behavior Tracks of Predator and Prey[4].

[4] Ti (i=1,2,3) denotes the time when a major event occurs. The arrows on Himiko and Takeru show the direction of their visions.

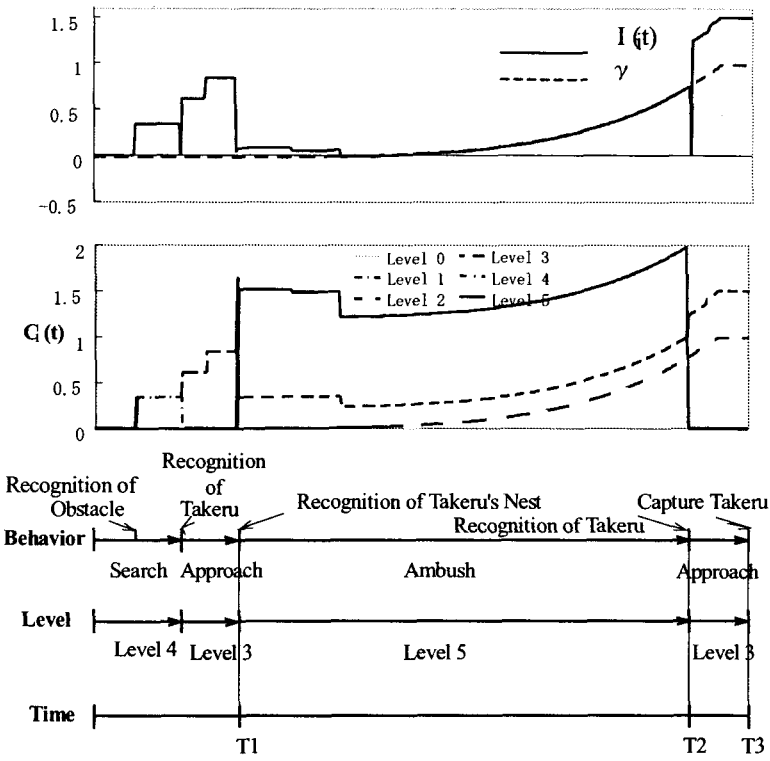


Fig. 2.4 Time Course of Himiko's Behavior Selection Criteria[5].

The first experiment is one in which Himiko ambushes Takeru to capture him. The schematic bird's view of the experimental setup and the computer output of the behavior tracks are shown in Fig. 2.3. While approaching Takeru at level 4, hungry Himiko found him hidden at T1 in the nest that she dislikes. Then Himiko's approach toward Takeru is inhibited because $I_3(t+1,B) \leq I_3(t)+I_e$ for any behavior B at level 3, and so Himiko's consciousness changed to a level 4 desire for favorite Takeru. But her hatred for Takeru's nest is so strong that Himiko immediately went up to level 5 because $I_4(t+1,B) \leq I_4(t)+I_e$ for any behavior B at level 4. The robot ambushes at level 5 as long as the condition C_5

[5] On the first panel, the maximal value of the action selection criterion Ii(t), i.e., the criterion value at the activated level of consciousness, is depicted. See which level of consciousness is activated in the middle and bottom panels.

(t)>C_i(t) (i=1,...,4) is true, until she recognized Takeru at T2. Himiko's recognition of Takeru maximized her C_3(t). In this experiment, β_4, Himiko's perception of Takeru at level 4, is defined as an increasing function of hunger so that $I_4(t+1,B) \leq I_4(t)+I_e$ can easily occur. This definition makes Himiko's hunger drive her more toward Takeru resulting in her skipping level 4 with no behavioral output at level 4.

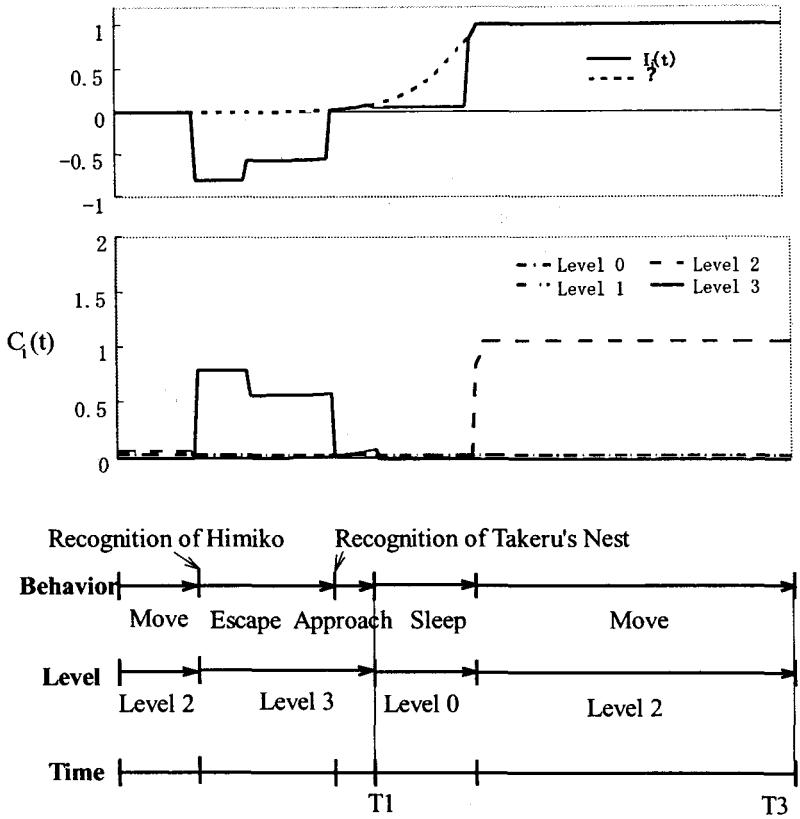


Fig. 2.5 Time Course of Takeru's Behavior Selection Criteria[6].

[6] On the first panel, the maximal value of the action selection criterion Ii(t), i.e., the criterion value at the activated level of consciousness, is depicted. See which level of consciousness is activated in the middle and bottom panels.

Takeru's recognition of both Himiko and his nest forced him to escape to the nest at level 3. The increase in hunger interrupted his sleep and made him move in search of food at level 2. The time course of Himiko's behavior selection criteria and that for Takeru are summarized in Fig. 2.4 and Fig. 2.5, respectively.

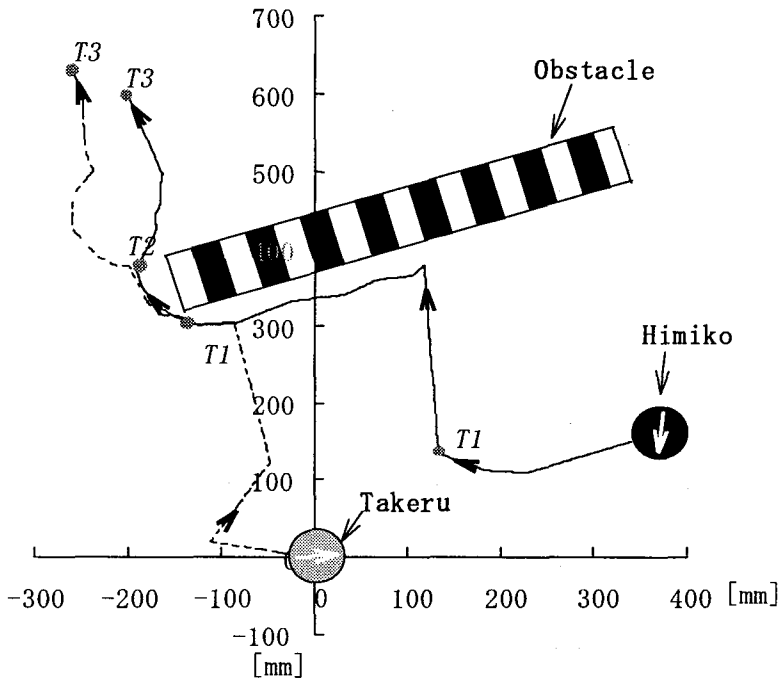


Fig. 2.6 Behavior Tracks of Pursuit by Detour at Level 5.

The second experiment is Himiko's pursuit of Takeru by detour at level5. The same parameters as those of the first experiment were set for both robots. Fig. 2.6 shows the behavior tracks of the two robots, and Figs. 2.7 and 2.8 show the time courses of their behavior selection criteria. After losing sight of Takeru at T1, she did not go into level 5 but into level 4 from level 3 and stayed there to conduct a detour. The different behavior of Himiko from the first experiment after T1 is because the striped object, a simple obstacle, did not move Himiko's emotion in this experiment. Therefore, Himiko moved into level 4 from level 3, staying at level 4 to conduct a detour satisfying $I_4(t+1,\text{Detour})>I_{34}(t)+ I_e$.

Each robot has its own contextual data of consciousness that is readable to the human interface. The human interface can pick up all the robot's memory from the memory belonging to the highest level of CBA since each level can access the memory of all the lower levels'. Contextual data includes obstructed behavior, environmental information, time, position, values of behavior selection criteria, and meaning of the obstruction in terms of emotional value when a behavior was obstructed.

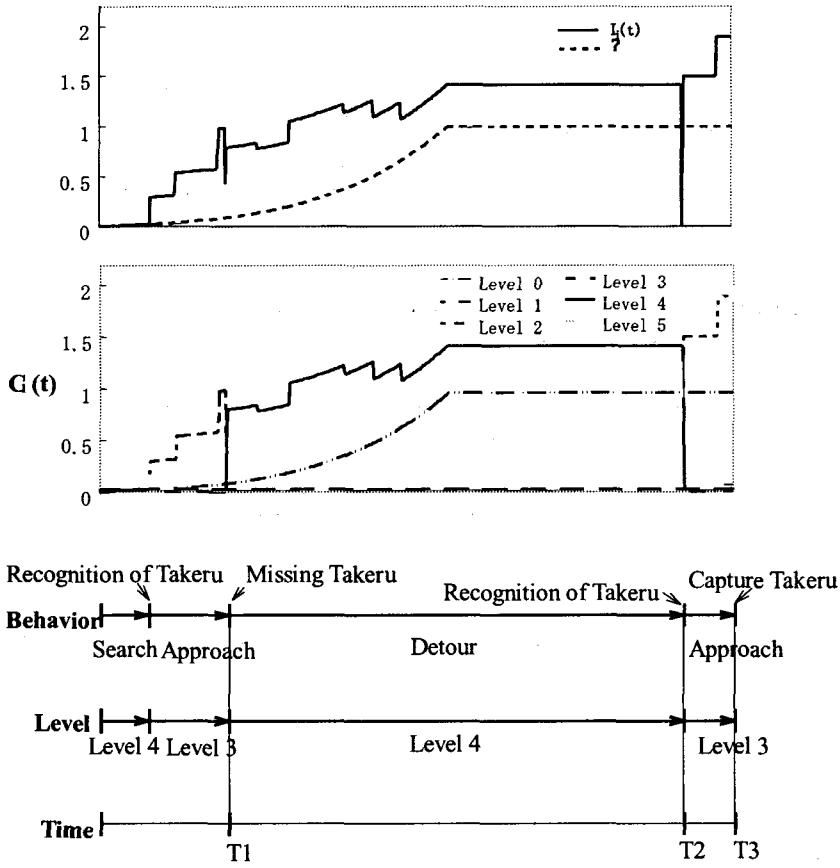


Fig. 2.7 Time Courses of Himiko's Behavior Selection Criteria for Detour Experiment.

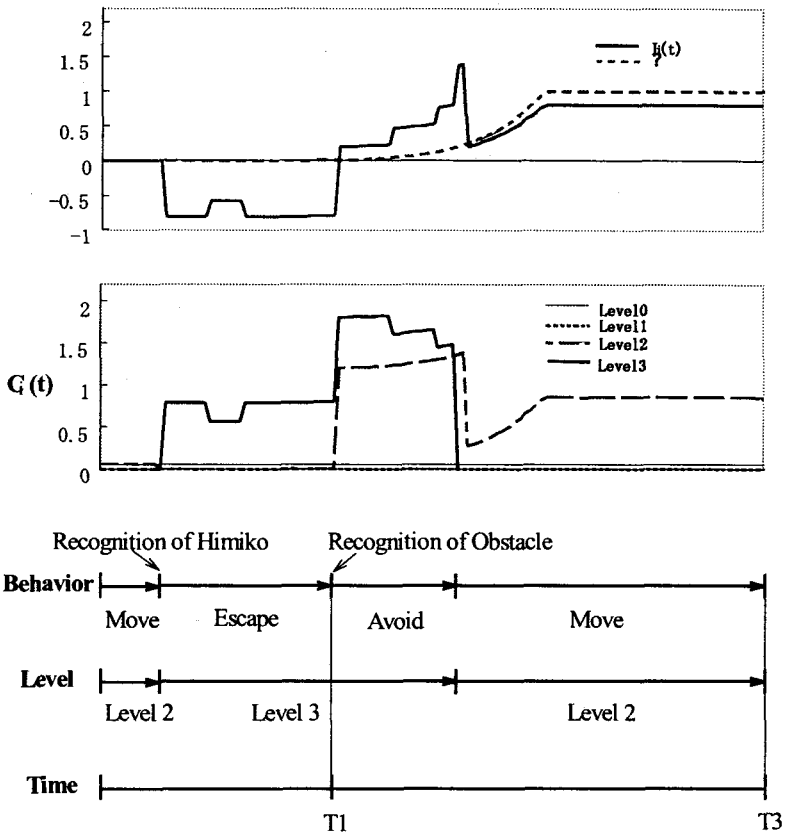


Fig. 2.8 Time Course of Takeru's Behavior Selection Criteria for Detour Experiment.

2.6 Discussion

The advantage to applying CBA to the animal-like behavior design of mobile robots is that it takes an efficient action selection strategy in that obstruction of behavior at one level activates an internal state for choosing a higher level of behavior arranged in the evolutionary hierarchy of CBA. Thus this architecture makes it possible to adaptively link reactive behavior and symbolic behavior such as ambush. To link a high level behavior to a lower one, the meaning of an object in terms of emotional value is organized in the behavior selection criteria.

Efforts to link a reactive behavior to a higher one have been made [4], but little attention has been paid to either making representations of a robot's internal process available for behavior design or taking into consideration the use of the emotional meaning of an environment for behavior selection.

The experiments using the two small robots showed that the linkage of reactive and symbolic behavior is successful in the following limited senses. First, the inhibition of Himiko's approach activated the consciousness level to select higher behaviors, such as detour and ambush. Second, the behavior and its subjective meaning in terms of emotional value can be designed by visualizing the consciousness and behavior transitions on the screen. Performing a symbolic behavior seems to require independence of physical needs to a large extent, and it may seem logical that a symbolic criterion should be designed for action selection at level 5 rather than using an emotional one. But further investigation is necessary to determine how symbolic behaviors at level 5 should be linked to emotions at lower levels.

Although Brooks's Subsumption Architecture (SSA) [5] and CBA appear similar due to the use of a hierarchy of behaviors, our approach should not be classified into a behavior-based approach. Our architecture has two central criteria for behavior selection by which a user can know the reasons for behavior selection. These central functions are linked to the representation of the consciousness, whereas Brooks assumed intelligence without representation in robots as well as animals [5]. When a behavior is inhibited, CBA activates a representation on the next higher level than the level of the inhibited behavior, while SSA assumes that behaviors are activated without representation. Our approach, however, is advantageous because the consciousness field representing a complex of its internal and external states is used as a human interface for behavior design.

An overall view of the design framework of CBA can be explained using the three dimensional design space illustrated in Fig. 2.9. The semantic subspace designed on the basis of CBA consists of the two axes, consciousness and behavior. This space is the equivelent to the sonsciousness field of CBA, in which the meaning of behavior is embedded by a robot designer. Ethologists, on the other hand, are interested in the subspace of behavior and physiological evolution, which in SSA is also designed to designed to combine reactive behaviors without use of symbolic representation for the control of behaviors, and therefore ignore the semantic subspace. Animal-like behavior design using CBA can be positioned and embodied in this 3-D space, where the specifications of the robot's hardware and behavior design should be given to the axes of body

evolution and behavior, respectively. The position of a bee and that of human are illustrated in the figure.

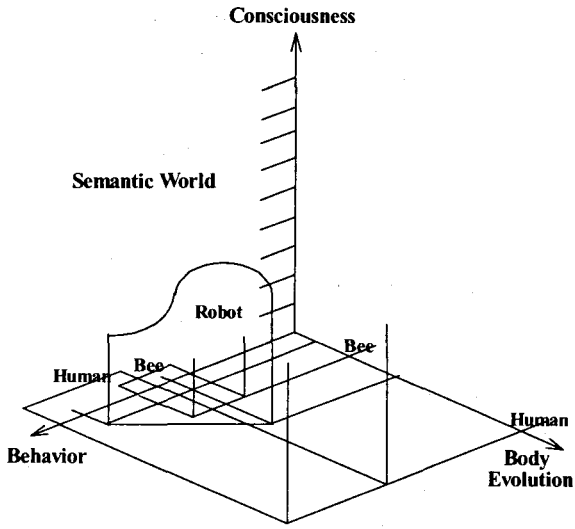


Fig. 2.9 3-D Design Space of CBA[7].

From a Husserlian phenomenological viewpoint, becoming conscious of an object is a feedback process in which meaning is given to the object. Anticipation plays an important role in controlling this process. An early stage of consciousness directed to an unidentified object is a process where the meaning of an object is retrieved from the subject's memory based on observation of it. In the meantime, a deeper meaning is anticipated as the object identification proceeds further. Then the anticipation becomes a reference to which different types of meaning of the object are removed. Unless the anticipation is disappointed, it becomes a belief about what the object really is. Contemplation of this view will provide a hint for a better approximation of consciousness in which an anticipation term as a function of past emotion to an environment should be added to the behavior selection criteria.

[7] The behaviors of a bee and human overlap for levels up to 4 on the axis of behavior, but they don't share anatomical structures on the axis of body evolution. A robot as an artificial animal may be constructed as a hybrid creature of human and bee in the semantic subspace depending on a designer.

2.7 Conclusion

CBA is a six-layered hierarchy of consciousness linked to behaviors including reflex action, approach, detour, and ambush. While a level of behavior is chosen so that the magnitude of the consciousness, a function of perception is maximized, an action at the level thus chosen is selected to raise the emotional value of the level higher than before. A level of consciousness in CBA is activated if a behavior is inhibited; then the activated consciousness selects and produces an action at the immediately higher level. The model design was tested using two small mobile robots, and a limited use of temporal and spatial information of environments was shown. Anticipation as short-term prediction should be embedded for a better approximation of animal behavior in future work. Anticipation would make it possible to perform symbolic behaviors, such as exploitation of a partner's behavior for capture of prey.

References

[1] Tran, D.T, Phenomenolgie et materialisme dialectique, Edition Minhtan, Paris, (1951) (Japanese Translation, Goudou-Shuppan, 1989).

[2] Kitamura, T., Imitation of Animal Behavior with Use of a Model of Consciousness-Behavior Relation for a Small Robot, Proc. 4th IEEE International Workshop on Robot and Human Communication, pp.313-316, (1995).

[3] Kitamura, T., Animal-like Behavior Design of Small Robots by Consciousness-Based Architecture, Advanced Robotics, Vol.12, No.3, pp.289-307, (1998).

[4] Connell, J., A Hybrid Architecture Applied to Robot Navigation, Proc. of IEEE Int. Conf. on Robotics and Automation, pp.2719-2724, (1992).

[5] Brooks, R.A., Intelligence without Representation, Artificial Intelligence, vol.47, pp.139-159, (1991).

[6] Griffin, D.R., The Question of Animal Awareness: the Evolutionary Continuity of Mental Experience, Rockefeller University Press, New York, (1976).

[7] Jerrison, H.J., Brain Evolution and Dinosaur Brains, Am. Nat. Vol.103, pp.575-588, (1969).

[8] Balch, T and Arkin, R.C., Behavior-based Formation Control for Multirobot Teams, IEEE Trans. Vol.14, No.6, pp.926-939, (1998).

This page is intentionally left blank

Chapter 3

A Computational Literary Theory: The Ultimate Products of the Brain/Mind Machine

Akifumi Tokosumi
Tokyo Institute of Technology

Abstract

This paper is an attempt to establish a new class of computation called literary computing. The author introduces a novel concept in the analysis of cognitive processes --- the reader's wish to understand a certain text in accordance with his/her own goal/plan knowledge structure. The possibility to model the wish generation process as well as its relationship to knowledge computing and affective computing is also discussed.

Keywords : literary computing, literary theory, affective computing, emotion, wish, desire, psycholinguistics, text understanding, film understanding, cognitive model, computational psychology

3.1 Introduction

3.1.1 *From Knowledge Computing to Affective Computing*

Expansions of the concept of "intelligence" may be expected to be one of the major outcomes of modern cognitive science. This concept, which originated from classical ideas about inference rules and knowledge representations, includes affective states of human mind, as it is our common understanding today that emotions are particular states of the human cognitive system. An expansion of knowledge computing to affective/emotional computing is also regarded as a natural development of intelligence computing research [1].

One of the reasons which make cognitive science so attractive is the surprising variety of the human mind. Let's explore the linguistic abilities. With only a piece of newspaper article, poem, or novel, our imagination can construct a

situation, a society, a culture, a world, free from definitions of time and place. This would be merely an example of how the powerful abilities of our symbolic process minimize various constraints stemming from environments and situations, and how the freedom of our mental activities arises. Language is certainly the field in which our mental capacities show their ultimate best. Language is also the field where knowledge-based intelligence meets affective/emotional intelligence in its precise resolution. With more than a hundred of affective words, language gives our intelligence an ability to distinguish, represent, and communicate precise states of our mind [2].

3.1.2 *From Affective Computing to Literary / Aesthetic Computing*

This paper presents yet another expansion in cognitive modeling, i.e. the expansion toward the poetic and artistic aspects of the human cognitive system. Our mental processes wouldn't stop when they recognize and understand the outer world. We always evaluate and appreciate the world. Sometimes we contemplate, mediate and are moved. Cognitive science needs to invent vocabularies to describe these aspects of mental processes. What the present paper calls literary computing is a class of computing to deal with these phenomena.

The concept of affective computing [1] has advanced the way we talk about cognitive activities. The addition of emotion processes to the classical concept of cognitive counterparts may not be sufficient, however. Cognitive scientists and psychologists have long used terms such as creativity and kansei(aesthetic sensitivity). These terms connotate similar sets of phenomena mentioned here. Literary computing may bring new perspectives into the tradition. This is why we need the new vocabulary including terms like literary computing, aesthetic computing, and kansei computing.

3.1.3 *The extremity of Literary Computing*

The fact that literature texts are in written format is of crucial importance to our analysis. Cognitive processes which may be different from those for spoken utterances are made possible by the written texts as external objects. The outcomes of mental activities can't be produced through real-time processing. These outcomes, which emerge only after repeated, concentrated mental processes, are the mental constructs discussed here. Most people have novels, music and pictures they often re-read, re-listen, and re-appreciate. Objects' re-examination may be an essential characteristic of the artistic component of our mind. Another characteristic of literary texts is their dependency on symbolic

expressions, that is the linguistic forms. Minimizing dependency on sensory channels and literary computing is on the extreme side of the aesthetic computing spectrum. Although we could certainly expect much from research in music and visual art computing, more direct results would be expected to be obtained from literary computing research. (Fig. 3.1 shows the hierarchical relation among various cognitive computing classes.)

Aesthetic Computing
literary computing, poetic computing, music computing, ...

Affective Computing
emotion/affect representation, emotional inference, motivation computing, ...

Knowledge Computing
knowledge representation, rule-based inference, case-based inference, ...

Fig. 3.1 Cognitive Computation Hierarchy.

3.1.4 *Research Strategies*

What would be the possible cognitive research topics of interest in literary computing? The present paper proposes three types of strategies:

(a) Literary cognition:
- literature experiences as cognitive phenomena
- cognitive literature theory
- psycho-literature theory

(b) Knowledge computing and literary computing:
- socio-cultural knowledge and language understanding [3]
- creative computing and language understanding [4]
- affective computing and language understanding [5]
- affective computing and literature appreciation [6]

(c) Compatibility to literature, music, and aesthetic theories:
The following section describes some attempts into the (a) and the (b) strategies.

3.2 Literary Text and Cognition

3.2.1 *Layers of Understanding*

The following is a scene from the French film *La Ballon Rouge*:

> *A boy attempted to bring a red balloon in the classroom. But the teacher didn't allow it, and the boy had to leave the balloon in the school yard.*

What would be the possible cognitive constructs made by the viewers of the above scene? Although they must be built through the complex interaction of everyday common sense knowledge, inference based on viewers' experiences, affective reasoning and so forth, we analyze them in four layers of different mental activities.

Understanding as problem solving

Understanding human activities, such as the actions of the boy in the above example, can be done with the help of the goal/plan knowledge structure paradigm [7]. [bring a balloon into the classroom] is a part of a plan to achieve a certain goal [BE-WITH the balloon]. Analysis based on this type of everyday knowledge structures leads us to mental products like:

```
[GOAL=BE-WITH(boy, balloon)]
        --> [plan=bring in the classroom]  -->  failed
[GOAL SUBSTITUTE=BE-NEAR(boy, balloon)]
        --> [plan=leave in the school yard]  -->  achieved
```

Understanding as affective reasoning

Our everyday knowledge about the human mind also includes emotional aspects of our life. We have plenty of knowledge about the correlation between action outcomes (achieved, failed, suspended) and mental states of the action agent [3]. When people fail to achieve a goal, we know that they generally feel sad, disappointed, and sometimes get angry when the failure was caused by others.

```
[GOAL]  -->  [PLAN EXECUTION]
        -->  failure  -->  {sad, disappointed, angry, ...}
```

We also know that some affective states can spawn other goals.

{negative state} --> [GOAL=REDUCE negative state]
--> [PLAN={revenge, wreak, ...}]

Understanding as emotion evokation

When we watch films or read novels, our mind works not just to understand correctly their contents. Our mental processes sometimes evoke emotions toward various objects. Some viewers may feel pity for the boy, while others may feel pity for the balloon as well.

GOAL failure --> [EMOTION pity-for boy]
GOAL failure --> [EMOTION pity-for balloon]

Understanding as evaluation evokation

A class of mental constructs different from interpersonal emotions may also arise -- interpersonal evaluations. The school teacher who didn't allow the balloon into the classroom could be evaluated as too restrictive, while the boy could be evaluated as too hedonistic.

prevent other's GOAL
--> [INTER-P-EVAL too-restrictive]
pursue only pleasure GOAL
--> [INTER-P-EVAL too-hedonistic]

These mental processes, which are called here emotion evokation and evaluation evokation, are not very helpful to understand the world (the film in our case). Rather, they work independently and heavily dependent on the cognizer's own reasons and logic. These processes, probably through the collaboration with other mental activities, would ultimately produce literary or artistic mental constructs. The term *appreciation* could be more suitable to understand these processes than *understanding/comprehension.*

3.2.2 *From Understanding to Appreciation*

Under the framework described above, we carried out a series of psychological experiments on film and text understanding/appreciation. As a result, transitions of mental activities through film watching were reported [6]. The experiment analyzed protocol data obtained by think-aloud technique while participants watched a 15 min. TV drama. We found the existence of overall transitions in the contents of participants' inferences:

(a) In the beginning of the drama inferences on the drama's title, background music, and such salient features were predominant. (Observation phase)
(b) Later inferences based on everyday knowledge and established schema about the usual roles in dramas became predominant. (Story understanding phase)
(c) Toward the end of drama protocol data showed viewers' wishes and desires. "... probably the drama will end like this, but I wish he (one of the characters in the drama) behaved the other way ..." was the archetypical response. (Wish generation phase)

3.2.3 *The Origin of Wish*

Among the transition [observation --> understanding --> wish], the latter part may represent more interest as it often accompanies evaluative attitudes, emotional attitudes or emotion evoking. A response like

> *If I were him (one of the characters in the drama), I would do ... first of all ...*

could be analyzed as:

> ROLE-X executed PLAN-1 to achieve GOAL-1, but I don't agree (evaluation evokation). I have a better PLAN-2 to achieve GOAL-1(evaluation evokation).

The mechanism proposed here is:

> Duplicate other's GOAL and propose better PLAN

Thus based on goal/plan knowledge structures, a wish production mechanism emerges. This mechanism, along with case-based understanding mechanism and the affective reasoning mechanism, is an important part of our literary text appreciation program known as KEWP (Knowledge and Emotion Workbench Project).

3.3 Literary Computing

3.3.1 *Novels as Knowledge Computing and Affective Computing*

The KEWP program parses and understands / appreciates a novel. The example

shown below is our treatment of Natsume Sohseki's short novel *Ten Nights of Dreams*. The limited abilities of our parser prevented us to use the exact wording of the text. The input text to a version of KEWP is a sequence of the propositions depicted in Fig. 3.2.

A goal recognizing mechanism constructs the goal/plan structure of the story as shown in Fig. 3.3.

The strange feeling most readers of the novel may have could be explained by its goal/plan structure. The novel suggests that the lily is the agent substituting for the dead woman, but not quite clearly. Readers' understanding processes leave the conclusion unsolved and affective reasoning processes also leave the final resolution suspended. However, here is the point a wish generation mechanism is evoked.

First Night
I dreamt.
A woman was whispering "I am dying."
She continued "Please bury me if I die. And wait for me beside the tomb.
Wait for me for a hundred years. I will be back to see you."
She died.
I buried her, and waited.
I waited, and waited.
A green stalk was approaching, and brought a white lily.
I suddenly realized that the hundred years had passed.

Fig. 3.2 Propositional Structure of *Ten Nights of Dreams: First Night.*

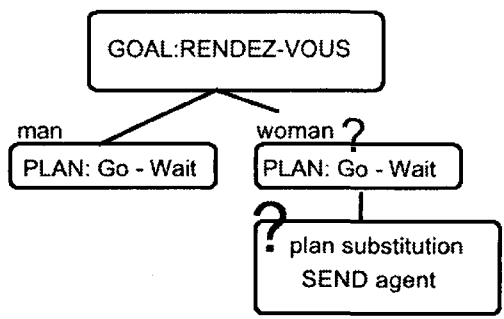


Fig. 3.3 Goal/Plan Structure of *Ten Nights of Dreams: First Night.*

After a certain amount of precondition checking processes, the program duplicates the goal of the dead woman [GOAL: RENDEZ-VOUS]. The duplication of the goal doesn't mean the duplication of the plans. The program has its own set of plans and can propose a better candidate through plan evaluation processes (Fig. 3.4). Human readers may have such individual sets of plans based on their own experiences. A plan proposed by the program is [PLAN: {send signals}] to notify him of her arrival, which roughly corresponds to the following thinking: "if I were the woman, I would send some signal so that he can recognize that the lily is in fact me...". This is the version of wish generation proposed by the program.

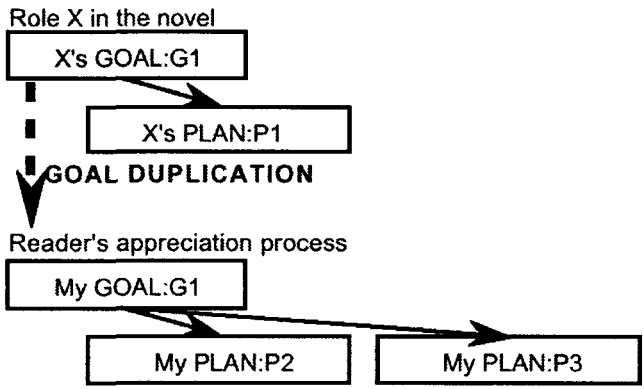


Fig. 3.4 A Computational Mechanism of Wish Generation.

3.4 Conclusions

Based on psychological evidences, we have proposed a wish generation mechanism as an aspect of literary cognition/computation. Although the mechanism presented in this paper is based on knowledge computation techniques, it could be interpreted as the model of an interesting phenomenon: the reader's wish or desire inspired by a literary text.

It is true that our literary experience covers far broader aspects of mental life. It is also true that our literary experience contributes to the development of our value systems which is core to human personality. Despite the limited scope of the presented results, it is very clear that future work in the area of literary computing will lead to important insight to human intelligence.

References

[1] Picard, R.W., Affective Computing, MIT Press, (1997).

[2] Tokosumi, A. and Hayakawa, T., Cognitive Components of Affective Words: Their Mental and Computational Representations, Abstracts for the First International Conference on the Mental Lexicon, pp.112, (1998).

[3] Dyer, M.G., In-depth Understanding: A Computer Model of Integrated Processing for Narrative Comprehension, MIT Press, (1983).

[4] Schank, R.C., Explanation Patterns: Understanding Mechanically and Creatively, Lawrence Erlbaum Associates, (1986).

[5] Mueller, E.T., Daydreaming in Humans and Machines: A Computer Model of the Stream of Thought, Ablex Publishing, (1990).

[6] Tokosumi, A., Cognitive Science of Literature, in Science of Literature (Eds. Iguchi, T., Tokosumi, A. and Iwayama, M.) Asakura Publishing (In Japanese) (1996).

[7] Galambos, J.A., Abelson, R.P., and Black, J.B. (eds.) Knowledge Structures, Lawrence Erlbaum Associates, (1986).

This page is intentionally left blank

Chapter 4

Cooperation between Neural Networks within the Brain

Michel Dufossé, Arthur Kaladjian, and Halim Djennane
University Pierre and Marie Curie

Abstract

Modelling human and robotics task involve the cooperation between neural networks, both in artificial and biological domains. For optimal motor control, the nervous system must take into account (implement) the external constraints. During learning, the dynamical model of the environment is implemented in the brain neural network. By contrast, the moon moving around the earth only takes into account the Newtonian law, without learning it. Soft Computing techniques already deal with the cooperation between neural nets. Brain model models also take into account the philogenetic development of brain structures, with different architectonic architectures and learning processes. So, the cooperation between brain structures is an essential goal for their understanding. Here, motor control is taken as an example of the environmental constraint. The Newtonian law (unconsciouly solved) must be computed by the brain, taking into account the Cartesian metric space (consciouly solved by the instructions).

Keywords : brain motor control, cooperative biological neural networks, cerebro-cerebellar interaction, cerebellum

4.1 Introduction

Modelling the brain assumes many assumptions. Many of them are now validaded by consisting experimantal data from many groups in the world. To simplify, the brain my be roughly divided into two parts: a computational part that we describe here for motor control, and a limbic part (motivations, emotions ...) depending more on Neurochemistry, which we ignore here. Our goal was to use Engineers knowledge for brain modelling. As an example, the aeronautics 'fly-by-wires' technique suggests two functional roles for cerebellum and for basal ganglia [30]. Before that example, only different clinical deseases and different architectonic anatomical structures were only shown, without understanding

their respective functions. By return, brain modelling was the source of many engineers results, such as distributed processing by Neural Networks and functional human reasoning by Fuzzy Logic. More recently, there was a great computer advance in Chess playing, a game that cannot be solved by pure algorithmic solutions.

Motor learning is performed by modifying the flows of signals transmitted between neural structures, mainly the motor areas of the cerebral cortex, the cerebellum and the basal ganglia. According to modern control theory, learning requires not only an adaptable system but also the possibility of changing the information processing rules [29].

In each cortical structure, it is possible to define a basic cristalline unit, consisting of several neuronal types [48], which recurs throughout the structure. This basic unit is the microcolumn of the cerebral cortex, the microzone of the cerebellar cortex, and the striato-element module of the basal ganglia. These morpho-functional structures, together with their activation and plastic rules, are taken here as the basic units of neural network modelling. Present models suggest how cellular mechanisms in various structures may be responsible for different types of adaptive process. Future models must explain how mutual interactions can produce automatic and refined motor sequences.

At the cerebral level, which is the first to operate, it is possible to compensate for any weaknesses in the spinal mechanisms by producing responses which are better adapted to disturbances caused by the environment. When a task is performed repeatedly, the cerebellum is able to deal with some of the repetitive aspects so that the motor response becomes more finely attuned and automatized. The cerebellum can thus free the cerebral cortex during sensorimotor or even mental tasks [39]. The basal ganglia nuclei are more involved in the postural stabilization during elementary movements and in the optimal destabilization between successive elementary movements. Both the cerebellum and the basal ganglia are involved in motor control, motor planning and cognitive aspects of action, helping the cerebral cortex.

It is here suggested how cellular mechanisms in these three neural structures may be responsible for different types of adaptive processes and how their mutual interactions may lead to automatic and refined motor sequences.

First, we give some clues for plausible neurobiological modelling, based on widely accepted neurophysiological concepts. Secondly, we describe a previously built model of the dialogue that is established between the two cortices, the cerebral cortex and the cerebellar cortex during the early phases of sensorimotor learning [5]. Thirdly, we extend this approcah to the role of the basal

ganglia in motor sequence learning and cooperation with the cerebral cortex.

4.2 Cerebral Cortex : the 'Pilot'

4.2.1 *Anatomical Data*

The cerebral cortex has six layers [4]. At the surface, the upper layer (layer 1) is mainly composed of axons of local cortico-cortical cell interactions. The "granular" layer (layer 4) receives the external inputs, all of them coming from or through the thalamus. This granular layer is much thicker in sensory areas than in motor areas. It is intermediate between two subsets of output pyramidal neurons, the one in the supragranular layer (layers 2 and 3), and the other in the infragranular layer (5 and 6). They form the two principal output pathways, inside and external to the cortex, respectively. Take into account that, modelling physiological data is pure assumption (based on many experimental data) ! (Fig. 2.9, page 91 of [4]).

The pyramidal cells are arranged in vertical columns, perpendicular to the surface, with strong interconnections, sharing afferents with the same sensory or motor significance [52], [44], [28]. These connections are accompanied by excitatory and inhibitory pathways formed by at least five main types of interneurons, which can couple or uncouple columns, and layers within each column, depending upon the patterns of activity involved. This cellular texture shapes the specific operations that each column effects.

Known physiological properties of the cortical column can be summed in a table which gives the two main outputs of the column (intra- and extra-cortical) along with the two main inputs (cortical and thalamic) and the previous state of the column [4]. Before any mathematical formalization or computer implementation, four main properties have to be considered: 1) the relationships between two columns can be either excitatory or inhibitory depending upon the level of activity; 2) the activity can spread through the cortical network even without any significant outputs sent outside the cortex; 3) an amplifying effect is produced when cortical and thalamic inputs are coactive; 4) the relative importance of these two inputs varies from one cortical area to another.

This system is basically an adaptive mechanism producing two types of responses to external events: either a specific cortical action when the inputs are coactive, or an intra cortical "call" to other columns. The "call" can remain in force until one of the columns called produces an extra cortical action which results in an extra-cortical input to the calling column. The action of each corti-

cal column constitutes an equilibrium position or a kind of goal. A call results in an exploration, a search through the possible actions that the cortex can command in order to reach the goal.

Memorization rules make it possible to store the appropriate patterns of activity in the connections between columns. Excitatory connections are strengthened when a called module produces an action outside the cortex that reactivates the calling module (the goal) by an extracortical feedback loop (external causal link).

These call trees have a top-down and a bottom-up activation. First in the top-down direction, calls emanate from possible actions and produce an anticipatory activation of a set of cortical modules which represent possible actions (or subgoals) that could reduce the distance from the goal. The call spreads until it is in keeping with the environmental conditions. Actions are then triggered both in parallel and in sequence, resulting in the attainment of subgoals and the goal (bottom-up direction).

4.2.2 *Kinematics*

Many neurophysiological studies have been devoted to analysing the relationships between neuronal activity and arm movement kinematics. These studies have been performed on the associative, premotor and primary motor areas of the cerebral cortex [18], [19], [20], [7] and on the cerebellum [14]. These authors using two- and three-dimensional reaching tasks, have shown that when an animal makes arm movements in various directions towards visual targets in extrapersonal space, the cell activity varies in an orderly fashion with the direction of the movement. The description of the broad directional tuning of cortical cells around their preferred direction (the direction of movement in which they discharge most during the reaction time period) led to directional information being treated like a population code, as opposed to a single cell code. This has provided a useful tool for interpreting directional relations, not only in various regions of the distributed motor system but also in associative areas dealing with visual information processing.

The importance of directional information in determining neuronal activity raises related questions concerning the coordinate system used by the frontal cortex to represent the direction of arm movements. In fact, muscle command and sensory information are not mapped in the same reference frame and need to be correlated by the information signalling the positions of arm segments relative to the body. This information comes from several sources, motor, so-

mesthetic and proprioceptive. When the initial arm position changes, the body-centered information about the target position is not invariant in an arm-centered coordinate system, and has to be combined with information about the initial position of the arm in order to compute the appropriate motor commands.

All the results obtained upon recording motor cortex cells are consistent with the following statements:

1) The motor cortical cells command muscle synergies, which can be represented by a vector, the "cell's preferred direction", corresponding to the overall effect of the motor command on the arm and hand positions.
2) The orientation of this synergy vector has invariant properties in an arm centered reference system. Consequently, it does not remain constant with respect to an extrapersonal coordinate system, but rotates with the initial position of the arm in space.
3) A cortical command sent to the shoulder and elbow joint muscles will change the arm position in a constant way within this arm-centered coordinate system.
4) The computation of the appropriate motor command, taking the cells' preferred directions into account is performed in the early phase of the reaction time, before the onset of the movement. This therefore results from a combination performed by cortical areas before the movement is initiated, without requiring a feedback loop when the movement is actually performed.

The main cortical operation performed in order to determine the initial direction of the forthcoming movement is a bilinear combination of the initial position and movement direction vectors [6]. The activity of a cell a is then given in a matrix form by:

$$A_a = {}^tP \cdot M_a \cdot V \qquad (1)$$

where P is the initial vector position ('index-t' means transposed matrix) resulting from arm proprioceptive informations, V the vector direction of movement resulting from the gaze direction to the target and Ma a matrix expressing the computational cell properties. This bilinear formulation predicts the main cell activities recorded in the motor area during visually guided 3-D arm movements:

a) Cell's preferred directions with a given initial arm position.

Given a constant initial position, the cortical operations will be linear and the

command can be expressed in terms of its effects on the trajectory in the 3-D space. Any cell a is broadly tuned to a preferred direction D_a which depends straightforwardly upon the initial position, with a fixed set of coefficients which can be tuned by learning.

$$D_a = {}^tM_a \cdot P \tag{2}$$

At any time, the cortical computation A_a of the muscle commands is equivalent to the projection of the desired trajectory vector on this preferred direction.

$$A_a = {}^tD_a \cdot V \tag{3}$$

b) Population vector.

The population vector always predicts the movement direction, even when the arm rotates. The maximum activity of cells corresponds to directions in the 3-D space which change with the initial arm position. However, in each part of space, all the cells' preferred directions form a uniform sampling of the 3-D space and consequently the population vector always predicts the direction of the movement, even when the initial position of the arm changes.

Under experimental conditions [7], the movement is performed in parallel directions in the external space and consequently the population vectors will represent this common direction and will stay parallel, even if they are computed from components which are not the same.

c) Invariant properties of cell activities in the motor cortex.

The preferred direction of each cell changes in an orderly way with the initial position of the arm. For each neuron, there exist specific a subspace (plane) in which the cell's preferred direction rotates exactly with the initial position of the arm. For a rotation defined by the matrix H_q, the new preferred vector direction is given by:

$$D'_a = H_q \cdot D_a \tag{4}$$

However, this property cannot be generalized to the whole 3-D space, since the cell's preferred directions are computed as the product of rotations which are not commutative.

$$D'_a = {}^tM_a \cdot H_q \cdot M_a \cdot D_a \tag{5}$$

d) Population code in the 3-D space.

As the distribution of the cells' preferred directions is uniform, the whole set of vectors will rotate in the same way as the arm in the 3-D space.

$$D_{pop} = S\, A_a \,.D_a = S\,(\,{}^{t}D_a \,.V\,)\,.\,D_a = V \qquad (6)$$

4.2.3 *Dynamics*

The muscle provides the body with a compliant interface with the environment. Together with the spinal reflexes, it constitutes a servomechanism for regulating body stiffness [25]. Nervous system permanently controls the equilibrium position of the servo [13]. The "muscular unit" is the elementary module at the spinal cord level of neural network models. It consists of 1) a single a motoneuron together with the muscular fibers it innervates (showing various mechanical properties from "slow tonic" to "fast phasic"), 2) the associated g fusimotor neurons which modulate the transfer functions of the fusimotor spindles (the muscular sensors), and 3) the associated spinal circuits. A disadvantage of this kind of interface is that the position of remote segments can be disturbed by the forces involved in the movement, causing a loss of both stability and equilibrium. These disturbances are lessened by anticipatory postural adjustments [41]. The movement dynamics, refering to the forces, torques and muscle activities which produce movement, has to be taken into account by central structures for an optimal performance.

Nervous system first selects a working point, such as the finger's tip during a reaching movement, and then determines its virtual trajectory, as defined by the succession of the instantaneous positions, the positions which would be reached at any time if commands were frozen. This operation probably take place in cerebral projecting areas of the basal ganglia, mainly the supplementary motor area (SMA).

The activity of many primary motor cells varies with parameters such as the level of muscle activity, the level and direction of static output force or torque, and their time derivatives [12], [51], [53].

A large population of motor cortex cells encodes motor behavior in a coordinate system which reflects movement dynamics. The broad tuning curves of many proximal-arm related motor cortex cells can be described in similar terms, either during movement kinematics for different directions of movements in the workspace, or during load compensation for different directions of a constant additional load [35]. Moreover, the spatial relation between movement direction-related and load direction related tuning is approximatively in opposite directions for most cells. As a result, the large single-cell load-related activity changes tend to sum, so that the "primary motor" population activity patterns

changed under different load conditions. An important population of proximal arm-related cells in primary motor cortex encoded arm movement in a coordinate system which covaried with the dynamics underlying movement, either by encoding dynamic parameters or by directly coding muscle activity levels.

Many other motor cortex cells are less affected by loads, and so appear to deal with the control of movement direction, trajectory and other kinematic variables, rather than with dynamics or muscle activity [53], [20]. This rule applies more generally in the premotor area, and even more strongly in the posterior parietal area, where the load sensitivity disappears at the cell population level [36].

4.3 Cerebellar Cortex : the 'Smoother Computer'

4.3.1 *Anatomical Data*

The cerebellar cortex is a mosaic structure, the whole surface of which consists of microzones. Cerebellar microzones are the basic units of the cortex of the cerebellum. Together with their output nuclear projecting zones, they form the so-called "cortico-nuclear microcomplex" [29]. Climbing inputs from specific parts of the inferior olive longitudinally organize the cerebellar cortex [45]. A second input is provided by the mossy fiber system, which is characterized by its considerable divergence onto wide regions on the cerebellar cortex. The set of mossy fiber inputs constitutes a general context about the present sensorimotor actions and intents for future movements, that is a large set of signals providing information about the states of activity in various nervous structures, from command structures to more sensory structures. These thousands of microzones work in parallel in a relatively independent manner.

Each microzone can be viewed as a three layer neural network [42], [1]: an input layer formed by cells which originate from the mossy fibers, an intermediate layer of granular cells, and an output layer of Purkinje cells that project to cerebellar output nuclei. Two pathways, one cortical, the other via subcortcal nuclei (red nucleus) transmit this information to the spinal cord. Since plastic synapses only exist between the second and third layers, it is similar to the original perceptron described by [47].

During the adaptive phase, the output layer utilizes an error signal conveyed by the climbing fibers, originating from a teacher, the inferior olive nucleus. Each output Purkinje cell receives one and only one climbing fiber. The long term effect of this error signal is a decrease in the synaptic efficiency between

parallel fibers, axons of granular cells, and the output Purkinje cells, whenever the parallel fiber activity is correlated with the error signal [42], [1], [16], [29]. The physiological properties of this long term depression have been described in detail [33], [31].

An important feature of the cerebellar design is the great number (10^{11}) of granular cells in the intermediate layer, which is of the same order of magnitude as the total number of cells in the nervous system. The role of this architecture is to provide an extended set of new combinations of inputs which are needed for bypassing the mathematical limitations of the classical two-layer perceptron for learning any arbitrary input/output function [42]. Without these combinations, wired at the glomerulus level, many functions cannot be learned (see the classical example of the XOR problem given by [43]).

The functional role of each microzone is then defined by the climbing fiber error signals which originate from a restricted part of the inferior olive.

4.3.2 *Cerebro-Cerebellar Cooperation*

During motor tasks, the cerebral cortex rapidly learns the appropriate motor commands for reaching a goal by supplying the spinal servomechanisms. When a task is performed repeatedly, the cerebellum gradually intervenes, finely adjusting and automatizing the motor response. The cerebellum thus helps the cerebral cortex during goal directed tasks [26] and during mental control of more cognitive tasks [39].

In the analysis of motor control, cybernetic models have been proposed in order to explain the respective roles of cerebral and cerebellar cortices and have been correlated with experimental data [29]; [38]; [32].

It is assumed (Fig. 4.1) that the cerebral cortex first learns an inverse dynamics model of the skeleto-muscular apparatus in order to translate a desired trajectory into the appropriate commands, corrections being effected through a feedback loop. Later, the cerebellar cortex builds a feedforward control which replaces this closed loop cerebral process. Two successive phases must be distinguished in this cerebellar take-over.

A) First, the external feedback loop with which the movement can be corrected is replaced by a 'direct dynamics model' of the skeleto-muscular apparatus which makes it possible to predict and anticipate such corrections. The functioning of this internal model imitates the functioning of the real skeleto-muscular apparatus. The internal model may be built up thanks to the adaptive mechanisms of the cerebellum.

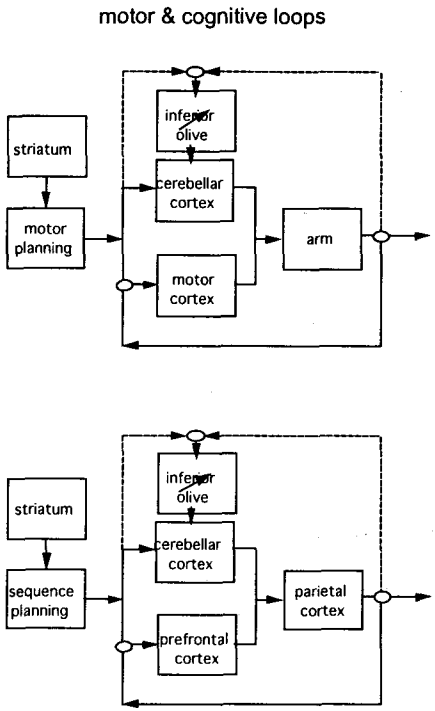


Fig. 4.1 Motor (top) and Cognitive Processes (bottom). Motor control involves a dual control: a feedback mode (cerebral) and a feedforward mode (cerebellar). The cerebral mode of control is faster but less accurate and less precise than the cerebellar one. Error signals result from the difference between desired and actual movement, executed (top) or planned (bottom). Cognitive processes (bottom) involves relations between lateral cerebellum and prefrontal cerebral coxtex, both structures "exploding" during philogenetic evolution. Role of the striatum is described in sec. 4.3 (figure modified from [32]).

At the end of this learning phase, in addition to the inverse dynamics model already learned by the cerebral cortex to command the movement, the cerebellum has learned the dynamics model which allows anticipatory corrections because it simulates the movement and at each step of the command, predicts the mechanical effect, without or before sending any command to the muscular apparatus.

B) Secondly, with practice, the performance of the command system performed by the cerebral cortex can be taken over by the cerebellum in a feedforward mode. In this case, the feedforward system can learn an inverse dynamics model of the skeleto-muscular apparatus, since the whole system is designed to make an actually performed trajectory equal to a desired trajectory. Numerical simulation has shown that this second phase produces a smooth, efficient trajectory [38]. This is probably learned by the lateral cerebellum, which progressively implement the inverse dynamics model previously learned by the cerebrum. The result is a completely automatic control of the movement, which will free the computational capacity of the cerebral cortex for other tasks.

These two phases are not necessarily separate but may overlap considerably. The distinction between their roles is more in terms of the relative importance in their behavioral effects, such as refining or automatizing a movement. In the first case, a representation of the effects of a command (direct dynamics) is learned. In the second case, a representation of the command itself (inverse dynamics) is learned.

The cortico-cortical processes are slower than the cerebellar ones since the response of a column is a decision process elaborated through the progressive recruitment of neuronal activities within the column; whereas the input:output relationships in the cerebellum result from a more direct process with no recruitment delays. If the thalamic input to the motor cortex suffices for an appropriate command to occur, this command will inhibit the subsequently useless cortical activities. The microcolumns influencing the motor cortex will progressively decrease their influence, since they are no longer needed to produce the appropriate motor commands. The error signal disappears with the progressive decrease in the redundant cortical influences on the motor cortex. As a result, learning is completely automatized and stored in the cerebellar circuits. In this way, the cerebellum has learned a function which was previously performed by the cerebral cortex. This function corresponds to the inverse dynamics model of the skeleto-muscular apparatus formerly held to operate at the cerebral level.

At the cerebellar level, a vector representation also shows a similar broad tuning curve in the case of individual cells and an overall activity varying with movement direction [14]. In a population of cerebellar neurons including cortical Purkinje and unidentified cells, the distribution of preferred directions covered all the possible movements. The proportion of directional versus nondirectional cells was constant over the four cell populations. Despite the metric hand movement deficits which occur following cerebellar damage [24], it is not

yet known whether some of the cerebellar cells are involved in the movement kinematics only. Oppositely, all the directional cells would be mainly involved in load compensation, i.e. they would contribute to the transformation between the kinematic and dynamic aspects of the arm movement, by taking into account postural perturbations.

4.3.3 *Connectionist Model*

A voluntary movement comprises several components [46]. It generally consists of a combination of muscle actions which make it possible to change the relative positions of body segments, in order to reach a goal. In addition, these displacements are accompanied by a set of postural adjustments. Due to the spring properties of muscles [25], the forces involved in the movement perturb the position of postural segments and may cause a loss of stability and balance. Learning the co-ordination between posture and movement is therefore essential for efficient motor performance to be possible [41].

The learning of the whole command is performed in parallel by both cortices and can be analyzed in terms of a dialogue between cerebral and cerebellar cortices, as described in Fig. 4.2. During this dialogue, each structure is guiding the processing and learning is taking place in the other structure. Here a distinction is made between the two phases described above.

During the first phase (Fig. 4.2, top), the various muscular actions which are necessary to reach a goal are learned in the cerebral cortex from the changes in the connective coefficients between microcolumns, since the learning rules in this tissue are well suited for selecting the appropriate patterns of local actions which produce the expected result.

If a goal is signalled by a strong activity in a subset of microcolumns in an associative cortex, learning will consist of increasing the connective coefficients between a subset of cerebral motor cells and another subset of premotor cells which commands the appropriate muscular contraction. These movements may produce several postural perturbations in other body segments. Displacement of a postural segment gives rise to an error message transmitted from the periphery to a specific cerebellar microzone which cooperates with the cerebral microcolumns of the motor cortex that are able to correct the error [5].

These cerebello-cerebral connexions can then guide activities in the corresponding microcolumns and produce the appropriate cerebral commands for postural corrections to be made. At this step, the postural adjustment is made in a closed loop mode by means of an error signal.

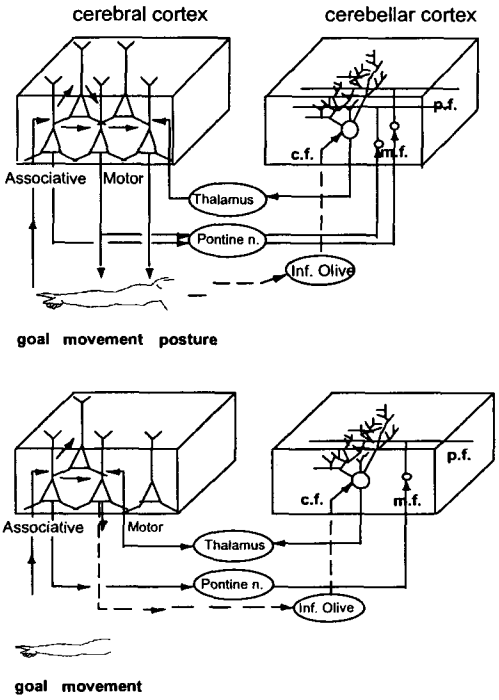


Fig. 4.2 Implementation of the Two Mechanical Models of Direct Dynamics (top), and Inverse Dynamics (bottom). Cerebral and cerebellar cortices cooperate during the learning of a reaching movement with postural shoulder control. Abbrev: cf, climbing fiber; mf, mossy fiber, pf, parallel fiber [5].

Within the cerebellar microzone which can participate in the corrections, two input signals interact on Purkinje cells: (1) mossy fiber activities and parallel fiber inputs conveying the signals from the cerebral cortex which are involved in the anticipatory aspects of the command; (2) error signals coming through the inferior olive from perturbed postural segments. This interaction will progressively shape the Purkinje cell response to the specific pattern of anticipatory inputs, until it suppresses the error signal. The movement will then be controlled in a feedforward mode by the cerebellar microzone.

At this step, the microzones have acquired an internal model of the effects of the voluntary movements and in this sense, the cerebellum has learned a direct dynamics model of the skeleto-muscular apparatus. At the end of this learning

phase, the appropriate motor commands are adequately tuned but the cerebral cortex is still in charge of the task.

Cerebral loading is not necessary here because it corresponds to a redundant control of the motor commands expressed by a whole set of cortico-cortical interactions which have favored the movement learning. The cerebellar learning can suffice to produce the general command and progressively replace the cortico-cortical influences on the microcolumns of the motor cortex. The output activities of these columns will therefore be due only to the cerebellar influences provided by the thalamic inputs.

The second phase (Fig. 4.2, bottom) of learning process can take place if the differences between the thalamic input and the motor cortex output give rise to an error signal which is sent to the inferior olive and then to the cerebellum through the climbing fiber system. These signals will progressively change the input-ouput transfer function in the microzone and consequently, the thalamic input to the motor cortex. The microzones involved are probably different from those involved in the first phase of learning, laying in a more lateral region of the cerebellar cortex.

4.3.4 *Origin of Error Signals*

How the olivary error signal needed for the adaptive process of a given Purkinje cell, can downsteam modify the motoneuronal activities, in order to produce an error correction which will then decrease the error signal of this specific Purkinje cell? More shortly, how error signals are generated? This question remains to be answered. We developed a hypothesis based on the connectionist idea of parsimony, trying to minimize the need of any genetic hardware in models.

At least, three successive although largely overlapping phases in the learning of compound movements exist (sec. 4.2.2). They result from the differences in the rates at which the different brain structures adapt. They are assumed to be the cerebro-spinal, the cerebello-cerebral and the cerebello-rubro-spinal pathways. During the second learning phase, the cerebellum fine-tunes the cerebral processes learned in the initial phase. Even though each cerebellar longitudinal microzone projects over a wide cerebral zone, a given Purkinje cell will only reinforce the strongly active pyramidal cells related to the actual or intended movement, by heterosynaptic long-term potentiation and long-term depression plasticity at the cerebral level. As a result, this cerebellar effect increases the actual cerebral output, unloading the cerebrum from its internal recurrent proc-

essing task [10].

Let us examine three examples. First, the increase in climbing fiber activity that occurs at the onset of a fast ballistic movement [40] may indicate: 'movement is not fast enough!'. By cerebellar disinhibition, it will increase the activities of pyramidal cells involved in the motor command. This climbing activity will never vanish, even after long training [50].

Secondly, any unexpected painful event, such as 'running into an obstacle', is also an olivary-mediated error. This repeated signal may gradually increase the cerebral control of the muscular braking process. A fairly small number of cerebellar modules is sufficient to modify the concomitant cerebral motor activities, wherever is the source of pain.

Thirdly, the cerebellum may contribute to the straight line hand path formation. [9] showed that a single global constraint of mimimum muscle tension change is sufficient to solve the four so-called 'ill posed' problems of arm movements: hand path and trajectory formations, coordinate transformation, and the calculation of muscle tensions. Olivary error signals may simply originate from a global summation of fusorial informations given by arm muscles, since each piece of fusorial information indicates a mismatch between the actual (muscular fiber) and the desired (gamma fusimotor) lengths [27], and each Golgi tendon afferent indicates a force change [34].

This conceptual framework does away the 'mystery' of error signal generation. The putative cerebral-cerebellar mechanism first selects the target pyramidal cells, then produces the motor error correction, and later reduces the error signal. No hardware is needed and all olivary afferents are potential error signals. Central or peripheral signals topographically project to the inferior olive, then to related cerebellar cortical beams, and further to their learned cerebral targets.

Cerebral overloading during movement or succession of movements may be the origin of error signals in the most lateral cerebellar modules [5], [15].

4.4 Basal Ganglia : the 'Security' Computer

The basal ganglia consist of five extensively interconnected subcortical nuclei that participate in the control of movement: the caudate nucleus, putamen, globus pallidus, subthalamic nucleus, and substantia nigra. The latter nucleus is itself divided in two heterogeneous parts: the substantia nigra pars reticulata and the substantia nigra pars compacta. They receive input from and project to the cortex by way of the thalamus. Almost all afferents to the basal ganglia

terminate in the neostriatum, which group the caudate and putamen nuclei [37].

In a modelling approach, only three set of structures are taken into account: 1) a main cortical-like processing layer, 2) an output layer and, 3) a nucleus which gives rise to reinforcement signals. Respectively, the three corresponding nervous structures are: 1) the neostriatum (caudate and putamen), 2) the globus pallidus (internal and external), together with the substantia nigra pars reticulata, and 3) the substantia nigra, pars compacta.

Despite the assembly of nervous nuclei, which is much more complex than the two cortical structures, the basal ganglia can be modelled as a three layer neural network. The first layer consists of the cerebral pyramidal cells sending projections to the neostriatum. The second layer consists of the neostriatum target cells. The third layer consists of cells located in both the globus pallidus and the pars reticulata of the substantia nigra, which project back via the thalamus to some restricted areas of the cerebral cortex.

Three main differences with the cerebellum should be noted. First, the cerebral input areas projecting to the basal ganglia are much wider than to the cerebellum. Secondly, the cerebral areas receiving from the basal ganglia are much more restricted than the cerebral areas receiving from the cerebellum. Third, the basal ganglia project almost exclusively to the cerebral cortex, as compared to the cerebellum which also projects to subcortical nuclei (the red nucleus, for example), which send signals to the periphery (the rubrospinal pathway, respectively).

The globus pallidus considered as the third output layer, is divided into two parts: the internal and external segments. Because of the striking similarities in cytology, connectivity, and function of the internal segment of the globus pallidus and the substantia nigra pars reticulata, these two nuclei can be considered as a single structure arbitrary divided by the internal capsule, much like the caudate and putamen [37]. They formed a kind of bistable ('push-pull') circuit. The first is aimed at activating a first set of cerebral motor cells, the second at inhibiting another set (largely overlaping).

In a more formal approach, the dual role of the motor part of the basal ganglia is to stabilize the ongoing movement and in turn to optimally destabilize the ongoing movement by selecting the postural and the moving joints for the next forthcoming movement [30]. Experimental data obtained in the cerebral 'supplementary motor area' (SMA), the major cerebral projecting area of the motor part of the basal ganglia support this idea [3], [54].

The "second layer" formed by the neostriatal spiny cells, consists of two nuclei, putamen and caudate. The putamen nucleus is more related to motor

system and the caudate nucleus more related to cognitive processes. Considered as the major structure, the neostriatum is approximated by a simple three-dimensional conglomeration of medium-sized spiny neurons, each receiving cortical input and sending it outward to the pallidum and the substrantia nigra. Interneurons comprise such a small portion of the neostriatal population that their numbers may be considered negligible [22].

Any model should take into account the basic striatal processing, which underlyies the arousing function of striatum, the psychophysiological context in which it takes place, and the message it can transmit to the motor system [8].

Within the striatum (Table 4.1), some medium spiny neurons, lying within more tightly packed aggregations of cells (islands) and situated in a milieu poor of acelytcholinestarase (striosomes) and rich in opiate receptors (patches), may be distinguished from all other medium spiny neurons lying in the surrounding matrix [22]. Recent evidence suggests that deep and superficial layers cortical layers within every cerebral cortical area innervate patch and matrix compartments, respectively [21].

striatum	striosomes	matrisomes
anatomy	patch / islands	surrounding matrix
cortex origin	infra (deep) layers	supra-granular layers
subs. nigra projections	pars compacta	pars reticulata
limbic afferences	amygdala cross-modal & affect learning	hippocampus spatial & factual memory

Table 4.1 Main Differences between the Striosomes and the Surrounding Matrix (matrisome). In this table, we insist upon the various levels of understanding: hardware (anatomy), computation (cerebral models), plasticity (reinforcement role of the pars compacta), and motivation for learning (limbic).

4.5 Conclusion

For Engineers and for NeuroScientits, knowledge of learning processes is more important that the study of the functional capacities of artificial or living 'machines'. In the present review we showed that several nets with different learning rules must cooperate for the learning of the movement.

	cerebral cortex	cerebellum	striatum
arrangement	cortico-cortical + thalamic input	mossy & climbing fibers	cortico-striatal + dopamine
functionning	combinatory operations	pavlovian conditioning	operant conditioning
mode	construction of sequences	error reduction	reinforcement
processing	recurrent cortico-cortical	parallel feedforward microzones	parallel feedforward modules
activities	sustained activity during delays	context dependent modulation	context dependent stabilization
goal	"fuzzy logic"-like qualitative	adaptive gain control	bistable pallidus ext. / int.
role	attentional control&planning	"control augmentation"	"stabilization augmentation"

Table 4.2 Complementary Roles of the Three Major Neural Structures: 1) a cerebral one which learns a goal, 2) a cerebellar one which continuously, finely adjust the motor commands, and 3) a basal ganglia one which optimally adjusts the time series of transitions between successive elementary actions, and stabilizes the processes between transitions.

When similar constraints are applied to artificial or living 'machines', such as the Newtonian Law, we do believe that the common goal between NeuroScientists and Engineers is to find a common solution. For Engineers, the solution is to solve a technical problem, such as the 'fly-by-wires' technique which allows to manage the six degrees of freedom of an airplane, but must secure his stability without the help of its pilot. For NeuroScientists, the solution is to get a deeper understanding of brain structures and of their cooperation during

movement learning. Here we spoke of the cerebral cortex as the pilot, of cerebellum for smooting the commands and of the basal ganglia for security (1- postural stability of the eyes or the trunck when arms or legs moves, 2- destabilizing this set for the mext ongoing planned movement).

More practically, we try to simplify this model for a smooth robot, buidt on an electrical wheel-chair for handicadded persons, helping them to open a door or to grasp a glass.

As an example (Fig. 4.3), the four models involved in movement control (cerebral, cerebellar, striatal, servo-spinal) are gathered to the understanding of the cooperation between brain structures during the learning and the execution of a bimanual load lifting task.

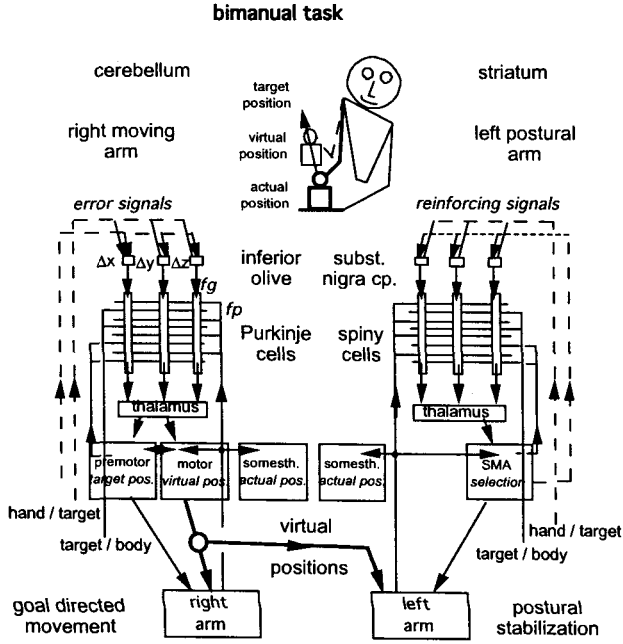


Fig. 4.3 The Equilibrium Point Model Takes into Account the Muscular Spring-like Properties. During a bimanual "load lifting task", when the right hand lifts a load, the left arm is stabilized by anticipatory postural adjustments. Three cerebral areas are related to the target (premotor), the virtual or equilibrium (motor) and the actual (somesthetic) positions. Cerebellar and striatal perceptron-like networks helps cerebral internal processing.

Modelling approach may explain: 1) how adaptive control is progressively transferred from cerebral to cerebellar and striatal levels during the acquisition of the co-ordination between posture and movement, and 2) how the whole movement could be automatized with a minimal load for the cerebral cortex. Each subsystem is essential for fast learning to occur and participates in the final result. If we attempt to establish a parallel with artificial systems, the cerebral cortex can be said to be a multiprocess but single-task (attentional) "central processing unit", devoted to a "foreground" task selected by attentional mechanisms.

This central processing unit needs both "mass memory" and "computational power", that are provided by the extraordinary storage capacity available thanks to the perpendicular arrangement of the cerebellar cortex and to a similar striatal architecture. These computer-like devices are content-address associative memories, consisting of multiple modules working in parallel in order to detect, memorize and automatize the repetetitive tasks processed by the central cerebral unit, which itself try to minimize the cost of the attentional "load".

References

[1] Albus, J., A theory of cerebellar function. Math. Biosci. 10, pp.25-61, (1971).

[2] Alexandre, F., Burnod, Y., Guyot, F., Haton, J.P., La colonne corticale, nouvelle unité de base pour les réseaux multicouches. C.R.Acad.Sci. Paris 309, III, pp.259-264, (1989).

[3] Brinkman, C., Lesions in supplementary motor area interfere with a monkey's performance of a bimanual coordination task. Neurosci. Lett., 27, pp.267-270, (1981).

[4] Burnod, Y., An adaptive neural network: the cerebral cortex, Masson, Paris, (1989).

[5] Burnod, Y., Dufossé, M., A model for the co-operation between cerebral cortex and cerebellar cortex in movement learning, In: Brain and Space, J. Paillard (Ed.), Oxford University Press, pp.446-458, (1990).

[6] Burnod, Y., Caminiti, R., Johnson, P., Granguillaume, P., Otto, I., Model of visuomotor transformations performed by the cerebral cortex to command arm movements at visual targets in the 3-D space. J. Neurosci, (1991).

[7] Caminiti, R., Johnson, P.B., Galli, C., Ferraina, S., Burnod, Y., Urbano, A., Making arm movements within different parts of space: the premotor and motor cortical rep-

resentation of a coordinate system for reaching to visual targets. J. Neurosci, (1991).

[8] Chevalier, G., Deniau, J.M., Disinhibition as a basic process in the expression of striatal functions. TINS 13, 7, pp.277-280, (1990).

[9] Dornay, M., Uno, Y., Kawato, M., Suzuki, R., Minimum muscle tension change trajectories. J. Motor Behav., (1996).

[10]Dufossé, M., How the cerebellum can match "error signal" and "error correction"? Brain Behav. Sci. (in press) Neural Network World 1996, 4, pp.545-551, (1996).

[11]Dufossé, M., Ito, M., Jastreboff, P., Miyashita, Y., A neuronal correlate in rabbit's cerebellum to adaptive modification of the vestibulo-ocular reflex. Brain Res., 150, pp.611-616, (1978).

[12]Evarts, E.V., Activity of pyramidal tract neurons during postural fixations. J. Neurophysiol. 32, pp.375-385, (1969).

[13]Feldman, A.G., Once more on the equilibrium-point hypothesis (l-model) for motor control. J. Motor Behav., 18, pp.17-54, (1986).

[14]Fortier, P.A., Kalaska, J.F., Smith, A.M., Cerebellar neuronal activity related to the whole-arm reaching movements in the monkey. J. Neurophysiol. 62, pp.198-211, (1989).

[15]Frolov, A.A., Roschin, V.Y., Biryukova, E.V., Adaptive neural model of multijoint movement control by working point analysis. Neural Network World 4, pp.141-156, (1993).

[16]Fujita, M., Adaptive filter model of the cerebellum. Biol. Cybern., 45, pp.195-206, (1982).

[17]Gellman, R.S., Miles F.A., A new role for the cerebellum in conditioning? TINS, 8, pp.181-182, (1985).

[18]Georgopoulos, A.P., Kalaska, J.F., Caminiti, R., Massey, J.T., On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. J. Neurosci., 11, pp.1527-1537, (1982).

[19]Georgopoulos, A.P., Schwartz A.B., Kettner R.E., Neuronal population coding of movement direction. Science 233, pp.1416-1419, (1986).

[20]Georgopoulos, A.P., Kettner, R.E., Schwartz, A.B., Primate motor cortex and free arm movements to visual targets in three-dimensional space. II. Coding of the direction of movement by a neuronal population. J. Neurosci., 8, pp.2928-2937, (1988).

[21]Gerfen, C.R., The neostriatal mosaic: Striatal patch-matrix organization is related to cortical lamination. Science 246, pp.385-388, (1989).

[22]Goldman-Rakic, P.S., Selemon, L.D., New frontiers in basal ganglia research. TINS 13, 7, pp.214-243, (1990).

[23]Gilbert, P.F.C., Thach, W.T., Purkinje cell activity during motor learning. Brain Res.

128, pp.309-328, (1977.)

[24]Holmes, G., The cerebellum of man. Brain 62, pp.1-30, (1939).

[25]Houk, J.C., Rymer, W.Z., Neural control of muscle length and tension. In: Handbook of Physiology, Brooks, V.B. (Ed.) Section I, Vol 2, pp.257-323, (1981).

[26]Holmes, G., The cerebellum of man. Brain 1, 62, pp.1-30, (1939).

[27]Houk, J.C., Rymer, W.Z., Neural control of muscle length and tension. In: Handbook of physiology, Brooks, V.B. (Ed.) Section I, Vol 2, pp.257-323, (1981).

[28]Hubel, D.H., Wiesel, T.N., Stryker M.P., Anatomical demonstration of orientation columns in macaque monkey. J. Comp. Neurol. 177, pp.361-380, (1978).

[29]Ito, M., The cerebellum and neural control, Raven Press, New York, (1984).

[30]Ito, M., Neural systems controlling movement. TINS oct-1986, pp.515-518, (1986).

[31]Ito, M., Long term depression. Annu. Rev. Neurosci. 12, pp.85-102, (1989).

[32]Ito, M., New concepts in cerebellar function. Rev. Neurol. 149, 11, pp.569-599, (1993).

[33]Ito, M., Sakurai, M. Tongroach, P., Climbing fiber induced depression of both mossy fiber responsiveness and glutamate sensitivity of cerebellar Purkinje cells. J. Physiol., London, 324, pp.113,134, (1982).

[34]Jami, L., Golgi tendon organs in mammalians skeletal muscle: functional properties and central actions. Physiol. Rev. 72, pp.623-666, (1992).

[35]Kalaska, J.F., Cohen, D.A.D., Hyde, M.L., Prud'homme, M., A comparison of movement direction-related versus load direction-related activity in primate motor cortex, using a two-dimensional reaching task. J. Neurosci. 9, pp.2080-2102, (1989).

[36]Kalaska, J.F., Cohen, D.A.D., Prud'homme, M., Hyde, M.L., Parietal area 5 neuronal activity encodes movement kinematics, not movement dynamics. Exp. Brain Res. 80, pp.351-364, (1990).

[37]Kandel, E.R., Schwartz, J.H., Jessel, T.M., Principles of Neural science. Prentice Hall Int. Inc. chap 42, pp.647-659, (1991).

[38]Kawato, M., Furukawa, K., Suzuki, R., A hierarchical model for control and learning of voluntary movement. Biol. Cybern. 57, pp.169-185, (1987).

[39]Leiner, H.C., Leiner, A.L., Dow, R.S., Reappraising the cerebellum: What does the hindbrain contribute to the forebrain. Behav. Neurosci. 103, 5, pp.998-1008, (1989).

[40]Mano, N., Kanazawa, I., Yamamoto, K., Complex-spike activity of cerebellar Purkinje cells related to wrist tracking movement in monkey. J. Neurophysiol. 56, pp.137-158, (1986).

[41]Massion, J., Dufossé, M., Co-ordination between posture and movement: Why and How? NIPS 3, pp.88-93, (1988).

[42]Marr, D., A theory of cerebellar cortex. J. Physiol. (London) 202, pp.437-470, (1969).

[43]Minsky, M., Papert, S., Perceptrons, MIT press, Cambridge, USA, (1969).

[44]Mountcastle, V.B., An organizing principle for cerebral function: the unit module and the distributed system. In: The mindful brain, F.O. Schmidt (Ed), MIT Press, Cambridge, USA, (1978).

[45]Oscarsson, O., The sagittal organization of the cerebellar anterior lobe as revealed by the projection patterns of the climbing fiber system. In: Neurobiology of cerebellar evolution and development, R. Llinas (ed.), American Medical Association, Chicago, (1969).

[46]Paillard, J., Apraxia and the neurophysiology of motor control. Phil. Trans. R. Soc. Lond. B-298, pp.111-134, (1982).

[47]Rosenblatt, F., The perceptron: a probabilistic model for information storage and organization in the brain. Psychol. Rev. 65, pp.386-408, (1958).

[48]Shepherd, G.M., The synaptic organization of the brain. Oxford Univ. Press, Oxford, (1969).

[49]Schmidt, R.A., A schema theory of discrete motor skill learning. Psychol. Rev. 82, 4, pp.225-260, (1975).

[50]Smith, A.M., What do studies of specific motor acts such as reaching and grasping tell us about the general principles of goal-directed motor behavior? In: Motor control: concepts and issues, Humphrey & Freund (eds), (1991).

[51]Smith, A.M., Hepp-Reymond M.C., Wyss U.R., Relation of activity in precentral cortical neurons to force and rate of force change during isometric contractions of finger muscles. Exp. Brain Res. 23, pp.315-332, (1975).

[52]Szentagothai, J., The 'module-concept' in cerebral cortex architecture. Brain Res. 95, pp.475-496, (1975).

[53]Thach, W.T., Correlation of neural discharge with pattern and force of muscular activity, joint position and direction of intended next movement in motor cortex and cerebellum. J. Neurophysiol. 41, pp.654-676, (1978).

[54]Viallet, F., Massion, J., Massarino, R., Khalil, R., Coordination between posture and movement in a bimanual unloading task: putative role of a medial frontal region including the SMA. Exp. Brain Res. 88, pp.674-684, (1992).

[55]Widrow, B., Hopf, M.E., Adaptive switching circuits. IRE Wescon Cov. Record 4, pp.96-104, (1960).

This page is intentionally left blank

Chapter 5

Brain-like Functions in Evolving Connectionist Systems for On-line, Knowledge-Based Learning

Nikola Kasabov
University of Otago

Abstract

The paper discusses some biological principles of the human brain that would be useful to implement in intelligent information systems (IS). These principles are used to formulate seven major requirements to the current and the future IS. These requirements are met in a new connectionist architecture called evolving connectionist systems (ECOS). ECOS are designed to facilitate building on-line, adaptive, knowledge-based IS. ECOS evolve through incremental, hybrid (supervised/unsupervised), on-line learning. They can accommodate new input data, including new features, new classes, etc. through local element tuning. New connections and new neurons are created during the operation of the system. The ECOS framework is presented and illustrated on a particular type of evolving neural networks - evolving fuzzy neural networks (EFuNNs). EFuNNs can learn spatial-temporal sequences in an adaptive way, through one pass learning. Rules can be inserted and extracted at any time of the system operation. ECOS and EFuNNs are suitable for adaptive pattern classification; adaptive, phoneme-based spoken language recognition; adaptive dynamic time-series prediction; intelligent agents.

Keywords : evolving connectionist systems, evolving fuzzy neural networks, on-line learning, spatial-temporal adaptation

5.1 Introduction: What Brain-like Functions and Principles to Implement in Intelligent Information Systems?

The human brain proved to be the best computational mechanism for many tasks, such as speech and language processing, image processing, navigation, control. One of the most important characteristics of the brain is its ability to learn in an on-line mode, in a lifelong mode, to adapt quickly, to make abstractions and represent them as knowledge, to evolve its structure and functions

during its lifetime in an interactive way and its innate way.

The following are some principles of the evolving brains:

1. Evolving is achieved through both genetically defined information and learning;
2. The evolved neurons have a spatial-temporal representation where similar stimuli activate close neurons;
3. Redundancy, i.e. there are many redundant neurons allocated to a single stimulus or a task; e.g., when a word is heard, there are hundreds of thousands of neurons that get immediately activated;
4. Memory-based learning, i.e. the brain stores exemplars of facts that can be recalled at a later stage;
5. Evolving through interaction with the environment and with other brains;
6. Inner processes, based on information theory, take place; these processes can be described as an "*instinct for information*"; they are based on information entropy and cause the brain to acquire information.
7. The evolving process is continuous, lifelong.
8. Evolving higher level functions, cognition and intelligence, i.e. higher-level concepts emerge that are embodied in the structure and can be represented as a level of abstraction at any time of the evolving process, e.g. acquisition and the development of speech and language, especially in multilingual subjects.
9. Evolving 'global brains' through interaction of individuals, i.e. an individual brain acts as an 'agent' in a collaborative environment of other agents. A collection of agents can be viewed as a 'global brain', that improves in a continuous, endless way, with the emergence of new individuals.

It is known that the human brain develops even before the child is born. During learning the brain allocates neurons to respond to certain stimuli and develops their connections [72,77,80]. Evolving is achieved through both genetically defined information and learning. The learning and the structural evolution coexist in ECOS. That is plausible with the co-evolution of structure and learning in the brain. The neuronal structures eventually implement a long-term memory. Biological facts about growing neural network structures through learning and adaptation are presented in [80, 82].

The observation that humans (and animals) learn through memorising sensory information and then remembering it when interpreting it in a context-driven way belongs to Helmholtz (1866). This is demonstrated in the consolidation principle that is widely accepted in physiology. It states that what has happened

in the first 5 or so hours after presenting input stimulus the brain is learning to 'cement' what has been learned. This has been used to explain retrograde amnesia (a trauma of the brain that results in loss of memory about events that occurred several hours before the event of the trauma).

During the learning process, exemplars (or patterns) are stored in a long-term memory. Using stored patterns is the bases for the Task Rehearsal Mechanism (TRM) [55]. The TRM assumes that there are long term and short term centers for learning. The TRM relies on long-term memory for the production of virtual examples of previously learned task knowledge (background knowledge). A functional transfer method is then used to selectively bias the learning of a new task that is developed in short-term memory. The representation of this short-term memory is then transferred to long-term memory where it can be used for learning yet another new task in the future. Notice, that explicit examples of a new task need not be stored in long-term memory, only the representation of the task which can be later used to generate virtual examples. These virtual examples can be used to rehearse previously learned tasks in a concert with a new 'related' task". But if a system is working in a real-time mode, it may not be able to adapt to new data if its speed of processing is 'too, when compared to the speed of the continuously incoming information. This phenomenon is known in psychology as "loss of skills". The brain has a limited amount of working or short term memory. And when encountering important new information, the brain stores it simply by erasing some old information from the working memory. The prior information gets erased from the working memory before the brain has time to transfer it to a more permanent or semi-permanent location for actual learning. These issues are also discussed in [55, 66].

The complexity and dynamics of real-world problems, especially in engineering and manufacturing, require sophisticated methods and tools for building on-line, adaptive intelligent systems (IS). Such systems should be able to grow as they operate, to update their knowledge and refine the model through interaction with the environment. This is especially crucial when solving AI problems such as adaptive speech and image recognition, multi-modal information processing, adaptive prediction, adaptive on-line control, intelligent agents on the WWW.

The above described biological principles and functions are used here to specify seven major principles of the current and the future intelligent information systems (IS). They are addressed later in the presented framework for evolving connectionist systems ECOS. These are:

(1) IS should *learn fast* from a large amount of data (using fast training, e.g. one-pass training).
(2) IS should be able to *adapt incrementally* in both real time, and in an on-line mode, where new data is accommodated as they become available. The system should tolerate and accommodate imprecise and uncertain facts or knowledge and refine its knowledge.
(3) IS should have an *open structure* where new features (relevant to the task) can be introduced at a later stage of the system's operation. IS should dynamically create new modules, new inputs and outputs, new connections and nodes. That should occur either in a supervised, or in an unsupervised mode, using one modality or another, accommodating data, heuristic rules, text, images, etc.
(4) IS should be *memory-based,* i.e. they should keep a reasonable track of information that has been used in the past and be able to retrieve some of it for the purpose of inner refinement, or for answering an external query.
(5) IS should improve continuously (possibly in a life–long mode) through active *interaction* with other IS and with the environment they operate in.
(6) IS should be able to *analyse themselves* in terms of behaviour, global error and success; to explain what has been learned; to make decisions about its own improvement; to manifest introspection.
(7) IS should adequately represent *space and time* in their different scales; should have parameters to represent such concepts as spatial distance, short-term and long-term memory, age, forgetting, etc.

Several investigations [18, 28, 43, 55, 65, 66, 67, 69, 74] proved that the most popular neural network models and algorithms are not suitable for adaptive, on-line learning, that includes multilayer perceptrons trained with the backpropagation algorithm, radial basis function networks [58], self-organising maps SOMs [47, 48] and these NN models were not designed for on-line learning in the first instance. At same time some of the seven issues above have been acknowledged and addressed in the development of several NN models for adaptive learning and for structure and knowledge manipulation as discussed below.

Adaptive learning is aiming at solving the well-known stability/plasticity dilemma [3, 4, 7, 8, 9, 13, 47, 48]. Several methods for adaptive learning are related to the work presented here, namely incremental learning, lifelong learning, on-line learning.

Incremental learning is the ability of a NN to learn new data without destroying (or at least fully destroying) the learned patterns from old data, and without a need to be trained on the whole old and new data. Significant pro-

gress in incremental learning has been achieved due to the Adaptive Resonance Theory (ART) [7, 8, 9] and its various models, that include unsupervised models (ART1, ART2, FuzzyART) and supervised versions (ARTMAP, Fuzzy ARTMAP- FAM). Lifelong learning is concerned with the ability of a system to learn during its entire existence in a changing environment [82, 69, 35, 36]. Growing, as well as pruning operation, are involved in the learning process. On-line learning is concerned with learning data as the system operates (usually in a real time) and the data might exist only for a short time. NN models for on-line learning are introduced and studied in [1, 2, 4, 7, 11, 17, 22, 28, 31, 35, 36, 42, 44, 46, 53, 69].

The issue of NN structure, the bias/variance dilemma, has been acknowledged by several authors [6, 7, 13, 65, 68]. The dilemma is concerned with the situation where if the structure of a NN is too small, the NN is biased to certain patterns, and if the NN structure is too large there are too many variances that result in over-training, and poor generalisation, etc. In order to avoid this problem, a NN (or an IS) structure should dynamically adjust during the learning process to better represent the patterns in the data from a changing environment. Three approaches have been taken so far for the purpose of creating dynamic IS structures: constructivism, selectivism, and a hybrid approach.

Constructivism is concerned with developing NNs that have a simple initial structure and grow during its operation through insertion of new nodes and new connections when new data items arrive. This approach can also be implemented with the use of an initial set of neurons that are sparsely connected and that become more and more wired with the incoming data [62, 73, 15, 19]. The latter implementation is supported by biological facts [62, 73, 77]. Node insertion can be controlled by either a similarity measure, or by the output error measure, or by both. There are other methods that insert nodes based on the evaluation of the local error, e.g. the Growing Cell Structure, Growing Neural Gas, Dynamic Cell Structure [19, 11, 13]. Other methods insert nodes based on a global error evaluation of the performance of the whole NN. Such method is the Cascade-Correlation [15]. Methods that use both similarity and output error for node insertion are used in Fuzzy ARTMAP [9]. Cellular automata systems have also been used to implement the constructivist approach [11, 4]. These systems grow by creating connections between neighbouring cells in a regular cellular structure. Simple rules, embodied in the cells, are used to achieve the growing effect. Unfortunately in most of the implementations the rules for growing do not change during the evolving process. This limits the adaptation of the growing structure. The brain-building system is an example of this class

[11].

Selectivism is concerned with pruning unnecessary connections in a NN that starts its learning with many, in most cases redundant, connections [26, 29, 49, 56, 59, 64]. Pruning connections that do not contribute to the performance of the system can be done by using several methods, e.g.: optimal-brain damage [50]; optimal brain surgeon [26]; structural learning with forgetting [29, 49]; training-and-zeroing [32]; regular pruning [56].

Genetic algorithms (GA) and other evolutionary computation techniques that constitute a heuristic search technique for finding the optimal, or near optimal solution from a solution space, have also been widely applied for optimising a NN structure [20, 23, 13, 39, 40, 71, 79, 80]. Unfortunately, most of the evolutionary computation methods developed so far assume that the solution space is compact and bounded, i.e. the evolution takes place within a pre-defined problem space and not in a dynamically changing and open one, therefore not allowing for continuous, on-line adaptation. The GA implementations so far have also been very time-consuming.

Some NN models use a hybrid constructivist/selectivist approach [52, 61, 70]. The framework proposed here also belongs to this group.

Some of the above seven issues have also been addressed in the knowledge-based neural networks (KBNN) [24, 33, 38, 63, 76, 83] as knowledge is the essence of what an IS system has learned. KBNN have operations to deal with both data and knowledge, that include learning from data, rule insertion, rule extraction, adaptation and reasoning. KBNN have been developed mainly as a combination of symbolic AI systems and NN [24, 30, 76], or as a combination of fuzzy logic systems and NN [25, 30, 33, 38, 39, 44, 45, 51, 63, 83], or as a combination of a statistical technique and NN [2, 4, 12, 57].

It is clear that in order to fulfil the seven major requirements of the current IS, radically different methods and systems are essential in both learning algorithms and structure development. A framework called ECOS (Evolving COnnectionist Systems) that addresses all seven issues above is introduced in the paper, along with a method of training called ECO training. The major principles of ECOS are presented in section 2. The principles of ECOS are applied in section 3 to develop evolving fuzzy neural network model called EFuNN. Several learning strategies of ECOS and EFuNNs are introduced in section 3. In section 4 ECOS and EFuNNs are illustrated on several case study problems of adaptive phoneme recognition, dynamic time series prediction, and intelligent agents. Section 5 suggests directions for further development of ECOS.

5.2 The ECOS Framework

Evolving connectionist systems (ECOS) are systems that evolve in time through interaction with the environment. They have some (genetically) pre-defined parameters (knowledge) but they also learn and adapt as they operate. In contrast with the evolutionary systems they do not necessarily create copies of individuals and select the best ones for the future. They emerge, evolve, develop, unfold through innateness and learning, and through changing their structure in order to better represent data [14, 31, 35, 36]. ECOS learn in an on-line and a knowledge–based mode, so they can accommodate any new incoming data from a data stream, and the learning process can be expressed as a process of rule manipulation.

A block diagram of the ECOS framework is given in Fig. 5.1. ECOS are multi-level, multi-modular structures where many neural network modules (denoted as NNM) are connected with inter-, and intra- connections. ECOS do not have a clear multi-layer structure, but rather a modular, "open" structure.
The main parts of ECOS are described below.

(1) Feature selection part. It performs filtering of the input information, feature extraction and forming the input vectors. The number of inputs (features) can vary from example to example from the input data stream fed to the ECOS.

(2) Presentation and representation (memory) part, where information (patterns) are stored. It is a multi-modular, evolving structure of NNM organised in spatially distributed groups; for example one module can represent the phonemes in a spoken language (one NN representing one class phoneme).

(3) Higher-level decision part that consists of several modules, each taking decision on a particular problem (e.g., phoneme, word, concept). The modules receive feedback from the environment and make decisions about the functioning and the adaptation of the whole ECOS.

(4) Action modules, that take the output from the decision modules and pass output information to the environment.

(5) Self-analysis, and rule extraction modules. This part extracts compressed abstract information from the representation modules and from the decision modules in different forms of rules, abstract associations, etc.

Initially an ECOS has a pre-defined structure of some NNMs, each of them being a mesh of nodes (neurons) and very few connections defined through

prior knowledge, or "genetic" information. Gradually, the system becomes more and more "wired" through self-organisation, and through creation of new NNM and new connections.

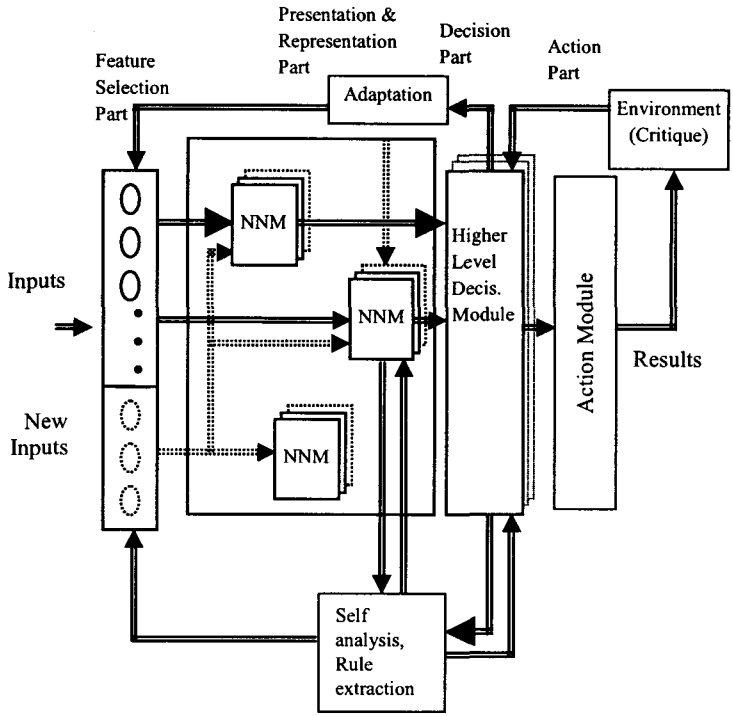


Fig. 5.1 Block Diagram of the ECOS Framework.

The ECOS functioning is based on the following *general principles*:

(1) ECOS evolve incrementally in an on-line, hybrid, adaptive *supervised/unsupervised mode* through accommodating more and more examples when they become known from a continuous input data stream. During the operation of ECOS the higher-level decision module may activate an adaptation process through the adaptation module.

(2) ECOS are memory-based and store exemplars (prototypes, rules) that represent groups of data from the data stream. New input vectors are stored in the NNMs based on their similarity to previously stored data both on the input and the desired output information. A node in an NNM

is created and designated to represent an individual example if it is significantly different from the previously used examples (with a level of differentiation set through dynamic parameters). Learning is based on *locally tuned* elements from the ECOS structure thus making the learning process fast for real-time parallel implementation. Three ways to implement local learning in a connectionist structure are presented in [6, 7, 47, 58].

(3) There are *three levels* at which ECOS are functionally and structurally defined:

(a) *Parameter (gene) level*, i.e. a chromosome contains genes that represent certain parameters of the whole systems, such as: type of the structure (connections) that will be evolved; learning rate; forgetting rate; size of a NNM; NNM specialisation, thresholds that define similarity; error rate that is tolerated, and many more. The values of the genes are relatively stable, but can be changed through genetic operations, such as mutation of a gene, deletion and insertion of genes that are triggered by the self analysis module as a result of the overall performance of the ECOS.

(b) *Representation (synaptic) level*, that is the information contained in the connections of the NNM. This is the long-term memory of the system where exemplars of data are stored. They can be either retrieved to answer an external query, or can be used for internal ECOS refinement.

(c) *Behavioural (neuronal activation) level*, that is the short-term activation patterns triggered by input stimuli. This level defines how well the system is functioning in the end.

(4) ECOS evolve through *learning (growing), forgetting (pruning), and aggregation,* that are both defined at a genetic level and adapted during the learning process. ECOS allow for: creating/connecting neurons; removing neurons and their corresponding connections that are not actively involved in the functioning of the system thus making space for new input patterns to be learned; aggregating nodes into bigger-cluster nodes.

(5) There are two global modes of learning in ECOS:

(a) *Active learning* - learning is performed when a stimulus (input pattern) is presented and kept active.

(b) *Passive (inner, ECO) learning mode* - learning is performed when there is no input pattern presented to the ECOS. In this case the process of further elaboration of the connections in ECOS is done in a passive learning phase, when existing connections, that store previously fed input patterns, are used as "echo" (here denoted as ECO) to reiterate the learning process (see for example Fig. 5.9 explained later).

There are two types of ECO training:

- *cascade eco*-training: a new connectionist structure (a NN) is created in an on-line mode when conceptually new data (e.g., a new class data) is presented. The NN is trained on the positive examples of this class, on the negative examples from the following incoming data, and on the negative examples from previously stored patterns in previously created modules.
- *'sleep' eco*-training: NNs are created with the use of only partial information from the input stream (e.g., positive class examples only). Then the NNs are trained and refined on the stored patterns (exemplars) in other NNs and NNMs (e.g., as negative class examples).

(6) ECOS provide explanation information extracted from the NNMs through the self-analysis/ rule extraction module. Generally speaking, ECOS learn and store knowledge, rules, rather than individual examples or meaningless numbers.

(7) The ECOS principles above are based on some biological facts and biological principles (see for example [31, 55, 62, 68, 72, 82]).

Implementing the ECOS framework and the NNM from it requires connectionist models that comply with the ECOS principles. One of them, called evolving fuzzy neural network (EFuNN) is presented in the next section.

5.3 Evolving Fuzzy Neural Networks EFuNNs

5.3.1 *General Principles of EFuNNs*

Fuzzy neural networks are connectionist structures that implement fuzzy rules and fuzzy inference [25, 51, 63, 83, 38]. FuNNs represent a class of them [38, 33, 39, 40]. EFuNNs are FuNNs that evolve according to the ECOS principles. EFuNNs were introduced in [31,35,36] where preliminary results were given. Here EFuNNs are further developed.

EFuNNs have a five-layer structure, similar to the structure of FuNNs (Fig. 5.2a). But here nodes and connections are created/connected as data examples are presented. An optional short-term memory layer can be used through a feedback connection from the rule (also called, case) node layer (see Fig. 5.2b). The layer of feedback connections could be used if temporal relationships between input data are to be memorised structurally.

The input layer represents input variables. The second layer of nodes (fuzzy input neurons, or fuzzy inputs) represents fuzzy quantization of each input variable space. For example, two fuzzy input neurons can be used to represent

"small" and "large" fuzzy values. Different membership functions (MF) can be attached to these neurons (triangular, Gaussian, etc.) (see Fig. 5.3).

The number and the type of MF can be dynamically modified in an EFuNN which is explained later in section 3. New neurons can evolve in this layer if, for a given input vector, the corresponding variable value does not belong to any of the existing MF to a degree greater than a membership threshold. A new fuzzy input neuron, or an input neuron, can be created during the adaptation phase of an EFuNN (see Fig. 5.10a, 5.10b and the explanation in section 3). The task of the fuzzy input nodes is to transfer the input values into membership degrees to which they belong to the MF.

The third layer contains rule (case) nodes that evolve through supervised/unsupervised learning. The rule nodes represent prototypes (exemplars, clusters) of input-output data associations, graphically represented as an association of hyper-spheres from the fuzzy input and fuzzy output spaces. Each rule node r is defined by two vectors of connection weights – W1(r) and W2(r), the latter being adjusted through supervised learning based on the output error, and the former being adjusted through unsupervised learning based on similarity measure within a local area of the problem space. The fourth layer of neurons represents fuzzy quantization for the output variables, similar to the input fuzzy neurons representation. The fifth layer represents the real values for the output variables.

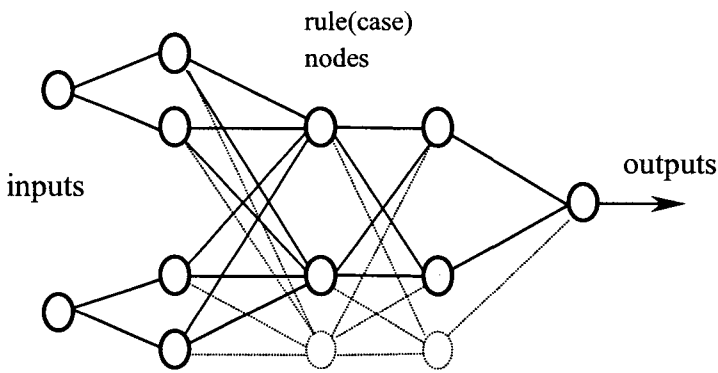


Fig. 5.2a The Five-layers Basic Structure of the EfuNNs.

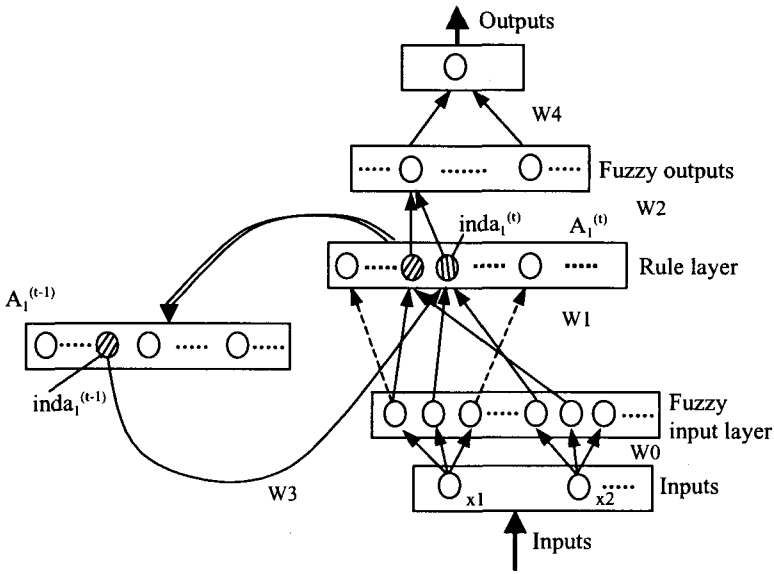
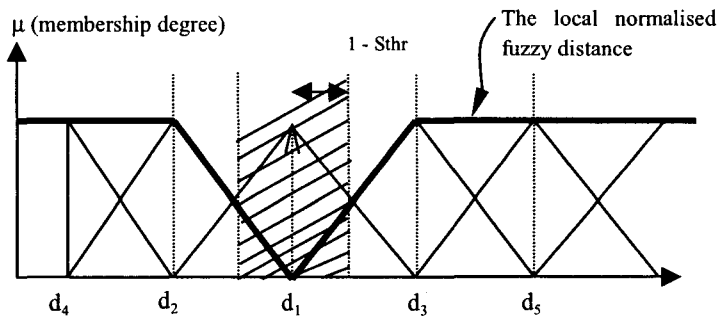


Fig. 5.2b EFuNNs with Recurrent Temporal Connections.



Dist(d1d2) = D(d1d3) = D(d1d5) =1

Fig. 5.3 Calculating Local Normalised Fuzzy Distance.

The evolving process can be based on two assumptions: (1) no rule nodes exist prior to learning and all of them are created (generated) during the evolving process; or (2) there is an initial set of rule nodes that are not connected to the input and output nodes and become connected through the learning (evolving) process. The latter case is more biologically plausible [82]. The EFuNN evolving algorithm presented in the next section does not make a difference between these two cases.

Each rule node, e.g. r_j, represents an association between a hyper-sphere from the fuzzy input space and a hyper-sphere from the fuzzy output space (see Fig. 5.4a), the $W1(r_j)$ connection weights representing the co-ordinates of the center of the sphere in the fuzzy input space, and the $W2$ (r_j) – the co-ordinates in the fuzzy output space. The radius of an input hyper-sphere of a rule node is defined as (1- Sthr), where Sthr is the sensitivity threshold parameter defining the minimum activation of a rule node (e.g., r1, previously evolved to represent a data point (Xd1,Yd1)) to an input vector (e.g., (Xd2,Yd2)) in order for the new input vector to be associated with this rule node. Two pairs of fuzzy input-output data vectors **d1**=(Xd1,Yd1) and **d2**=(Xd2,Yd2) will be allocated to the first rule node r_1 if they fall into the r_1 input sphere and in the r_1 output sphere, i.e. the local normalised fuzzy difference between Xd1 and Xd2 is smaller than the radius *r* and the local normalised fuzzy difference between Yd1 and Yd2 is smaller than an error threshold Errthr. The local normalised fuzzy difference between two fuzzy membership vectors **d1f** and **d2f** that represent the membership degrees to which two real values d1 and d2 data belong to the pre-defined MF, are calculated as D(**d1f**,**d2f**) = sum(abs(**d1f** - **d2f**))/sum(**d1f** + **d2f**)). For example, if **d1f**=(0,0,1,0,0,0) and **d2f**=(0,1,0,0,0,0) (see Fig. 5.3a), than D(d1,d2) = (1+1)/2=1 which is the maximum value for the local normalised fuzzy difference (see Fig. 5.3a, 5.3b).

If data example **d1** = (Xd1,Yd1), where Xd1 and Xd2 are correspondingly the input and the output fuzzy membership degree vectors, and the data example is associated with a rule node r_1 with a centre r_1^1, than a new data point **d2**=(Xd2,Yd2), that is within the shaded area as shown in Fig. 5.3a and Fig. 5.4a, will be associated with this rule node too. Through the process of associating (learning) of new data points to a rule node, the centres of this node hyper-spheres adjust in the fuzzy input space depending on a learning rate lrn1, and in the fuzzy output space depending on a learning rate lr2, as it is shown in Fig. 5.4a on the two data points **d1** and **d2**. The adjustment of the centre r_1^1 to its new position r_1^2 can be represented mathematically by the change in the connection weights of the rule node r_1 from $W1(r_1^1)$ and $W2(r_1^1)$ to $W1(r_1^2)$ and

$W2(r_1^2)$ according to the following vector operations:

$$W2(r_1^2) = W2(r_1^1) + lr2.\ Err(Yd1,Yd2).\ A1(r_1^1),$$
$$W1(r_1^2) = W1(r_1^1) + lr1.\ Ds(Xd1,Xd2),$$

where: Err(Yd1,Yd2)= Ds(Yd1,Yd2)=Yd1-Yd2 is the signed value rather than the absolute value of the fuzzy difference vector; $A1(r_1^1)$ is the activation of the rule node r_1^1 for the input vector Xd2. The learning process in the fuzzy input space is illustrated in Fig. 5.4b on four data points d1,d2,d3 and d4. Fig. 5.4c shows how the centre of the rule node r_1 adjusts after learning each new data point when two-pass learning is applied. If lrn1=lrn2=0, once established, the centres of the rules nodes do not move.

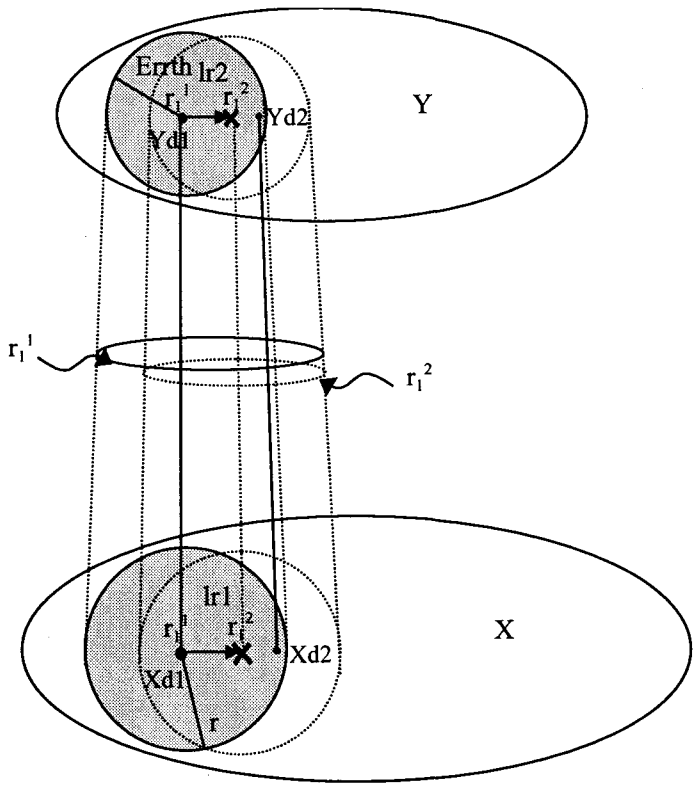


Fig. 5.4a Input / Output Mapping and Association.

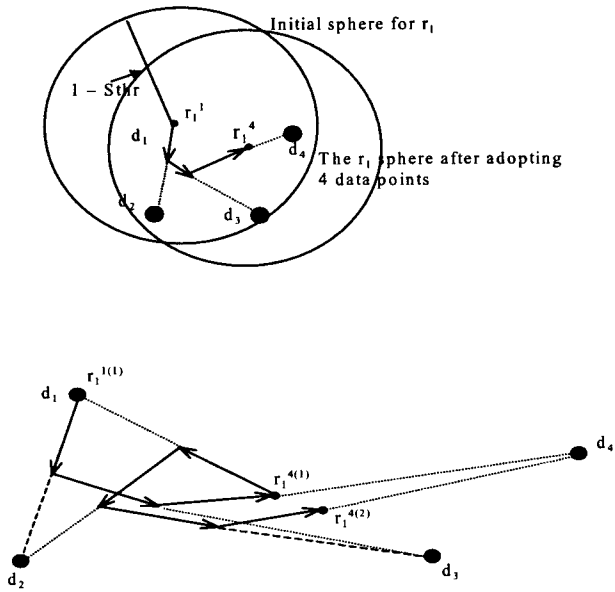


Fig. 5.4b, and -c The Process of Moving a Lyper-sphere Center of a Rule Node When Accommodation More Similar Data Vectors in : (a) one-pass learning; (b) in two-passes learning.

While the connection weights from W1 and W2 capture spatial characteristics of the learned data (centres of hyper-spheres), the temporal layer of connection weights W3 from Fig. 5.2b captures temporal dependencies between consecutive data examples. If the winning rule node at the moment (t-1) (to which the input data vector at the moment (t-1) was associated) was r1=inda1(t-1), and the winning node at the moment t is r2=inda1(t), then a link between the two nodes is established as follows:

$$W3(r1,r2)^{(t)} = W3(r1,r2)^{(t-1)} + lr3.\ A1(r1)^{(t-1)}\ A1(r2))^{(t)},$$

where: $A1(r)^{(t)}$ denotes the activation of a rule node r at a time moment (t); lr3 defines the degree to which the EFuNN associates links between rules (clusters, prototypes) that include consecutive data examples (if lr3=0, no temporal asso-

ciations are learned in an EFuNN structure and the EFuNN from Fig. 5.2b becomes the one from Fig. 5.2a).

The learned temporal associations can be used to support the activation of rule nodes based on temporal, pattern similarity. Here, temporal dependencies are learned through establishing structural links. These dependencies can be further investigated and enhanced through synaptic analysis (at the synaptic memory level) rather than through neuronal activation analysis (at the behavioural level). The ratio spatial-similarity/temporal-correlation can be balanced for different applications through two parameters Ss and Tc such that the activation of a rule node r for a new data example d_{new} is defined as the following vector operations:

$$A1\ (r) = f(\ Ss.\ D(r, d_{new}) + Tc.W3(r^{(t-1)}, r))$$

where: f is the activation function of the rule node r, $D(r, d_{new})$ is the normalised fuzzy distance value and $r^{(t-1)}$ is the winning neuron at the previous time moment.

Fig. 5.5a, 5.5b show a schematic diagram of the process of evolving of four rule nodes and setting the temporal links between them for data taken from consecutive frames of hypothetical speech (phoneme) data.

Several parameters were introduced so far for the purpose of controlling the functioning of an EFuNN. Some more parameters will be introduced later, that will bring the EFuNN parameters to a comparatively large number. In order to achieve a better control of the functioning of an EFuNN structure, the three-level functional hierarchy is used here as defined in section 2 for the ECOS architecture, namely: genetic level, long-term synaptic level, and short- term activation level.

At the genetic level, all the EFuNN parameters are defined as genes in a chromosome. These are:

(a) structural parameters, e.g.: number of inputs, number of MF for each of the inputs, initial type of rule nodes, maximum number of rule nodes, number of MF for the output variables, number of outputs.

(b) functional parameters, e.g.: activation functions of the rule nodes and the fuzzy output nodes (in the experiments below saturated linear functions are used); mode of rule node activation ("one-of-n", or "many-of-n", depending on how many activation values of rule nodes are propagated to the next level); learning rates lr1,lr2 and lr3; sensitivity threshold Sthr for the rule layer; error threshold Errthr for the output layer; forgetting rate;

various pruning strategies and parameters, as explained in the EFuNN algorithm below.

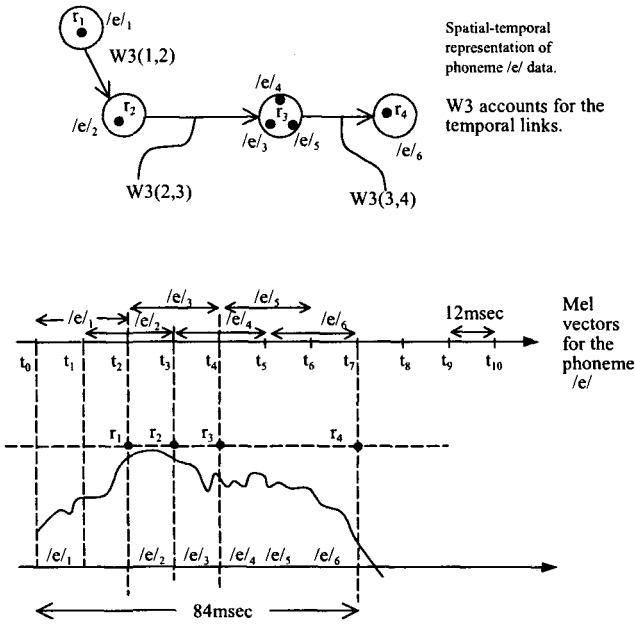


Fig. 5.5a, and -b Evolving Four Rule Nodes from a Hypothetical Speech Data (e.g., phoneme /e/).

5.3.2 *The EFuNN Learning Algorithm*

The EFuNN algorithm, to evolve EFuNNs from incoming examples, is based on the principles explained in the previous section. It is given below as a procedure of consecutive steps. Matrix operation expressions are used similar to the expressions in a matrix processing language such as MATLAB.

1. Initialise an EFuNN structure with a maximum number of neurons and no (or zero-value) connections. Initial connections may be set through inserting fuzzy rules in the structure [44]. If initially there are no rule (case) nodes connected to the fuzzy input and fuzzy output neurons, then create the first node rn=1 to represent the first example d1 and set its input W1(rn) and output W2(rn) connection weight vectors as follows:

<Create a new rule node rn>: W1(rn)=EX; W2(rn) = TE, where TE is the fuzzy output vector for the current fuzzy input vector EX.

2. WHILE <there are examples in the input stream> DO

Enter the current example (Xdi,Ydi), EX denoting its fuzzy input vector. If new variables appear in this example, which are absent in the previous examples, create new input and/or output nodes with their corresponding membership functions.

3. Find the *local normalised fuzzy distance* between the fuzzy input vector EX and the already stored patterns (prototypes, exemplars) in the rule (case) nodes rj=r1,r2,...,rn

D(EX, rj)= sum (abs (EX - W1(j))) / sum (W1(j)+EX)

4. Find the activation A1 (rj) of the rule (case) nodes rj, rj=r1:rn. Here radial basis activation function, or a saturated linear one, can be used, i.e. A1 (rj) = radbas (D(EX, rj)), or A1(rj) = satlin (1 – D(EX, rj)). The former may be appropriate for function approximation tasks, while the latter may be preferred for classification tasks. In case of the feedback variant of an EFuNN, the activation A1(rj) is calculated as:

A1 (rj) = radbas (Ss. D(EX, rj) - Tc.W3), or A1(j) = satlin (1 – Ss. D(EX, rj) + Tc.W3) .

5. Update the pruning parameter values for the rule nodes, e.g. *age*, *average activation*, as pre-defined in the EFuNN chromosome.

6. Find all case nodes rj with an activation value A1(rj) above a sensitivity threshold Sthr.

7. If there is no such case node, then <Create a new rule node> using the procedure from step 1 in an unsupervised learning mode

ELSE

8. Find the rule node inda1 that has the maximum activation value (e.g., maxa1).

9. (a) in case of "one-of-n" EFuNNs (as it is in [9, 27, 47]) propagate the activation maxa1 of the rule node inda1 to the fuzzy output neurons:

A2 = satlin (A1(inda1) . W2(inda1)

(b) in case of "many-of-n" mode, the activation values of all rule nodes that are above an activation threshold of Athr are propagated to the next neuronal layer (this case is not discussed in details here; it has been further developed into a new EFuNN architecture called dynamic, 'many-of-n' EFuNN, or DEFuNN [42]).

10. Find the winning fuzzy output neuron *inda2* and its activation *maxa2*.

11. Find the desired winning fuzzy output neuron *indt2* and its value *maxt2*.

12. Calculate the fuzzy output error vector: Err=A2 - TE.

13. IF (*inda2* is different from *indt2*) or (D(A2,TE) > Errthr) <Create a new rule node>

ELSE

14. Update: (a) the input, (b) the output, an (c) the temporal connection vectors (if such exist) of the rule node *k=inda1* as follows:

(a) Ds(EX,W1(k)) =EX-W1(k); W1(k)=W1(k) + lr1.Ds(EX,W1(k)), where *lr1* is the learning rate for the first layer;

(b) W2(k) = W2 (k) + lr2. Err. maxa1, where *lr2* is the learning rate for the second layer;

(c) W3(l,k)=W3(l,k)+lr3. A1(k).A1(l) $^{(t-1)}$, here l is the winning rule neron at the previous time moment (t-1), and A1(l) $^{(t-1)}$ is its activation value kept in the short term memory.

15. Prune rule nodes j and their connections that satisfy the following fuzzy pruning rule to a pre-defined level:

IF (a rule node rj is OLD) AND (average activation A1av(rj) is LOW) and (the density of the neighbouring area of neurons is HIGH or MODERATE (i.e. there are other prototypical nodes that overlap with j in the input-output space; this condition apply only for some strategies of inseting rule nodes as explained in a sub-section below)

THEN the probability of pruning node (rj) is HIGH

The above pruning rule is fuzzy and it requires that the fuzzy concepts of OLD, HIGH, etc., are defined in advance (as part of the EFuNN's chromosome). As a partial case, a fixed value can be used, e.g. a node is OLD if it has existed during the evolving of a FuNN from more than 1000 examples. The use of a pruning strategy and the way the values for the pruning parameters are defined, depends on the application task.

16. Aggregate rule nodes, if necessary, into a smaller number of nodes (see the explanation in the following subsection).

17. END of the while loop and the algorithm

18. Repeat steps 2-17 for a second presentation of the same input data or for an ECO training if needed.

5.3.3 *Strategies for Locating Rule Nodes in the Rule Node Space*

There are different ways to locate rule nodes in an EFuNN rule node space as it is explained here. The type selected depends on the type of the problem the EFuNN is designed to solve. Here some possible strategies are explained as illustrated in Fig. 5.6:

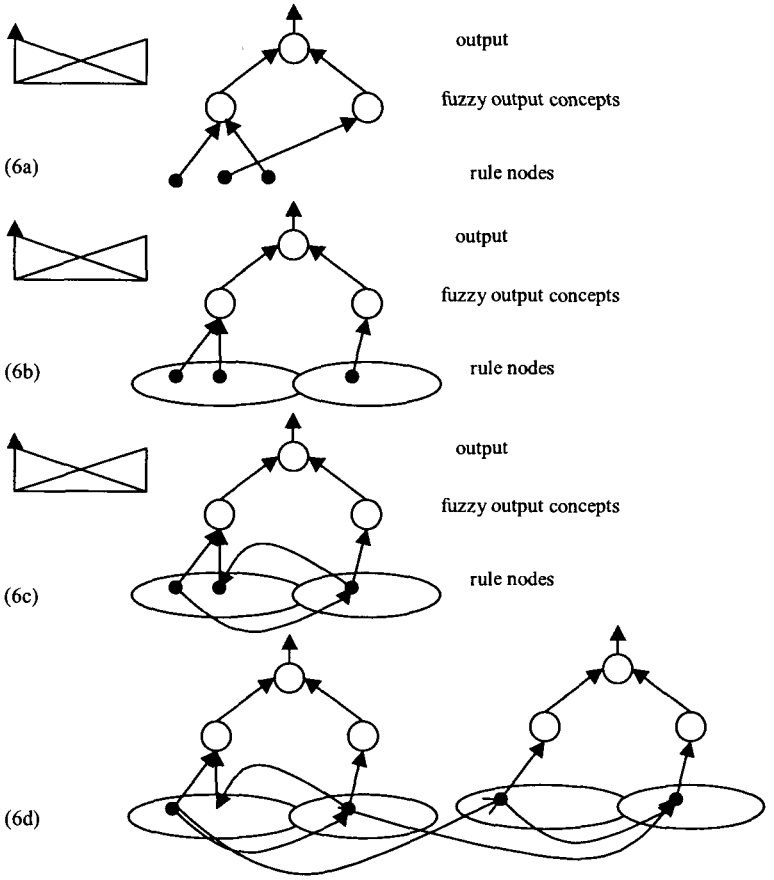


Fig. 5.6 Strategies for Allocating Rule Nodes in the Rule Node Space.

(a) Simple consecutive allocation strategy, i.e. each newly created rule (case) node is allocated next to the previous and the following ones in a linear fashion. That represents a time order. The following statement is valid if no pruning technique is applied, but aggregation technique instead, to optimise the size of the rule layer: at least one example that was associated with rule node r_j was presented to the EFuNN before at least one example that was associated to the rule node (r_j+1) (see Fig. 5.6a).

(b) Pre-clustered location, i.e. for each output fuzzy node (e.g. NO, YES) there is a predefined location where the rule nodes supporting this prede-

fined concept are located. At the center of this area the nodes that fully support this concept (error 0) are placed; every new rule node's location is defined based on the fuzzy output error and the similarity with other nodes (Fig. 5.6b);

(c) Nearest activated node insertion strategy, i.e. a new rule node is placed nearest to the highly activated node which activation is still less than the Sthr. A connection between the neighbouring nodes can be established similar to the temporary connections from W3.

(d) As in (c) but temporal feedback connections are set as well (see Fig. 5.2b and Fig. 5.6c). New connections are set that link consecutively activated rule nodes through using the short term memory and the links established through the W3 weight matrix; that will allow for the evolving system to repeat a sequence of data points starting from a certain point and not necessarily from the beginning.

(e) The same as above, but in addition, new connections are established between rule nodes from different EFuNN modules that become activated simultaneously (at the same time moment) (Fig. 5.6d). This would make it possible for an ECOS to learn a correlation between conceptually different variables, e.g. correlation between speech sound and lip movement.

5.3.4 *An Example of Using the EFuNN Algorithm in an EFuNN Simulator*

Here, a small speech data set of 400 phoneme data examples is used to illustrate the EFuNN learning algorithm. 100 examples of each of the four phonemes /I/ (from 'sit'), /e/ (from 'get'), /ae/ (from 'cat'), and /i/ (from 'see'), which are phonemes 25,26,27 and 31 from the Otago Speech Corpus available from the WWW http://kel.otago.ac.nz/, are extracted from the speech data of two speakers of NZ English (one male and one female, numbers 17 and 21 from the Corpus). Each data example used in the experiment described below consists of 3 time lags of 26-element mel-scale vectors, each representing the speech signal within a time frame of 11.6msec, and an output label giving the phoneme class. The speech data is segmented and processed with the use of a 256-point FFT, Hamming window, overlapping of 50% between the consecutive time frames, each of them being 11.6msec long (see Fig. 5.5b).

An EFuNN with 78 inputs and 4 outputs was evolved on the 400 data examples and tested on another set. Fig. 5.7 shows the growth of the number of the rule nodes with the progress of entering data examples for one pass of training and the root mean square error RMSE. The parameter values for the EFuNN

parameters (e.g. number of evolved rule nodes rn, learning rates lr1,lr2 and lr3, pruning parameters) are shown on the display of the EFuNN simulator which is available from: http://divcom.otago.ac.nz/infosci/kel/CBIIS/RICBIS/).

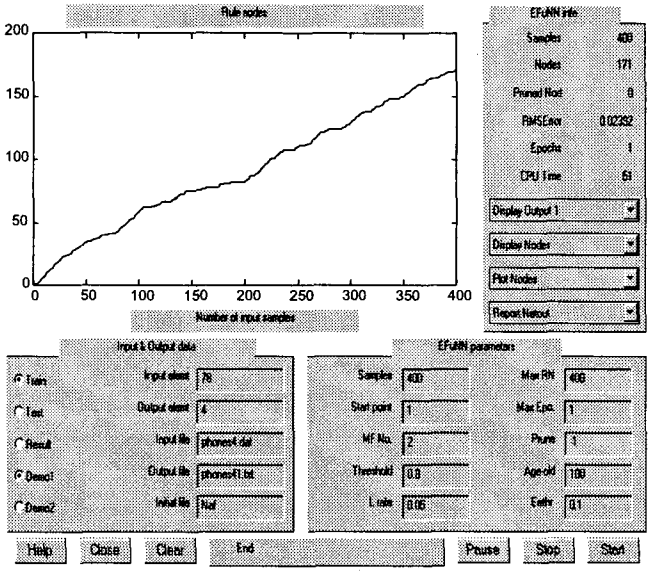


Fig. 5.7 A Simulation Process of Evolving an EFuNN for Four Vowels /I/,/e/, /ae/ and /i/ from 400 Data Examples Taken from the Otago Speech Corpus Data.

5.3.5 *Learning Modes in EFuNN. Rule Insertion, Rule Extraction and Aggregation*

Different learning, adaptation and optimisation strategies and algorithms can be applied on an EFuNN structure for the purpose of its evolving. These include:

- *Active learning* , e.g. the EFuNN algorithm;
- *Passive learning* (i.e., *cascade-eco,* and *sleep-eco* learning) as explained in section 2;
- *Rule insertion* into EFuNNs [44]. EFuNNs are adaptive rule-based systems. Manipulating rules is essential for their operation. This includes rule insertion, rule extraction, and rule adaptation. At any time (phase) of the evolving (learning) process fuzzy or exact rules can be inserted and extracted. Insertion of fuzzy rules is achieved through setting a new rule no-

de rj for each new rule R, such that the connection weights W1(rj) and W2 (rj) of the rule node represent the rule R. For example, the fuzzy rule *(IF x1 is Small and x2 is Small THEN y is Small)* can be inserted into an EFuNN structure by setting the connections of a new rule node to the fuzzy condition nodes x1- Small and x2- Small and to the fuzzy output node y-Small to a value of 1 each. The rest of the connections are set to a value of zero. Similarly, an exact rule can be inserted into an EFuNN structure, e.g. *IF x1 is 3.4 and x2 is 6.7 THEN y is 9.5,* but here the membership degrees to which the input values x1=3.4 and x2=6.7, and the output value *y=9.5* belong to the corresponding fuzzy values are calculated and attached to the corresponding connection weights.

- *Rule extraction and aggregation.* Each rule node *r*, which represents a prototype, rule, exemplar from the problem space, can be described by its connection weights W1(r) and W2 (r) that define the association of the two corresponding hyper-spheres from the fuzzy input and the fuzzy output problem spaces. The association is expressed as a fuzzy rule, for example:

  IF x1 is Small 0.85 and x1 is Medium 0.15 and x2 is Small 0.7 and x2 is Medium 0.3

  THEN y is Small 0.2 and y is Large 0.8

The numbers attached to the fuzzy labels denote the degree to which the centers of the input and the output hyper-spheres belong to the respective MF.

The process of rule extraction can be performed as aggregation of several rule nodes into a larger hyper-spheres as it is shown in Fig. 5.8a and Fig. 5.8b on an example of three rule nodes r1, r2 and r3 (only the input space is shown there). For the aggregation of two rule nodes r1 and r2, the following aggregation rule is used [44]:

> IF (D(W1(r1),W1(r2)) < = Thr1) AND (D(W2(r1),W2(r2)) <= Thr2)
> THEN aggregate r1 and r2 into r_{agg} and calculate the centres of the new rule node as: $W1(r_{agg})$ = average (W1(r1),W1(r2)), $W2(r_{agg})$ = average (W2(r1),W2(r2))

Here the geometrical center between two points in a fuzzy problem space is calculated with the use of an average vector operation over the two fuzzy vectors. This is based on a presumed piece-wise linear function between two points from the defined through the parameters Sthr and Errthr input and output fuzzy hyper-spheres.

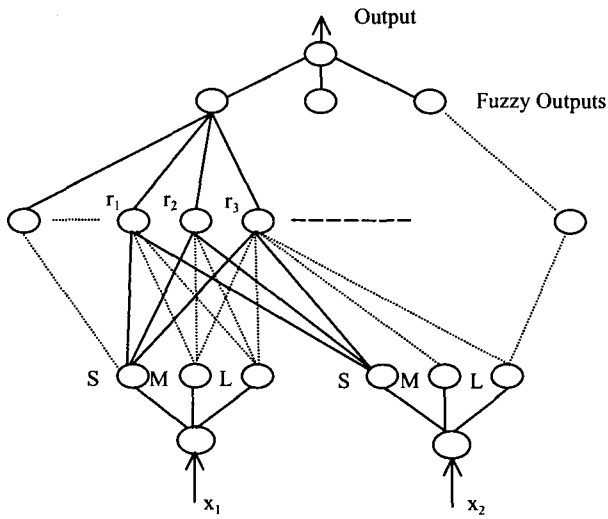


Fig. 5.8a An Evolving EFuNN Where Currently Three Rule Nodes Are Allocated in a Neighbourhood to Represent Three Close Exemplars.

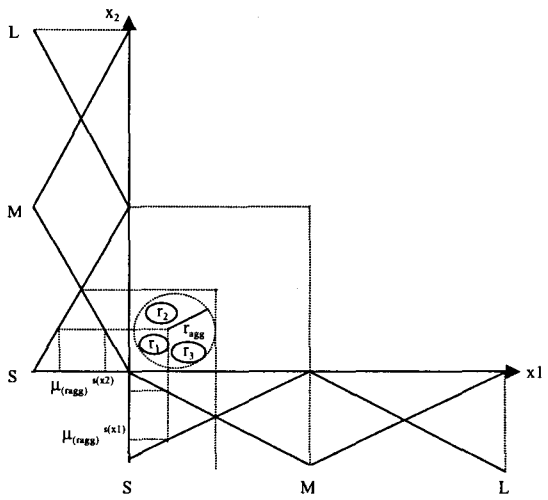


Fig. 5.8b The Process of Rule Node Aggregation Illustrated on Three Rule Nodes.

Example: The following two rules (rule nodes) r1 and r2 can be aggregated for Thr1=0.15 and Thr2=0.05 into a new rule r_{agg} as it is shown below:
r1:IF x1 is Small 0.85 and x1 is Medium 0.15 and x2 is Small 0.7 and x2 is Medium 0.3 THEN y is Small 0.1 and y is Medium 0.9
r2: IF x1 is Small 0.80 and x1 is Medium 0.2 and x2 is Small 0.8 and x2 is Medium 0.2
THEN y is Small 0.12 and y is Medium 0.88
D(W1(r1),W1(r2))=(0.05+0.05+0.1+0.1)/2/2=0.075 < Thr1=0.15;
D(W2(r1),W2(r2))=(0.02+0.02)/2/1=0.005 < 0.02 < Thr2=0.05;

r_{agg}: IF x1 is Small 0.825 and x1 is Medium 0.175 and x2 is Small 0.75 and x2 is Medium 0.25 THEN y is Small 0.11 and y is Medium 0.89

Through node creation and consecutive aggregation an EFuNN systems can adjust over time to changes in the data stream. Fig. 5.8c shows a hypothetical case of how a rule node rj, which represents a phoneme data cluster, would shift in the phoneme data space with new speakers of different accents talking to the system over time and the system adapts to them.

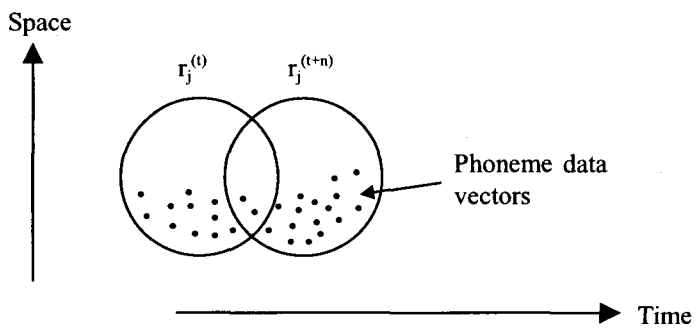


Fig. 5.8c The Process of Moving Rule Node Clusters through Aggregation.

- *Aggregation and abstraction through ECO-learning*: Aggregation of rule nodes to represent association of larger hyper-spheres from the input and the output space can be achieved through the use of the ECO learning method, when the connection weights $W1^{(1)}$ and $W2^{(1)}$ of an evolved

EFuNN1 are used as fuzzy exemplars to evolve an EFuNN2 for smaller values of the sensitivity threshold Sthr and the error threshold Errthr (see Fig. 5.9). This process can be continued further to evolve a new EFuNN3 with smaller number of rule nodes, therefore smaller number of rules, and so on. In case of function approximation tasks, the accuracy of the generalisation in this case may decrease depending on the chosen thresholds Thr1 and Thr2 as aggregation means creation of larger prototypes that accommodate more examples having similar input vectors and similar output vectors. For classification tasks where the output value is a symbolic (e.g., 'yes'/'no' class label) the aggregation may not affect the accuracy.

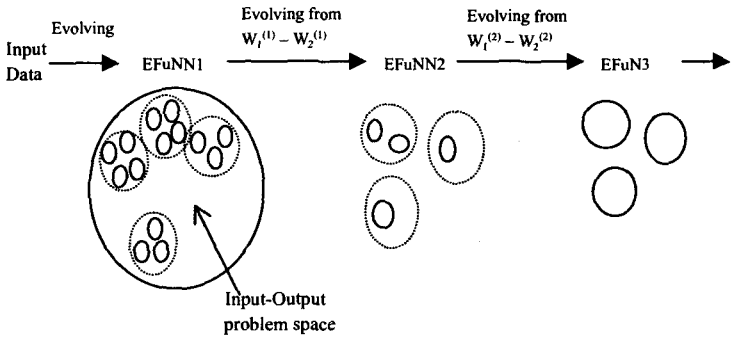


Fig. 5.9 The Process of Growing (evolving) and Shrinking (aggregation); also the Process of Sleep-ECO Learning.

- *Extracting rules for learning temporal pattern correlation:* Through analysis of the weights W3 of an evolved EFuNN, temporal correlation between time consecutive exemplars can be expressed in terms of rules and conditional probabilities, e.g.:

IF (W1(r1),W2(r1)) $^{(t-1)}$
THEN (W1(r1),W2(r2)) $^{(t)}$ (0.3)

The meaning of the above rule is that examples that belong to the rule (prototype) r1 follow in time examples from the rule prototype r2 with a relative conditional probability of 0.3.

- *Changing MF during operation.* This operation may be needed for a refined performance after certain time of the system operation. For example, instead of three MF, the system may perform better if it had five MF for some of the variables. In traditional fuzzy neural networks this change is either not allowed, or is extremely difficult to implement. In EFuNNs there are several possibilities to implement such dynamical changes of MF. These are: (a) The stored fuzzy exemplars in W1 and W2 that have three MF are defuzzifyied (e.g., through the center of gravity deffuzification technique) and than used to evolve a new EFuNN structure that has, for example, five MF; (b) New MF can be created (inserted) without a need for the old ones to be changed.The degree to which each cluster centre (each rule node) belongs to the new MF can be calculated through defuzzifying the centres as in case (a); (c) When aggregation of rule nodes is applied after many epochs, it is possible that input or output MF become fuzzy as the centers of the rule hyper-spheres move, so that there is no one-to-one defuzzification procedure from the connection weights back to the real input values.
- *On-line parameter optimisation.* Once set, the values for the EFuNN parameters will need to be optimised during the learning process. Optimisation can be done through analysis of the behaviour of the system and through a feedback connection from the higher level modules. Genetic algorithms (GA) can also be applied to optimise the EFuNNs structural and functional parameters based on either standard GA algorithms, or on their possible modifications for dynamic, on-line application. The latter case is concerned with an optimisation of parameters to adjust to a continuously incoming stream of data with changing dynamics and changing probability distribution. In this case a segment of the most recent data is stored regularly into an additional memory and a GA is applied on this data to optimise the EFuNN.

With the learning and pruning operations as part of the EFuNN learning algorithm, and with some additional adaptation techniques, an EFuNN can dynamically organise its structure to learn from data in an adaptive, continuous, incremental, life-long learning mode.

5.4 EFuNNs as Universal Learning Machines. Local and Global Generalisation

EFuNNs are designed to work in an on-line mode, with a continuous input data

stream. An EFuNN is trained (evolved) on input-output vectors of data available over time. Then it is used to generalise on new incoming data Xd for which the output is not known. Once the output vector Yd for the new input data becomes known, the input-output pair (Xd,Yd) is accommodated in the EFuNN structure, which is then used on the next input data, and so on. EFuNNs are memory-based systems, i.e. they store the incoming information as associated input-output clusters (fuzzy rules, prototypes) organised in hyperspherical forms. The clusters (their centres) are adjustable through the learning parameters lr1 and lr2, so they can 'move' in the problem space in order to accommodate new examples as such become available from the input stream. This continuous, learning process depends very much on the values set for the learning and pruning parameters. The optimal performance of EFuNNs in terms of learning error, generalisation, forgetting and convergence can be achieved through varying their structural and functional parameters. The generalisation ability of EFuNNs depends on the learning and pruning coefficients which can be dynamically adjusted in an ECOS architecture through a feedback connection from the higher level decision module or through optimisation techniques (see Fig. 5.1). It will be shown here that EFuNNs are universal learning machines that can learn, subject to a chosen degree of accuracy, any data set D, regardless of the class of problems (function approximation, time series prediction, classification, etc.).

In an on-line learning an EFuNN is evolved incrementally on different segments of data from the input stream (as a partial case this is just one data item). Off-line learning can also be applied on an EFuNN, when the system is evolved on part of the data and then tested on another part from the problem space, which completes the training and testing procedure as it is the case in many traditional NN models.

When issues such as universality of the EFuNN mechanism, learning accuracy, generalisation and convergence for different tasks are discussed, two cases must be distinguished:

(a) The incoming data is from a compact and bounded data space. In this case the more data vectors are used for evolving an EFuNN, the better its generalisation is on the whole problem space (or an extraction of it). After an EFuNN is evolved on some examples for the problem space, its *global generalisation error* can be evaluated on a set of p new examples from the problem space as follows:

$$GErr = \text{sum } \{Err_i\}_{i=1,2,...p},$$

where: Err_i is the error for a vector $\mathbf{x}_i$ from the input space X, which vector has not been and will not be used for training the EFuNN before the value GErr is calculated.

After having evolved an EFuNN on a small, but representative part of the whole problem space, its global generalisation error can become sufficiently small. This is valid for both off-line learning mode and on-line learning (when an EFuNN is evolved on k examples and then used to generalise on the next p examples).

For an on-line learning mode in which the EFuNN is adjusted incrementally on each example from the data stream the generalisation error on the next new input vector (for which the output vector is not known) is called *local generalisation error.* The local generalisation error at the moment t, for example, when the input vector is Xdt, and the calculated by the evolved EFuNN output vector is Ydt', is expressed as Err_t. The cumulative local generalisation error can be estimated as:

$$TErr_t = \text{sum } \{Err_t\}_{,t=1,2,...i}.$$

In contrast to the global generalisation error, here the error Err_t is calculated after the EFuNN has learned the previous example (Xd(t-1), Yd(t-1)). Each example is propagated only once through the EFuNN, both for testing the error and learning (after the output vector becomes known). The root mean square error can be calculated for each data point i from the input data stream as:

$$RMSE(i) = \text{sqrt } (\text{sum}\{Err_t\}_{t=1,2,..,i}) / i),$$

where: $Err_t = (d_t - o_t)^2$, d_t is the desired output value and o_t is the EFuNN output value produced for the t_{th} input vector.

(b) Open problem space, where the data dynamics and data probability distribution can change over time in a continuous way. Here, local generalisation error only can be evaluated.

For the two cases (a) and (b) above the following two theorems are valid.

Theorem 1. For any stream of input-output data from a compact and bounded problem space, there is an EFuNN system that can approximate the data to any desired degree of accuracy ξ after a certain time moment T defined by the distribution of the incoming data if the data represents a continuous function in the problem space.

Proof. The proof of the theorem, which is outlined here, is based on the following assumptions. After a time moment T, each of the fuzzy input and the fuzzy output spaces (they are compact and bounded) will be covered by the fuzzy hyper-spheres of the rule nodes generated over time, with a resolution accuracy of r=1-Sthr and Errthr respectively. After a sufficient number of examples from the stream presented by a certain time moment T, both the global generalisation error and the total local generalisation error will saturate to a value E proportional to the chosen value for the error threshold Errthr, therefore each of them will become less than the desired accuracy ξ. This is valid in case of the data stream approximating a continuous function, so that any two data points from a sufficiently small fuzzy input neighbourhood will have sufficiently small difference in the fuzzy output space. It can be precisely proved that any two associated compact and bounded fuzzy spaces X and Y can be fully covered by associated (possibly, overlapping) fuzzy hyper-spheres [38]. A similar theorem for multi-layer perceptrons with sigmoidal activation functions was proved in [10, 21]. But here, the on-line learning mode is covered too.

The EFuNNs can also be used to learn sequences from open spaces (case (b)), where the probability distribution and the dynamics of the data sequence can change over time. In this case the system will learn rules and prototypes and the generalisation accuracy will depend on the closeness of the new input data to already evolved prototypes both in space and time.

Theorem 2. For any continuous stream of input-output data from an open problem space, used to evolve an EFuNN, the local generalisation error at a time moment (t+1) will be less than a predefined value ξ if at the time moment t there is a rule node rj = (W1(rj), W2(rj)), such that D(W2(rj).(1-Dx), Ydt) < ξ, when Dx = D(W1(rj,Xdt))= min {D(W1(ri), Xdt))}, for i= 1,2,...,rn (rn is the number of the rule nodes evolved in the EFuNN structure until the time moment t).

The proof of this theorem uses the definition of local generalisation and the operations from the EFuNN learning algorithm.

5.5 Conclusions and Directions for Further Research

This paper presents some biological principles and functions of the brain and their implementation in a framework ECOS for evolving connectionist systems, and in evolving fuzzy neural networks EFuNN, in particular, for building on-line, knowledge-based, adaptive learning systems. ECOS have features that address the seven major requirements to the next generation of intelligent in-

formation systems as derived from several principles and functions of the human brain. A significant advantage of ECOS and EFuNNs is the local learning procedure which allows for a fast learning (possibly – one pass) after every new data item is entered and only few connections and nodes are changed. This is in contrast to the global learning algorithms where, for each input vector, all connection weights change thus making the system prone to catastrophic forgetting when applied for adaptive, on-line learning tasks.

In spite of the advantages of ECOS and EFuNNs when applied for on-line, adaptive learning, there are some difficulties that should be addressed in the future research. These include finding the optimal values for the evolving parameters, such as the sensitivity threshold Sthr, the error threshold Errthr, learning rates lr1, lr2 and lr3, forgetting rate, pruning, etc. For example, pruning of rule nodes has to be made specific for every application, thus depending on the definition of age and the other fuzzy variables in the pruning rule. One solution is to regularly apply genetic algorithms and evolutionary computation as optimisation procedures to the ECOS and EFuNN structures.

Evolving connectionist systems could be viewed as a new AI paradigm. They incorporate important AI features, such as: adaptive learning; non-monotonic reasoning; knowledge manipulation in the presence of imprecision and uncertainties; knowledge acquisition and explanation. ECOS are knowledge-based systems, logic systems, case-based reasoning systems and adaptive connectionist-based systems, all together. Through self-organisation and self-improvement during its learning process, they allow for simulations of emerging, evolving intelligence to be attempted.

At present more theoretical investigations on the limitations of ECOS and EFuNNs are needed and also more analysis on their biological plausibility.

Acknowledgements

This research is part of a research programme funded by the New Zealand Foundation for Research Science and Technology, contract UOO808.

References

[1] Albus, J.S., A new approach to manipulator control: The cerebellar model articulation controller (CMAC), Tarns. of the ASME: Journal of Dynamic Systems, Measurement, and Control, pp.220-227, Sept. (1975).

[2] Amari, S. and Kasabov, N. eds, Brain-like Computing and Intelligent Information Systems, Springer Verlag, (1998).

[3] Amari, S., Mathematical foundations of neuro-computing, Proc. of IEEE, 78, 9, Sept. (1990).

[4] Arbib, M. (ed) The Handbook of Brain Theory and Neural Networks,The MIT Press, (1995).

[5] Bollacker, K., S.Lawrence and L.Giles, CiteSeer: An autonomous Web agent for automatic retrieval and identification of interesting publications, 2nd International ACM conference on autonomous agents, ACM Press, pp.116-123, (1998).

[6] Bottu and Vapnik, Local learning computation, Neural Computation, 4, pp.888-900, (1992).

[7] Carpenter, G. and Grossberg, S., Pattern recognition by self-organizing neural networks, The MIT Press, Cambridge, Massachusetts, (1991).

[8] Carpenter, G. and Grossberg, S, ART3: Hierarchical search using chemical transmitters in self-organising pattern-recognition architectures, Neural Networks, 3, 2, pp.129-152, (1990).

[9] Carpenter, G., Grossberg, S, Markuzon, N., Reynolds, J.H., and Rosen, D.B., FuzzyARTMAP: A neural network architecture for incremental supervised learning of analog multi-dimensional maps, IEEE Transactions of Neural Networks, Vol.3, No.5, pp.698-713, (1991).

[10]Cybenko, G., Approximation by super-positions of sigmoidal function, Mathematics of Control, Signals and Systems, 2, pp.303-314, (1989).

[11]DeGaris, H., Circuits of Production Rule - GenNets - The genetic programming of nervous systems, in: Albrecht, R., Reeves, C. and Steele, N. (eds) Artificial Neural Networks and Genetic Algorithms, Springer Verlag, (1993).

[12]Duda and Hart, Pattern classification and scene analysis, New York: Willey, (1973).

[13]Edelman, G., Neuronal Darwinism: The theory of neuronal group selection, Basic Books, (1992).

[14]Elman, J., Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D., and Plunkett, K., Rethinking Innateness (A Connectionist Perspective of Development), The MIT Press, (1997).

[15]Fahlman, C., and Lebiere, C., The Cascade - Correlation Learning Architecture, in: Turetzky, D (ed) Advances in Neural Information Processing Systems, vol.2, Mor-

gan Kaufmann, pp.524-532, (1990).

[16]Farmer, J.D., and Sidorowitch, Predicting chaotic time series, Physical Review Letters, 59, 845, (1987).

[17]Freeman, J., Saad, D., On-line learning in radial basis function networks, Neural Computation vol. 9, No.7, (1997).

[18]French, Semi-destructive representations and catastrophic forgetting in connectionist networks, Connection Science, 1, pp.365-377, (1992).

[19]Fritzke, B. A growing neural gas network learns topologies, Advances in Neural Information Processing Systems, vol.7, (1995).

[20]Fukuda, T., Komata, Y., and Arakawa, T., Recurrent Neural Networks with Self-Adaptive GAs for Biped Locomotion Robot, In: Proceedings of the International Conference on Neural Networks ICNN'97, IEEE Press, (1997).

[21]Funuhashi, K., On the approximate realization of continuous mappings by neural networks, Neural Networks, 2, pp.183-192, (1989).

[22]Gaussier, T., and Zrehen, S., A topological neural map for on-line learning: Emergence of obstacle avoidance in a mobile robot, In: From Animals to Animats No.3, pp.282-290, (1994).

[23]Goldberg, D.E., Genetic Algorithms in Search, Optimisation and Machine Learning, Addison-Wesley, (1989).

[24]Goodman, R., Higgins, C.M., Miller, J.W., and Smyth, P., Rule-based neural networks for classification and probability estimation, Neural Computation, 14, pp.781-804, (1992).

[25]Hashiyama, T., Furuhashi, T., Uchikawa, Y., A Decision Making Model Using a Fuzzy Neural Network, in: Proceedings of the 2nd International Conference on Fuzzy Logic & Neural Networks, Iizuka, Japan, pp.1057-1060, (1992).

[26]Hassibi and Stork, Second order derivatives for network pruning: Optimal Brain Surgeon, in: Advances in Neural Information Processing Systems, 4, pp.164-171, (1992).

[27]Hech-Nielsen, R., Counter-propagation networks, IEEE First int. conference on neural networks, San Diego, vol.2, pp.19-31, (1987).

[28]Heskes, T.M., Kappen, B., On-line learning processes in artificial neural networks, in: Math. foundations of neural networks, Elsevier, Amsterdam, pp.199-233, (1993).

[29]Ishikawa, M., Structural Learning with Forgetting, Neural Networks 9, pp.501-521, (1996).

[30]Kasabov, N., Adaptable connectionist production systems. Neurocomputing, 13 (2-4) pp.95-117, (1996).

[31]Kasabov, N., The ECOS Framework and the ECO Learning Method for Evolving Connectionist Systems, Journal of Advanced Computational Intelligence, 2, 6, pp.1-

8, (1998).

[32]Kasabov, N., Investigating the adaptation and forgetting in fuzzy neural networks by using the method of training and zeroing, Proceedings of the International Conference on Neural Networks ICNN'96, Plenary, Panel and Special Sessions volume, pp.118-123, (1996).

[33]Kasabov, N., Learning fuzzy rules and approximate reasoning in fuzzy neural networks and hybrid systems, Fuzzy Sets and Systems 82, 2, pp.2-20, (1996).

[34]Kasabov, N., A framework for intelligent conscious machines utilising fuzzy neural networks and spatial temporal maps and a case study of multilingual speech recognition, in: Amari, S. and Kasabov, N. (eds) Brain-like computing and intelligent information systems, Springer Verlag, pp.106-126, (1998)

[35]Kasabov, N., ECOS: A framework for evolving connectionist systems and the ECO learning paradigm, Proc. of ICONIP'98, Kitakyushu, Japan, IOS Press, pp.1222-1235, Oct. (1998).

[36]Kasabov, N., Evolving Fuzzy Neural Networks - Algorithms, Applications and Biological Motivation, in: in: Yamakawa and Matsumoto (eds), Methodologies for the Conception, design and Application of Soft Computing, World Scientific, pp.271-274, (1998).

[37]Kasabov, N., E. Postma, and J. Van den Herik, AVIS: A Connectionist-based Framework for Integrated Audio and Visual Information Processing, in Proc. of Iizuka'98, Iizuka, Japan, Oct. (1998).

[38]Kasabov, N., Foundations of Neural Networks, Fuzzy Systems and Knowledge Engineering, The MIT Press, CA, MA, (1996).

[39]Kasabov, N., Kim, J.S, Watts, M., A. Gray, FuNN/2- A Fuzzy Neural Network Architecture for Adaptive Learning and Knowledge Acquisition, Information Sciences - Applications, 101, 3-4, pp.155-175, (1997).

[40]Kasabov, N., M. Watts, Genetic algorithms for structural optimisation, dynamic adaptation and automated design of fuzzy neural networks, in: Proceedings of the International Conference on Neural Networks ICNN'97, IEEE Press, Houston, (1997).

[41]Kasabov, N., Kozma, R., Kilgour, R., Laws, M., Taylor, J., Watts, M., and Gray, A., A Methodology for Speech Data Analysis and a Framework for Adaptive Speech Recognition Using Fuzzy Neural Networks and Self Organising Maps, in: Kasabov and Kozma (eds) Neuro-fuzzy techniques for intelligent information systems, Physica Verlag (Springer Verlag), (1999).

[42]Kasabov, N., Song, Q. Dynamic, evolving fuzzy neural networks with 'm-out-of- n' activation nodes for on-line adaptive systems , TR 99/04, Department of Information Science, University of Otago, (1999).

[43]Kasabov, N., Watts, M. Spatial-temporal evolving fuzzy neural networks STE-FuNNs and applications for adaptive phoneme recognition, TR 99/03 Department of Information Science, University of Otago, (1999).

[44]Kasabov, N., Woodford, B. Rule Insertion and Rule Extraction from Evolving Fuzzy Neural Networks: Algorithms and Applications for Building Adaptive, Intelligent Expert Systems, in Proc. of Int. Conf. FUZZ-IEEE, Seoul, August (1999).

[45]Kasabov, N., Learning fuzzy rules and approximate reasoning in fuzzy neural networks and hybrid systems, Fuzzy Sets and Systems 82, 2, pp.2-20, (1996).

[46]Kawahara, S., Saito, T., On a novel adaptive self-organising network, Cellular Neural Networks and Their Applications, pp.41-46, (1996).

[47]Kohonen, T., The Self-Organizing Map, Proceedings of the IEEE, vol.78, N-9, pp.1464-1497, (1990).

[48]Kohonen, T., Self-Organizing Maps, second edition, Springer Verlag, (1997).

[49]Krogh, A., and Hertz, J.A., A simple weight decay can improve generalisation, Advances in Neural Information Processing Systems, 4, pp.951-957, (1992).

[50]Le Cun, Y., Denker, J.S., and Solla, S.A., Optimal Brain Damage, in: Touretzky, D.S., ed., Advances in Neural Information Processing Systems, Morgan Kaufmann, 2, pp.598-605, (1990).

[51]Lin, C.T. and Lee, C.S.G., Neuro Fuzzy Systems, Prentice Hall, (1996).

[52]Maeda, M., Miyajima, H. and Murashima, S., A self organizing neural network with creatinga nd deleting methods, Nonlinear theory and its applications, 1, pp.397-400, (1996)

[53]Mandziuk, J., Shastri, L., Incremental class learning approach and its application to hand-written digit recognition, Proc. of the fifth int. conf. on neuro-information processing, Kitakyushu, Japan, Oct. pp.21-23, (1998).

[54]Massaro, D., and Cohen, M., Integration of visual and auditory information in speech perception, Journal of Experimental Psychology: Human Perception and Performance, Vol 9, pp.753-771, (1983).

[55]McClelland, J., McNaughton, B.L., and Reilly, R.C., Why there are Complementary Learning Systems in the Hippocampus and Neo-cortex: Insights from the Successes and Failures of Connectionist Models of Learning and Memory, CMU TR PDP.CNS.94.1, March, (1994).

[56]Miller, D.J., Zurada and Lilly, J.H., Pruning via Dynamic Adaptation of the Forgetting Rate in Structural Learning, Proc. IEEE ICNN'96, Vol.1, p.448, (1996).

[57]Mitchell, M.T., Machine Learning, MacGraw-Hill, (1997)

[58]Moody, J., Darken, C., Fast learning in networks of locally-tuned processing units, Neural Computation, 1, pp.281-294, (1989).

[59]Mozer, M., and Smolensky, P., A technique for trimming the fat from a network via

relevance assessment, in: D. Touretzky (ed) Advances in Neural Information Processing Systems, vol.2, Morgan Kaufmann, pp.598-605, (1989).

[60]Murphy, P. and Aha, D., UCI Repository of machine learning databases, Irvin, CA: University of California, Department of Information and Computer Science, (1994), (http://www.ics.uci.edu/~mlearn/MLRepository.html).

[61]Port, R., and T.van Gelder (eds) Mind as motion (Explorations in the Dynamics of Cognition), The MIT Press, (1995).

[62]Quartz, S.R., and Sejnowski, T.J., The neural basis of cognitive development: a constructivist manifesto, Behavioral and Brain Science, (to appear).

[63]Jang, R., ANFIS: adaptive network-based fuzzy inference system, IEEE Trans. on Syst., Man, Cybernetics, 23, 3, May-June, pp.665-685, (1993).

[64]Reed, R., Pruning algorithms - a survey, IEEE Trans. Neural Networks, 4, 5, pp.740-747, (1993).

[65]Robins, A. and Frean, M., Local learning algorithms for sequential learning tasks in neural networks, Journal of Advanced Computational Intelligence, vol.2, 6, (1998).

[66]Robins, A., Consolidation in neural networks and the sleeping brain, Connection Science, 8, 2, pp.259-275, (1996).

[67]Rummery, G.A., and Niranjan, M., On-line Q-learning using connectionist systems, Cambridge University Engineering Department, CUED/F-INENG/TR, 166, (1994).

[68]Joseph, S.R.H., Theories of adaptive neural growth, PhD Thesis, University of Edinburgh, (1998).

[69]Saad, D. (ed) On-line learning in neural networks, Cambridge University Press, (1999).

[70]Sankar, A., and Mammone, R.J., Growing and Pruning Neural Tree Networks, IEEE Trans. Comput. 42, 3, pp.291-299, (1993).

[71]Schiffman, W., Joost, M., and Werner, R., Application of Genetic Algorithms to the Construction of Topologies for Multilayer Perceptrons In: Albrecht, R.F., Reeves,

[72]Segalowitz, S.J., Language functions and brain organization, Academic Press, (1983).

[73]Segev, R. and Ben-Jacob, E., From neurons to brain: Adaptive self-wiring of neurons, TR /98 Faculty of Exact Sciences, Tel-Aviv University, (1998).

[74]Selverston, A. (ed) Model neural networks and behaviour, Plenum Press, (1985).

[75]Sinclair, S., and Watson, C., The Development of the Otago Speech Database, In Kasabov, N. and Coghill, G. (Eds.), Proceedings of ANNES '95, Los Alamitos, CA, IEEE Computer Society Press, (1995).

[76]Towel, G., Shavlik, J., and Noordewier, M., Refinement of approximate domain theories by knowledge-based neural networks, Proc. of the 8th National Conf. on Artificial Intelligence AAAI'90, Morgan Kaufmann, pp.861-866, (1990).

[77]Ooyen, V., and Pelt, J.V., Activity-dependent outgrowth of neurons and overshoot phenomena in developing neural networks, Journal Theoretical Biology, 167, pp.27-43, (1994).

[78]Waibel, A., Vo, M., Duchnovski, P., Manke, S., Multimodal Interfaces, Artificial Intelligence Review, (1997).

[79]Watts, M., and Kasabov, N., Genetic algorithms for the design of fuzzy neural networks, in Proc. of ICONIP'98, Kitakyushu, Oct. (1998).

[80]Whitley, D., and Bogart, C., The evolution of connectivity: Pruning neural networks using genetic algorithms. Proc. Int. Joint Conf. Neural Networks, No.1, pp.17-22, (1990).

[81]Woldrige, M., and Jennings, N., Intelligent agents: Theory and practice, The Knowledge Engineering review (10), (1995).

[82]Wong, R.O., Use, disuse, and growth of the brain, Proc. Nat. Acad. Sci. USA, 92, 6, pp.1797-99, (1995).

[83]Yamakawa, T., Kusanagi, H., Uchino, E. and Miki, T., A new Effective Algorithm for Neo Fuzzy Neuron Model, in: Proceedings of Fifth IFSA World Congress, pp.1017-1020, (1993).

This page is intentionally left blank

Chapter 6

Interrelationships, Communication, Semiotics, and Artificial Consciousness

Horia-Nicolai L. Teodorescu
Technical University of Iasi

Abstract

The aim of this chapter is to refine some questions regarding AI, and to provide partial answers to them. We analyze the state of the art in designing intelligent systems that are able to mimic human complex activities, including acts based on artificial consciousness. The analysis is performed to contrast the human cognition and behavior to the similar processes in AI systems. The analysis includes elements of psychology, sociology, and communication science related to humans and lower level beings. The second part of this chapter is devoted to human-human and man-machine communication, as related to intelligence. We emphasize that the relational aspects constitute the basis for the perception, knowledge, semiotic and communication processes. Several consequences are derived. Subsequently, we deal with the tools needed to endow the machines with intelligence. We discuss the roles of knowledge and data structures. The results could help building "sensitive and intelligent" machines.

Keywords : artificial intelligence, psychology, semantic, computer semiotic, syntax, behavior, sensation basis, sensitivity, knowledge, personality, emotion, relationship, self-representation, communication, subliminal communication, interactivity, speech, natural language, perceptive computer, parallel language, annex (side) language, subliminal language, recognition, group relationship, connotation, morality, consciousness, representation systems, environment, wholarchic groups, representation basis, adaptability, tools, limits

6.1 Introduction

6.1.1 *Aim*

The three main questions addressed in this chapter are:

1. Why we need to reconsider computer intelligence in view of human and other natural intelligence.
2. What makes intelligence.
3. How to implement more intelligence using more refined order relations,

logic, more refined and complex structures, and specialized sensing-, action-, and computation-means.

We stress three key points in developing artificial intelligence:

- finding suitable explanations of what intelligence is and how it forms and acts;
- building structures of hierarchies that are suitable to accommodate complex processes;
- imbedding enough "intelligence" in the structures and sub-structures thus formed.

The idea flow in this chapter is: i) we stress some limits of current approaches, then ii) we evidence some of the conceptual missing pieces, and iii) we introduce some tools to overcome these limits, and to implement the new concepts.

6.1.2 *Terminology*

We shall use the following terms, with the corresponding meanings:

- Semiotic (semeiotic): science of signs, their production and interpretation.
- Semantics: the field dealing with the significance (with no direct reference to 'physical' signs, i.e. to the way significance is carried).
- Sensiology: refers to analysis of sensitivity mechanisms (basic level), plus to their representations.
- Communication: refers here to all levels of communications, moreover to the transmission (channel) issues.
- Secondary language: a language that is used simultaneously with another language, carrying information directly related to the message transmitted by the main language used in the communication. Gesticulation, when not used alone, moreover when used to supplement and increase the impact of the message carried by the main language, represents a secondary language.
- Side-language (annex language): a language that is used simultaneously with another language, carrying different massages than (unrelated information to) the message transmitted by the main language used in the communication. Gesticulation, when not used alone, moreover when used to carry messages not directly represented in the main language, is a side-language.
- Subliminal language: a language that is used simultaneously or not with another language, possibly carrying the same or different massages than the main language, moreover which is not consciously perceived and not

addressing our consciousness. These languages are more numerous than are generally considered: smelling languages, taste language, melodic languages, rhythm languages, fall into this category. All the three categories will be referred to as parallel languages.

6.2 State of the Art

6.2.1 *General*

AI includes more philosophy, psychology, sociology, medicine, natural languages, common sense and qualitative reasoning than it includes classic numerical mathematics, differential equations and geometry. AI should in fact develop as the theory of personality and common sense, in the first place, because these are more related to what we currently name *intelligence* than numerical mathematics or chess. Agripa [Agripa, 1530] stressed that "*Arithmetic is not less superstitious than futile.*" The sense of this assertion relates to the fact that classic mathematics can not account for the complexity of the human reasoning and behavior, nor solve all the problems related to humans. This is a lesson to remember for AI.

History and literature are rich in stories about humanoid machines, thought reading machines and similar replacements of the humans. We try to show some of the features that make the difference between machine intelligence and the human intelligence, and mainly how to reduce this distance. The gap seems so large that we need to analyze the differences, and the specific features of humans that common machines have not yet acquired, but that they could have in a near future. In the last part of this chapter, we try to give some more technical insights. The field is vast, and it is related to humans, to biology, to psychology, and to several philosophical issues. One of the possible mistakes, moreover a possible source of misunderstanding in such an approach, is the schematic dealing with the topic.

Sophists and followers of Democrites in Ancient Greece were stressing on "sensations" (sensorialism): soul is sensations. Part of our approach could be related to this trend. Indeed, we argue that computer need sensations in order to be more intelligent, more human-like, more sociable, and finally more communicative. We defend the point of view that present computers are primary limited because of lack of appropriate sensors, meaning by this: sensors (in the technical) *plus* appropriate information processing. Therefore, the first essential point in our discourse is the *artificial sensitivity*.

In contrast to the sensualist viewpoint, the rationalism of Socrates, Plato and others put in front the concept of manipulation ability. This could be traced to the "Artificial intelligence" of today. Computer Science and robotics has been involved mostly in this direction; it is why we do not need to add much. Starting with Aristotle, *structures* gained priority. Science (mathematics, psychology, and psychiatry) evolved in this new direction during the last two centuries: toward *relations and structures*, including Piaget's structuralism, modern algebra, or sociology. The trend is paralleled by Minsky's theory of frames, to object-oriented programming, and to several other issues in AI. However, the trend was active at the software level only, not at the system (computer) level. We believe that we have to make a great departure from these *data structure* and *concept structures*, and to introduce the paradigm of *computer interrelationship and sociability*.

The "wholist" paradigm is not so new - it probably can be traced to the Middle Ages. It is maybe also reflected (in the broad sense, as a thinking paradigm, not as a specific theory), to characteriologists. This way of thinking is probably very far - and the most unacceptable - by computer scientists. We neither argue for a Montaigne of the Computer Morals, nor the La Bruyere for the Computer Characters. However, we will defend this point of view asking for *computer personality*, at least as much as *needed to improve communication*. In psychology and psychiatry, the concept of bio-psycho-social paradigm in describing man and psychic maladies was supported by several scientists (see [Brânzei, 1975]). We argue for a similar concept, which could be named *sensorial - intelligence - sociability* of computers. Finally, we stress that for every aspect of the above, both new types, specific hardware and software supports are needed. We will deal with some ways and receipts to go from general ideas to actual realization of such a computer type.

6.2.2 *The Program of Patrik Winston*

Artificial Intelligence has the roots in ideas extending back to Leibniz, Pascal, Babbage, and Turing, among many others. The main aim of all those initially interested in building computing machines was, in the first place, to build artificial intelligence, although the term was not yet born and the difference between *computation* and *intelligence* was unclear. The past thirty years saw a burst of activities in which conceptual innovation plaid an essential role. However, the artificial intelligence field has not a definite identity and is changing at a fast pace.

In his seminal book on "Artificial Intelligence" [14], Winston presented a program of goals to be reached by computers and by A.I. in the future. We briefly review this program. According to Winston, in the future (after 1977), computers could do: "i) in agriculture, cut branches...; ii) work in mines and undersea; iii) enable large scale robotics in manufacturing; iv) plan people and teams activities; v) correct documents; vi) in schools, help students to correct, provide teaching aids etc. vii) in hospitals, help diagnosis, monitor patients, decide treatment, and arrange beds; viii) at home, help cooking, advise the house-keeper, laundry etc." (abridged quotation.) Winston says "*Of course, nothing of this is possible now, but AI could help.*" Winston's program, although sometimes a little confuse, was right. Robotics allows us now (i), (ii), (iii), partly (viii). ("Partly" because of cost reasons, moreover due to some limits due to the complexity of the tasks.) CAE and multimedia almost solve (vi). Text processors solve (v), at least partly. Expert systems, decision support systems etc. solve partly (iv) and (vii). Moreover, many other things are done.

We try to explain why some tasks are only partly fulfilled. Let us look what remains to be done, or better, establish a program to accomplish in the future 10 or 20 years. We say "partly," because some tasks were solved at the surface level only.

Point (iv), regarding people and teams planning problem, is far from being solved at the deep level. This task asks for a good understanding of human beings, of good *interactivity* with them, and of judgments far beyond present machine capabilities. Point (vi), regarding helping students to correct (we understand this task far beyond the "error detection" in scores machines used in examinations) is a completely different level problem than providing electronic books. The second task is an almost robotic one (if the writing of books is not considered), while the first task is difficult even for good teachers: it needs the deep understanding of student knowledge acquisition process, and students' psychology. Regarding point (vii), the discussion should be much refined. In hospitals, helping diagnosis by expert systems is today a reachable goal, and one chapter in the volume by Winston deals with MYCIN and other developments. However, Winston was dissatisfied with that state of the art, at least for the limited capabilities of such system in diagnosis, and for the inability to decide treatment. Today, these are still unsolved or partly solved problems. Moreover, we are far from being able to conceive a true clinic diagnostic by machines: they perform today only some limited "pre-clinical / laboratory diagnostic" as the doctors name

it. There is no way to say the machine is able to determine psychiatric diagnosis, or clinic diagnosis. And the highly precise, although "low level" task of arranging beds, with disabled peoples on the bed, is still far from machine abilities.

6.2.3 *The Review by Heer and Lum*

In the chapter suggestively titled "Toward Intelligent Robot Systems in Aerospace" [6] present a history as well as a perspective of the future of AI, as related to robots. According to them: *"Developments in AI have flowed into three relatively independent areas:*

i) computer programs that can read, speak, or understand language as people use it in everyday conversation (natural language processing);
ii) smart robots and programs that can sense and understand the environment and perform goal-oriented operations; and
iii) computer programs that use symbolic knowledge to simulate the behavior of human experts, i.e., the expert system (ES) or knowledge-based system (KBS)."

Although it is a nice summary, it should be amended, because it is in some respect too optimistic, moreover it offers a too restrictive point of view. The aim (i) is much too enthusiastic: machines are far from really understanding natural languages, and very far from understanding the richness of language as conglomerate of languages (with intonation, gesticulation and other secondary, annex and subliminal languages). The same is true regarding machines speaking like humans – they could not be able to speak like humans before being equipped with emotions, intentions, beliefs, temper etc.

The aim (ii) is more realistic – to some extent, computers can partly sense the environment and schematically, goal-oriented interpret it. Moreover, they can do some more or less basic operations, i.e., operations that can be or not precise, but that do not carry much intelligence, or high complexity. However, still, the degree of complexity of the robot movements is much lower than the movements of a fly, or of a bug.

Regarding the imitation of the human expertise, much disappointment exists last years, though indisputable advances were performed. Anyway, computer expertise is not yet true human-like expertise, and even if one accepts it as human-like expertise, it is quite restricted. Indeed, only a few ways of building and using knowledge by humans were until now implemented in machines.

6.2.4 *Some Conclusions on the State of the Art*

Heer and Lum quote Aristotle saying: *"If every instrument could accomplish its own work, obeying or anticipating the will of others ... if the shuttle could wave, and the pick touch the lyre, without a hand to guide them..."* Heer and Lum conclude: "*What Aristotle supposed, we can do*".

Actually, we can not. Still, there is no shuttle waving by itself – and many human driven shuttles still disappear in the storms. In many respects, we are not fundamentally more advanced than in the 1920-1930 period, when feedback control and remote control allowed toy shuttles to move around on the lakes, or compared to the 1940-1950 period when the first (rudimentary) self-guided missiles flight in the Second WW. There is no pick touching the lyre much more human-like than in the mechanical music boxes of the 17th to 19th centuries. And, surely, no machine is able to *anticipate* humans' will. We have to conclude, rather pessimistically, that we are not fundamentally more advanced than 50 or 200 years before, and that the program of Aristotle is far from being accomplished.

A more realistic approach, quoted by Heer and Lum, is due to Wiener. The last had the idea to inverse the problem: define *machine-like* activities as the goal of machines. He is quoted as saying: *"People should not perform like machines. If a job is machine-like, a machine should perform it."* However, we have not much advanced: are "machine-like activities" those activities humans do not like? Alternatively, are these activities those that people think that can be performed by machines? In the first case, there is no relation to machines. In the second case, the definition becomes a circular one. Finally, is "machine-like activity" one reachable today by a machine, or reachable in the future? The above discussion is intended to reveal some of the main limits of current approaches. These limits will be discussed in more detail in subsequent sections.

6.3 Several Desirable Properties and Current Limits of the Current Machines

We summarize below the main concepts that base this approach and possibly allow us to partly eliminate the existing limits. We regard the respective desirable properties as some of the essential, but missing properties of the current machines.

Machine at the basic level: machine is sensitive (in the sense applied to humans, namely, in the sense that it has sensations).

From machine to user relation: i) recognition of users' bio-psychic states; ii) recognition of users' personalities and of every user as a bio-psychic personality (not merely as a physical entity); iii) recognition of users' group relationships.

Inter-relations: iv) machine as entity in a couple; v) machine as a group member; vi) machine as an evolutionary entity; vii) machine as a socialization developer (group and meta-group organizer);

Advanced personality: viii) reflexiveness: it has a set of reflexes; ix) emotivity: develops patterns of sensations; x) behaviorality: it has a specific behavior; xi) morality.

In brief, a machine should include and exhibit:

- human-like variability induced by different mechanisms: due to time (hour, season, cyclic time), influences on its state, human-like state of fatigue, recent history, and random variability;
- "behavior", in general: basic rules of behavior in a group, in a specific relation with partners, at work, in society etc., as well as adaptability, changes and variations of the behavior;
- temperament;
- "relationship": establish, detect, improve etc. relationship, as used in its social behavior as well as in its "internal life" and evolution;
- evolution.

It should be obvious that the main questions related to how to make computers more intelligent are questions related to humans and other natural beings. The answers to questions related to humans, animals and insects have to be transferred to machines.

Knowledge about how humans are and how they behave, and knowledge about how to make a computer to behave like a human, how to make it sensible and responding to human behavior are intimately related. Moreover, these knowledge bodies will grow together. Indeed, computer is asked to watch, detect, classify and recognize human behavior, thus increasing our knowledge about how we are and how we behave. In turn, this knowledge will be used to make the computer to behave in a human-like manner.

6.4 The Sensitive Computer

6.4.1 *The Sensible Information*

This discussion is related to the meaning of "information", which is basing the

"decision". Information becomes itself multi-dimensional. There is a semantic information, coded in some specific way, there is some "linguistic" (symbolic) information, and there is some *relational*, *affective*, and *sensible* information. The information, in the broader sense, is multi-component.

6.4.2 *Sensitivity*

A *sensitive machine* is defined as fulfilling some basic requirements an ordinary computer, or a today "robot" does not:

- to imitate a specific behavior, depending or its experience;
- to be able to imitate low level stimuli reception, (i.e., stimuli such as physical, thermal, chemical etc.), and to produce complex responses, affecting its behavior,
- to be able to mimic high level stimuli reception, such as voice, images, changes in the above and to produce complex responses, affecting its behavior,
- being able to mimic sensible responses to behavioral stimuli.

By *behavioral stimuli,* we understand all the information content in stimuli that is not directly related to its main semantic content. For example, in voice messages, all the information in frequency, tonality, loudness, accent etc. that is not a part of the meaning of the sentence, or of the word pronounced. It is known that this amount of information is higher than the semantic rest: it is why it is so difficult to get the semantic context by simple methods (related for instance only to the frequency spectrum of the sound).

6.4.3 *A Simple Example of Sensitive Machines*

An obvious application of a *sensitive machine* is in creating a *socialized* machine, that interacts easier, less stressing, more human-like, and especially more accurate and more efficient to human requirements, and to specific human communication needs.

The simple application previously addressed in our research was related to speech synthesis. The aim was to create methods to make speech less boring, less stressing for humans. To achieve this aim, one proposed method aimed to add to the semantic content some 'mechanical sensibility' in voice signal, i.e. some information that is as in the human speech, but generated without some specific content, although including typical specific information. The primary goal was to create a speech synthesizer that is *adaptive to the ambient*, moreover has a human-like variability in speech. The simplest case is to create noise adaptability similar

to the one used by humans, moreover adding the variability.

It is known that humans adapt both the spectrum and the loudness of their voice in presence of ambient noise. The most important adaptation consists in change of spectrum content and consists in shifting the whole spectrum to higher frequencies, as well as increasing the ratio of higher to lower frequencies. Such an adaptation was suggested in many papers. Variability – that prevents stress in humans – was induced taking into account the chaotic and random components in speech.

6.5 Perception, Self-representation, and Self-relating

6.5.1 *Differences between Machines and Living Beings: What Living Beings are doing and Machines do not*

To help the intuition, we start by discussing an example. Suppose in John's room are Taylor, John's friend, John's dog, John's boy's cat, John's friend's dog, a fly, and John's computer. The friend is waiting for John and he is reading a book, being seated in a chair, turned with the back to the door. The cat is staying on his legs (the cat can not directly see you entering the room). The dog is almost sleeping, closed eyes, under the table. A little apart stays John's dog. The fly is staying on the table. John's computer is on the table, the screen looking to the door. Imagine what happens when a person enters the room. Already the sounds generated by the door opening yield the following reactions:

Taylor turns his had, sees his friend and says "Hello", etc. The dog recognizes *his master*, he is the first to hear the person coming and rises a little his head and moves his tie. The cat sees *a family person* and closes again his eyes. The friend's dog recognizes *a friendly person* and slowly moves toward him to smell him. The fly, due to movement around, is possibly a little afraid: it has no friends, but *moving potential enemies* around; so, it probably flies away. The computer is doing nothing.

Now, suppose that instead of John, the master of the house, comes his friend, Roger. He never met Taylor or his dog, but often met John's dog and cat and worked with John's computer. The reactions of the actors will be completely different, except for the fly, which is doing the same action, and for the computer, who is doing again nothing.

Notice that all attributes as "his master", "friend", etc. denote *interrelationship*, as perceived as the subject. Characterization and recognition *can not* be independent on relationship. This is a great departure of the classic,

"objective" recognition process, where subjects are related to (and dealt with) objects, or symbols (even if named such as "John", "this dog", "that cat" etc.). We argue that at least in *communication oriented recognition*, the relationship between recognizer and recognized is the first and possibly the main part of the process, and possibly its main goal.

Let us analyze this example at a more basic level. Higher level actors (friends, dogs, and cats) react *in essentially almost the same way*, namely:

- are getting information of changes in situation;
- evaluate the change → somebody is coming?
- categorize "somebody" = known/unknown;
- if known → recognize "somebody" (this operation can be considered part of the previous one);
- determine relationship to "somebody": master, friend, familiar, known person, enemy, nothing – for the computer; (this operation can be seen as associative recalling, if the scheme of associative memory is accepted);
- establish appropriate behavioral strategy according to the *relationship* and according to the *current situation*;
- establish appropriate series of actions according to the *relationship* and according to the *current situation*, generally *before* the person acts in some way (i.e., *not as a response to the person's action*);
- react to people (and machines) response and actions etc.

Consequently, the main point is that higher level beings *first* detect a new situation, *second* establish the *relationship medium and actions*, and *third* react. Unfortunately, the second step is always omitted in researches on man and computer, as well as in many man-to-man communication studies (or it is dealt very schematically). Also note that the possible number of "attitudes" (i.e., acting patterns *in relationship to*) is depending on the number of elements in the set of possible relationships:

{*unknown person; son; ...; from the family; friend; close friend; master; friend of my friend; some person I met; enemy; ...*}.

A human has a richer set of relationship than a dog or a cat. Therefore, one could expect a richer palette of attitudes from humans. Consider the fly (a lower-level being). Its reaction includes only two of the above steps: the inter-relationship set includes only a few classes:

{*neutral* (dead things, not moving); *other flies*; *"small" animals* (not flies, not enemies); *enemy*}.

Consequently, the actions of the fly are much simpler. Of course, this is a schematic set and experts in flies can improve it. However, the main idea, of

poorer set of relationships is clear.

Finally, the computer set of relationships is void: it establishes no relationship. Indeed, although one often says "man-machine communication," there is no true communication between them, because communication fundamentally includes inter-relationship.

We notice that one often says that there is a "computer-user" relationship, but this is fooling: it is *the user* (human) who *thinks* such a relationship between him and the computer, not vice versa. The human user always tries to establish a relation of him to the world. This is specific today to life and living beings, but was not included in and extrapolated to existing machines. Concluding, one should ask to a machine to establish 1) a quasi-permanent activity and attention (even if at lower level), and 2) a relationship to everything around. Regarding the first request, notice that any living being is almost continuously checking and scrutinizing the environment, while a computer does not – a situation that should – and can – be corrected easily using current technology.

6.5.2 *Another Example of What a "Sensitive" Computer should do*

We detail the discussion above, making again a simple scenario, aiming to get an insight of what a machine can be expected to do (and currently it does not). The technical means to allow us the computer to do all this will be briefly discussed in the subsequent sections.

The scenario is as follows. A research assistant (the equivalent of the computer) is sleeping in the laboratory. At 6 p.m., someone, (the professor), comes to the laboratory. What the assistant does:

- He/she detects that the door opens, that some sounds inside the room were produced and that light is turned on.
- He/she analyses events, for example if a person is coming.
- He/she understands (this is more than a yes/no switch) someone entered the laboratory and determines if the person the professor, or some colleague, or a student (categorizes, i.e. evokes possible categories, and recognizes); activates (partly awakes) itself, in a way appropriate to the event.
- He/she interacts correspondingly to every category and to the specific person (two-stage strategy of action; two types of knowledge involved).
- He/she reacts correspondingly to actions (questions, requests etc.), taking into account the category of person and the specific person. (Again, different knowledge types are used: related to questions, to requests, or to requests for actions etc., moreover, related to a specific category of asking persons,

moreover specific to some person, according to the existing inter-relationship.)

- If several persons are coming, then the responses will be more or less changed, according to the homogeneity of the inter-relations the system has with them. For instance, the way of answering, and the answer content is changed if the professor comes with a student, or if a friend is coming, or if students come, or if students and friends are coming.

The following is a summary of what a computer (= machine research assistant) may be expected to do as a minimal response– similarly to a human being – if placed in the position of a human research assistant.

- It detects that door opens (new, specific sounds in the proximity? sounds coming from the door direction?), that some sounds inside the room were produced (sounds close to computer? inside?) and possibly that light is turned on (change of light condition).
- It analyses if a person is coming, or a different event happens.
- It understands that someone entered the laboratory; activates (partly awakes) himself.
- It recalls the information about types of persons; recognizes the actual person.
- It interacts correspondingly to every category and to the specific person: is it "the master"? is "allowed user", is it a "partly allowed user" (restricted user), or is it a "not allowed user"?
- It reacts correspondingly to actions (questions, requests etc.), taking into account the category of person and the specific person (gives free access to computer memory, working power etc., or gives partial access or no access).
- If there is more than a person coming, then the responses will be more or less slightly changed, according to the homogeneity of the inter-relations in the group of persons coming (for instance, does not give free access to "the master" if there are not-allowed-users in the room).

The above scenario and its discussion evidenced part of capabilities a computer should have and acts it should do. In the following, a simplified scheme of what a computer *has not* is presented:

- It has no sensing ability (sensors to detect environment, and user, beyond his simplest commands).
- It has no attention.
- It has no user-specific reaction.
- It has no specific inter-relationship with a *specific user*.
- It has no specific (in comparison to other computers) inter-relation with

(the) user(s), and no different user-dealing paradigm.
- It has no own (self) consciousness.
- It has no group consciousness.
- It has no evolution.

6.5.3 *A More Detailed List of Desirable Features*

Based on the above discussion, we revise the desirable capabilities and the acts of a computer, in accordance to the main features:

A. The *perceptive* computer should:

i) sense the environment (by environment one understands here the whole environment except the partner(s) of communication).
ii) detect the interlocutor(s), identify it (them) and establish relationship;
iii) detect at least several parallel (annex, secondary, subliminal) language communications, and understand them together, and in relation to the spoken (linguistic, main) message;
iv) at the technical level, to include technical means (sensors, appropriate software) to perform the above operations.

B. The *communicative* computer should:

i) have the ability to generate secondary (including side- and subliminal) language messages;
ii) adapt to the partner and to the his/her specificity;
iii) recognize specific states of the partner, and adapt to them;
iv) recognize specific attitudes and goals of the partner and adapt to them;
v) be able to recognize groups of partners, their relationship, and to establish a *group relationship* (with that group), i.e., has ability to report itself to the group;
vi) adapt to the specificity of group communication;
vii) at the technical level, to include technical means (sensors, appropriate software) to perform the above operations.

C. The *embedded-consciousness* computer should:

i) be able to identify itself;
ii) be able to manifest itself in a different way that other machines;
iii) be able to create and structure its own "life experience"; and possibly:
iv) be able to learn its own goals and strategies.

D. The perception and the representation systems. In every communication, even in any action, there are – and should be – present:

i) the perception system;

ii) the self-reporting, representation system, including the interpretation system.
The last feature is *constructive*, in the sense it should be built by the computer itself, during its life.

6.6 Relationship and Relationship Representation: Key Factors in Intelligence and Communication

In this section, we propose what we believe to be a part of the solution to the increasing of the intelligence of machines. Moreover, this section aims to indicate some major differences in human-human communication and behavior, when contrasted to man-machine communication.

6.6.1 *Main Characteristics of Self-representation and Self-relating*

Every communication has a "header", giving information on the relationship and the bounds between the communicating subjects. We suggest that every message have two parts: the declarative part, and the subjacent part. The declarative part may be considered the "direct", "conscious", "intended" part of the message, while the subjacent part reflects unconscious, non-intended, subjective information, related, for example, to the relationship of the message provider to the topic of the message, to the message receiver, to the state of the message generator and possibly to the general context. This part of the message is carried mostly by secondary-languages. The relationship, as well as the message, may include a sentient, conscious part, and a subjacent part that we are not aware of.

Regarding the relationships, we conjecture that:

1. In any human communication, the first, very basic act is the self-reporting, and reporting to the partners (action / communication partner).
2. The result of the self-reporting, and of partner-reporting *is communicated*, moreover it is *a very important part of the communication*.
3. There is a great number of "channels", instruments in communicating *this* part of knowledge.
4. The instruments are not so efficient regarding the channel capacity, they are slow; nevertheless, because of intrinsic *redundancy*, they are they are very reliable, much safer than other communication means.
5. The "alphabet", or better saying, the dictionary of relative positions is very rich.
6. An essential aim of communications, in human-to-human interaction, is

the communication of the relationship and of the inter-relationship. This is the basic message, and probably it comes from a very long time ago, and a long way from phylogenesis. Probably it is not acquired during ontogenesis, but it is embedded (innate).

We conclude that present machines completely lack the capability of self-reporting and of partner reporting, as well as the possibility of the inter-relationship communication. This is possibly the main reason they are perceived and are "non-socialized", "cold" etc.

At our best knowledge, these hypotheses, although in some respect already envisaged by other authors, were never formulated. There are few directed researches carried to validate them. However, they seem to be indirectly validated by empirical observations. An extensive research in fields like psychology and sociology is needed to clarify several issues related to the above discussion.

The relationship representation and communication are dynamic processes. Any human communicator performs two tasks during communication: the communication itself and the trimming of the inter-relationship. These tasks are dynamically performed, according to some rules. Two main types of dynamics may be considered.

i) If the two communicators meet for the first time (no previous relationship already established), the starting point of the inter-relationship is generally neutral (formal, for instance: formal-polite), or is dependent on external, circumstantial, conjuncture-related factors. For example, such circumstances may be reflected in relationship of disturbed, angry-by-disturbance man / disturbing man, or scared by sudden occurrence of someone else / sudden occurring man etc. After some time, by evolution depending on the communicated messages, the inter-relationship tends to get a specific character (in the alphabet of inter-relationships), if the messages remain in a given class. This is the *long-time* stability (limit, limit point) of the communication relationship. It could be also named *frame of relation*.

ii) If the two communicators already met (previous relationship already established), the starting point of the inter-relationship is the limit, long-time stable relationship. The relationship is modified, is submitted to fine, context-dependent adjustments, getting short-time characteristics. The specific types of these relationship forms are decided by the long-term inter-relationship. This means, for instance, that if the friendship-relationship is already established, the variations are according to this frame.

We suggest that in most cases, the essential question be not whether or not context can be computed, but if whether or not we are able to represent and "compute" relationships.

6.6.2 *Relationship Communication*

An important question is how relationship is communicated. We already argued that probably the most important means for communicating this type of message is not specific to humans, but comes from the deep history of life. Possibly, this is one of the most specific features of life. It also refers to self-reporting, that is, to the roots of the consciousness. Animals exchange such messages, at a more basic level. We have no knowledge about, but it would not be surprising to learn that such message exchange exists for protozoan too – even if it as elementary as a secretion of one of the some two or three specific substances the protozoan is able to generate outside itself. It is also acceptable to conjecture - with rather low error possibility - that even the immunity system of animals and plants is based on such messages. Maybe the understanding of such languages will help medicine and psychology as much as A.I., some day. Coming back to the question of communication "secondary languages" (because there is surely more than one), one can systematize:

i) sub-speech messages, for instance carried by modulation, melody (accents), ratio of high to low frequencies, pitch etc.
ii) gesticulation (gesture): head position, face expression, hands movements, body movements, speed of movements, lack of movements, readiness, even playing etc.;
iii) touching and related messages (kissing, caressing etc.);
iv) finally, speech-related messages, expressed by using: a) some stereotypes of language, b) using some specific expressions from a set of possible ones, c) some non-formally informative, but intentionally informative sentences, e.g. "thank you" (no factual meaning, but relational meaning), *"good morning", "how are you", "well! well! well!* etc. d) use of interjections (*eh! bah! oh!* etc.);

Other languages may exist to carry the relational messages.

Some examples of using specific expressions from a set of possible ones, with the intent to communicate the position of the speaker with respect to the topic or the listener are shown by the following sentences, which formally have the same meaning and carry the same formal message, but carry different annex information.

"- *No giving, no getting. / - You don't give, you don't get. / - My dear, you can't get without giving. / - I would like to give you if you could get me something. / - Rubbers wish to get without giving.*" etc. Obviously, the (main, at least) difference in meaning is the position (cold, ironical, moral, paternal, aggressively moral) of the speaker toward the listener.

6.6.3 *Some Types of Relations: "Mood-related" Relationships*

Beyond the relationship, and connected to this, the communication includes information about the affective position of the communicators with respect to the discussed topic, and with respect to the way the partner is responding to the subject of the conversation (instant response). Such communication (adjacent message) includes agreement, disagreement, surprise, comprehension, doubt, strong approving, irritation, fury, disappointment, sense of triviality of the communicated message, amusement, pleasure, raillery, enjoyment, delight, solemnity etc. Note that all these words all express exactly – and only, for the most of them - attitudes during communication. Such messages are richer for rich "communicative personalities" and are, in general, an increasing function of the cultural level of the communicator. Several other "mood-related" relationships, affecting the behavior and the communication, are:

i) affective relation: a) unbounded (love); b) sympathy; c) antipathy; ...
ii) collaborative: a) polite; b) professional; c) could but constructive; d) helping; e) paternal;
iii) neutral: "mechanical", machine-like (the one "implemented" by machines, or similar), objective;
iv) "active": rewarding; punitive; vengeance;
v) including one's side superiority;
vi) giving credit, or with disbelief on one's side;
vii) parasite (collaborative only from one side);
viii) "in force".

Each type of the above supposes different reciprocal adaptation strategies, i.e., response strategies.

The "annex" messages regarding the position representation are sometimes more important than "direct" the message itself. They communicate more than the logical sense of the words. A simple "Yes" or "No" includes less information – only general information. However, the ways of pronouncing the "yes" or the "no" can be very rich in meanings. There is no true human communication without these messages. Good novelists obviate their need, when describ-

ing the dialogues: they include much information about these states.

6.6.4 *Examples of Self-mood and Position Messages*

For instance, the following segments of a book show the importance of side- and subliminal-languages, and ways to carry annex messages referring to the mood and the self-representation of the speaker (if speaking!):

"Sir Henry *stared at* her. [Notice that *starred at* represents an action, gesture, that is part of the side-language in communication.] 'Why did the plan go wrong?'...Miss Marple said rather *apologetically*: ... 'I must say', said Sir Henry *ruefully*, 'that I do dislike...' Miss Marple shook her head *sadly*.... Sir Henry said *distastefully*: '....' There was *a moment's pause*, and then Miss Marple resumed. ... Miss Murple *interrupted* him. ... He said *heavily* ... Jefferson said *skeptically*... Sergeant Higgins, *rather breathless*, stood *at Harper's side. A flash* came over Harper's heavy features... Raymond *considered.* [a pause, with the significance of considering] ... Harper said *sharply*: ... Harper *nodded*... "(All quotations are from a few pages of "The Body in the Library", Agatha Christie, Phoenix Publ. Co. Paris (Sherz & Hallwag, Berne, Paris, London). [No year of publication.]) Notice some of the communicating side- or subliminal languages in the fragments above:

- *breath*: breathless, i.e., stop of breath for a while = deep surprise; breath stop and puffing = angry; deep breath = angry at the limit, or sorrow, etc.;
- *relative position to the other*: in front of him, aside, taking some distance, going away, etc.
- *way of performing some related activity*: "Slack [a subordinate of the inspector Harper, that is present] had *carefully* noted *all* the names mentioned." [A. Christie, ibid., p. 91].
- Way of verbal reaction: one speaker *interrupted* the other. We are interpreting here "interrupting" not in the simple meaning, as action of interrupting, but in a complex sense. Indeed, when someone interrupts another one's talk, he also does it in a specific "acoustic"-"melodic" way, by increasing the high frequency and the power of his voice, and by talking "abruptly" (short, quick vowels). Moreover, he/she tries to impose his/her ideas as a way to impose his own personality.

Note that even *performing / not performing specific activities, and the way of doing them, is a true language.* This language is most important in societies that have a great deal of hierarchy, and is related exactly to the communication of messages regarding the rank. For instance, taking his notebook and writing

down has no meaning in general, except the interest of the writer in preserving some information. However, taking his notebook and writing down (*attentively, painstakingly, with great care* etc.) what another is saying *and* in his presence is equivalent to one or several messages. Such messages are: "I am obedient", "I will strictly follow your instructions", or "You are my teacher" (I noted this last meaning to students in Japan), or "I am an applied and hard worker", or "It is interesting what you are saying."

There are many parallel (side- and subliminal) languages in conversations and in communication in general. It is time to address the question what makes the difference between conversation and communication. Some can understand the difference in the classic sense that communication is necessary informative, while conversation is not. Some others, more pragmatic (technical, engineers) can say that communication is just the *optimized* conversation, i.e. pure message, information transmission. Conversation, some say, needs a specific natural language and supposes human beings, while communication can be performed by numbers, bits etc., and can be (even better) supported by machines. This dispute has partly meaning only if one excludes the relational information carried in a dialogue that, maybe, carries no factual information. A dialogue is rather a *tool for building* a relation, not a tool to convey factual information.

One can accept the paradoxical say: "The main dialogue with the patient is mute." [7]. The meaning of this say is that talking to a patient to ask him about his state, the doctor has to watch, in the first place, the parallel languages, which are much more informative about the state than the spoken messages of the patient.

6.6.5 *Multiple-relationship*

The interrelation is context-dependent: one can be financially, wealth-related superior to someone, but at the same time, inferior from the point of view of the knowledge (less knowledgeable, spiritually inferior), and be aware of both of them. In some context, the first may be essential, while the second is unimportant.

6.6.6 *Computer Semiotics: Based on Relationship Representations*

There are many new disciplines in computer science, which are trying to mimic human abilities of intellectual work, as well as human communication. The multimedia, infographics, virtual worlds, speech and graphics interfaces, learning theories, computed aided instruction and education - to name just a few.

These look to be rather divergent, because the methods they use are tied to computer-related specific ways of understanding the problems, and not to their basics: the semiotic. *Computer semiotics* and *computer semantic*, related to human semiotic / semantic, does not exist yet. Computer semantics and semiotics may become a necessity because new types of relationships are established for computers, differing from the human-type relationships. Therefore, new (artificial) meanings and related signs will be established, asking for specific semiotics and semantic. Specific features may produce a large departure from the human semiotic. In his *Treatise of General Semiotics*, Eco is offering a view on the human semiotics only. We argue that the "general semiotics" has to go beyond the specificity of human semiotics, offering an insight that is more general, allowing different types of animal- and "artificial-semiotics" that can be developed by computer, or related to computers, and possibly learned by the humans too.

On the other hand, the first goal of the computer semiotics is to show how computers can cope with human semiotics miming. Only after surpassing this stage, computer science will be able to generalize semiotics. However, we should note that synthesized music, as well as virtual worlds, are potentially able to induce in humans, as well as to propose themselves *new semiotics* (in the form of new human-computer related semiotic codes). This topic, though beyond the interest in this chapter, is an exiting field of research and it is hoped it will be developed elsewhere.

We agree with Suppes [10] that "*...The meaning of a word, or of a sentence, or of a utterance is private and probabilistic for every individual... This does not negate the important public aspects of the language, but asks for an explicit theory of the communication referring to the way the listeners understand the speakers and to the way the speakers check this understanding...*" In addition, Suppes emphasizes the importance of "*assent*" and a complex assent that can not be reduced to the belief or agreement or logic "yes". Indeed, "assent" is the representation of the relation of the subject toward the (logic) validity of an assertion. This type of relationship is not discussed here. Interesting enough, Suppes has the intuition of annex and subliminal languages, stating that "the production and the reception of the verbal utterances is not controlled by the conscience." However, stating that the "*meaning is ... private and probabilistic,*" and developing a purely probabilistic theory of the meaning, Suppes fails to determine the complex way of representations of the meaning, and its intricate levels, related to the subject, his/her self representations and his/her relationships to the world.

We also agree with Draganescu [4] who emphasizes that the meaning is phe-

nomenological and non-formal. Moreover, we agree with Bacon [2] who (in his "*Cogitata et visa de interpretatione naturae sive de inventione rerum et operum*") talks about the "vague and undefined nature of the words that plays with the intelligence of the humans and, in some way, tyrannizes it...." In Plato's tradition, Francis Bacon (in his "De interpretatione naturae sententiae XII,") insists on the humans as *interpreters* of the nature and facts. All these quotations show roots of the concepts presented in this chapter, or similar ideas for the case of humans. We are advocating for the expansion and transposition of these basic concepts to computers. To perform that, we need specific tools.

6.7 Methods to Embed Relationships

6.7.1 *Relationship for Machines*

Machines can easily be provided with some basic "relationship characteristics." In fact, even mechanical, classic machines, could be provided with some means to react to the specificity of an individual partner, for instance to adapt to the speed in reaction etc. Examples of ways to endow machines with basic relationships are summarized below, for the case of a simple example.

Step 1. Determine physical-level features of (all) persons nearby: voice parameters; image: face, and body characteristics; gesture characteristics; general movements characteristics, etc.

Step 2. Characterize at the basic (physical level) the person(s): features extraction.

Step 3. Identify any known person; label "unknown" the others.

Step 4. Characterize the bio-psychical state of the known persons; whenever possible, of the others too.

Step 5. Characterize the relationship between persons.

Step 6. Identify messages.

6.7.2 *Endowing Machines with Relationship Capabilities*

The first condition is to provide the computer with representation means, for various levels of representations and for various classes of occurrences to be represented. More specifically, we have to provide the computer with the ability to represent knowledge related to every sense: smelling, hearing, seeing, tactile sense. (Taste is less important in the inter-relationship.)

The representation has several levels:

A. The objective level, as dealt by the classification and recognition tools (like the ones already existing).
B. Internal levels:
- the "*I like it*" level (pleasure, sympathy, well-living etc. sub-levels);
- the "*I remember it / it remembers me*" level (own history correlation level);
- the "*I know it* (*I am knowledgeable about*)" level – knowledge relationship level.

C. Internal-to-external level
- the "*he/she knows it*" level;
- the "*he/she likes/dislikes it*" level;
- the "*he/she has a good relationship with them/he/she*" level etc.

Besides the "external"-type characteristic, all these levels have a meaning related to the subject relationship. A mechanism for each of the above, moreover an integrating mechanism should be developed. A simplified example is as follows:

If "he is John" & "he is a nice person" &"every time he has a new, nice joke" & "I have not seen John for a long time", THEN

My attention will be keener & I am already happy & I am keen to relax now & it will be a rewarding meeting & better to forget my argument I had half an hour ago with my colleague Pete" & I am keen to discuss with someone & ...

Furthermore, the following mechanisms should be added:
- mechanism to update the relationships (relationships that are connected to the recent experience);
- mechanism to identify any other entity of similar or close type (same specie, or similar specie), moreover any "significant" object;
- mechanism to generate new relationships every time a new situation occurs, including the case a new being is identified, etc.

We conclude that this is a problem of complexity of representation and of integrating the representations. The "ego" is not factorable to its components; it is why it can not be reduced to components and has to have an integrating level.

6.7.3 *Complexity of the Representations*

Cognitive sciences were seen as the major breakthrough in the field, during the period 1980-1990. However, the promises were higher than the results. Similarly, the expectations of the fifth computer generation were higher than the results, possibly due to the low complexity of the structures proposed for the

"intelligent" systems. We need to adopt methodologies with a degree of complexity that is better suited to the task. Nowadays, based on the current state of the art, we can produce systems whose overall degree of intelligence is lower than that of insects, with the addition of one or a few several high level expertise, in a very narrow field. We can generate an entity having the general level of intelligence much less than that of a worm, or a fly, but with the expertise of an international master in chess. These are anomalies, created by the technique. Maybe this is a necessary the first step in the development of A.I.

We can contemplate the task of creating *intelligence* at a level similar to that of insects, for example, using the methodology presented in this research. Possibly, the can contemplate the creation of systems with the ability of relationships, *in a limited context*, similar to humans. However, we are far from creating a very complex intelligence, close to the human or to a mammal, because the complexity is huge and our knowledge about intelligence is too elementary yet. We have to wait for more progresses in human-related sciences, namely psychology, sociology, and cognitive sciences.

One of the main criteria that makes a difference between the classic adaptive systems and the "intelligent" ones is that the internal (or, sometimes, subjective) representation overwhelms the "objective", external representation the belief is stronger than the external reality, the "internal reality" is stronger than the external one. The actions of the humans are directed by the internal beliefs, not by the measures of the objective, but "external" measurements. The objective world rests outside the internal *ego*, it is always less important. The only "real" world to a human is his own, including his representation about the external world. The only "truth" is the subjective one. The only reality is the "virtual" one ("virtual" not in the nowadays sense, but in the sense it is subjective, made by the subject).

6.7.4 *The Machine as a Member of a Group*

To be able to perform relationships, the machine should have:

- a "library" of "personality" features of people;
- a library of classic circumstances of interaction with people, including the master;
- a library of human behaviors in classic circumstances;
- a "library" of relationships with people (master, friend, enemy etc.).

We considered above only features, personalities, behaviors, relationships and interaction related only to people. However, similar "libraries" may be provided

for other beings or specific objects (e.g., machines). The term "library" is used in a general sense, without care for the exact meaning. Details on the meaning will be provided in the sections on sensorial bases and relation bases.

6.7.5 *User Models*

The approach we denote by "user model" means, briefly: developing a "model" of the *types of users*, with the aim to adapt the response of the machine to the specific type of user.

The main problems of the "user model" approach are:

i) how to define the user categories, as related to the task to be performed in cooperation of the machine;
ii) how to identify the particular user type;
iii) how to build (the specific variations of) the responses of the machine, adapted to the user type.

The user's model approach is not new. It was considered in expert systems interfaces and in data bases interfaces in the eighties (for instance, Pejtersen A.M., Austin J.: Fiction retrieval: experimental design and evaluation of a search system based on user's values criteria. J. of Documentation, 39 (1983), 4, 230-246; quoted by [8]).

This approach is under consideration or development in several forms and at several laboratories. One of the limits of the current approaches is the conceptual starting point, namely the way of defining the user's behavior and his responses. As pointed out by [8], the main way of evaluating the user is by his *mistakes* (!), seen as acts contradicting the strategies that *the machine considers as optimal*. This leads, again as pointed by [8], to the very unnatural definition of the user "by his in-adaptation to the machine, and a huge energy is disposed to help the user to adapt to the machine" [[8], p. 188]. Therefore, these systems force the user to obey to the naive psychology of a hypothetical user that is just equal (in the best case) to the machine. One of the first ways to improve the situation, as presented in this chapter, is to eliminate this primitive approach.

The second limit of the existing approaches is the testing of the user: by his very constrained responses (both under very severe syntax and content constraints). No natural response and mainly no parallel (non-linguistic) responses are used. We try to alleviate this drawback in our approach.

6.7.6 *Speech and Personality*

"*The expression by words reveals an individual, his way of thinking and of*

manifesting toward the exterior" [[3], p. 17]. Here there are three independent problems:

i) "expression by words" should not interpreted as "expression by written words", but as "expression by spoken words";
ii) moreover, "expression by spoken words" should be interpreted as "expression by spoken words, with all connotations";
iii) "expression" both includes and reveals (communicates) information about both: a) the subject of thinking (i.e. reasoning results, let us denote this information: logic, objective information), b) indications of actions commanded by the speaker related to the subject of thinking (i.e., orders resulted from the reasoning, relations derived from the result, between the speaker and the world), c) relations related to the speaker and the listener.

Eco in his well-known Treatise [5] analyzed part of this information. He evidences in an example (Chapter 1 and Chapter 2) the connotation as related to the commands that are logically derived as consequences. In his example, "danger" has also the connotation "evacuation of population" (in his example, the message is transmitted by someone monitoring the water level in an accumulation lake, and the peril refers to imminence of flaws in the downstream). He speaks about superposed codes (also denoted as connotative semiotics). He differences between such connotations that are derived as a logical consequence from the actual message, and the "emotional" (as a counterpart as "referential"), or the "vague" (as a counterpart of "univoque") connotations. However, Eco fails in identifying the basis of these connotations, namely the self-representation and the relationship representations.

It is important to derive all connotations, including the "emotional", as related to subject position toward the matter, including its consequences, connotations: the voice change is crying "peril" is related to the degree of peril, as well as to the order "evacuate". However, the voice change will much depend upon the other types of connotations (information), for instance the relative position of the speaker and listener: more in a higher command position is the speaker, more his voice will be imperative. But, if the relative position of speakers is not allowing the speaker to command, the logical consequence connotation "evacuate" just disappears. Consequently, the information is more intricately related, and the connotations should be considered mutually dependent.

6.8 Technical Means

To cope with the complexity of the problem, we need tools of appropriate com-

plexity. [9] emphasized that there is a need for multi-track approach: "*There now exist at least a dozen well developed and different ways to represent knowledge. These include Natural Language, Semantic Networks, Rule-based systems, Logic Programming, Conceptual Dependency, Frame-based systems, Object Oriented Programming, Fuzzy Logic, Neural Networks and quite a few others.*"

However, we believe that many of the tools have yet to be developing. We suggest the following new tools. We suggest that data and knowledge bases should be complemented by:

- Sensation bases
- Relation bases: control the knowledge bases and the interface bases
- Representation bases, and
- Interface bases

A ***sensation basis*** is a mechanism including databases, knowledge bases and that receives data from interfaces, fuse the data and represent them in relation with acquired (learned) or fixed patterns. Moreover, a sensation base sends control information to the sensor interfaces to control them, and to higher level bases (knowledge bases, relation bases etc.). A sensation basis differs from a knowledge basis by the following characteristics:

i) it acts as an interface between post-sensors processing elements and other higher level bases;
ii) it sends information upwards, to various knowledge bases, inference systems, relational bases, downwards, to sensors, and possibly horizontally, to other sensation bases;
iii) it is controlled by specialized, upper level knowledge bases, and by the relation bases;
iv) it may include one or several knowledge bases, processing or memory-type neural networks, fuzzy systems etc.

A *representation basis* is a mechanism including databases, knowledge bases, and various types of inference engines, and that receives data from the sensation bases, knowledge bases and the relational bases and processes it to generate representations of the environments and the knowledge. The results are send horizontally, at bases of the same hierarchical level (relation bases, knowledge bases).

The *knowledge bases* are specialized for different types of knowledge. They are situated on a medium hierarchical level, representing tools for higher level bases such as relation bases, representation bases, and interface bases. The knowledge bases use one or several types of inference mechanisms, related to

the data they process. A detailed presentation of various such bases will be given elsewhere.

Data structuring has also to be appropriately developed, with various degrees of complexity. We may organize data as a *drawer chest*: a structure of differences, i.e., no order is assumed between the drawers; however, objects in a drawer are similar, while dissimilar to objects in other drawers. Another way to organize data is similar to a *topical library*: an associative memory, each object having connections with the close objects (same or similar topics), yet preserving some vague difference; moreover, relationships exist between various places in the library, reflected either in the multi-criterion subject index of the library, or in the *books* themselves. Note that a *library* is similar to a *drawer chest* but there is no boundary between the drawers, moreover it is somewhat similar to a two-dimensional, fuzzy (overlapping) *grid*. The structure may look like a *book*; a book is a structured object with knowledge passing from one *chapter* to another, yet some parts being loosely related to the others. The vague, loosely connections in the books and in the library are essential, allowing "creative": new correlations to be established; new correlations are embedded in the library arrangement, yet their "strength" can vary in time or from the perspective of the knowledgeable subject reading the book. The search in a book is performed according to some index of the book and library (contents, index of topics), but based on a *key* possessed by the user. Consequently, two users will have two different index-keys, in general. Here, the "user" is the information or production query itself, and the key is related to the information content. From the classic databases or knowledge bases viewpoint, the library is an ambiguous structure.

We see as one of the major limits of the current AI methods the use of fixed relations (in the classic sense) between various classes of data. This situation is quite different to the inherent "subject-related relations" in real life, and should be corrected in the first place. The use of "uncertainty" to eliminate this drawback is still a blind method of taking into account the flexibility of the relationships and production methods. The productions should be related to the producing information. All production systems and declarative structures – semantic networks, production rules, etc. should be endowed with specific coding/decoding mechanisms, able to adapt the productions and the declarative content to the "user" and to the data. In addition, the use of various logics in the productions and the selection-by-the-user of the logic may improve the power (thus, "intelligence") of the production-based methods.

There is evidence that not in all civilizations (mainly in those "primitive"), or

under all circumstances, the order relation is considered transitive by humans. Moreover, when various criteria are considered, the ordering is not ensured. Therefore, we have to consider the case: $a \prec b \ \& \ b \prec c \not\Rightarrow a \prec c$. An order that is not transitive may be used, for example, in indexing databases. In such a case, the order does not insure the identification of exactly one piece of information, but of a cluster of information. Moreover, it leaves the possibility to generate new searches every time, in a manner possibly miming the way of finding the information. The lexicographic order brakes down to several possibilities and the structure of the data become more intricate. Consequently, a database or a knowledge base using indexing form this type of "lexicographic" order will go through more states. Supplementary mechanism may help "choosing" the next "index" among the possible ones. Such mechanisms may be hidden Markov chains, genetic algorithms, or any other theoretically appropriate tool.

In case of real *human and animal* world, there is no production rule always viable. For instance, the rule: "*If offended, offend/hurt/punish.*" has no relevance in general, moreover, it has no probabilistic (statistical) meaning. Regarding the "general meaning": its truth depends on the individual and context. Some individuals, related to mood, age, religion etc., will behave according to the contrary rule: "*If offended, forgive/forget/do not react/offer help.*" Regarding the statistics: once you performed a statistic on an individual, it is already obsolete, because of dynamic changes of the "probabilities", that is, because of the evolution (age, culture, experience, mood related to external factors). The process is neither "stationary", nor "ergodic." Under such circumstances, statistics is irrelevant. We may conclude that the way from knowledge representation (structuring) to knowledge non-representation (destructuring) should be also considered in some cases.

Regarding the purely technical means to increase intelligence and communication, including by parallel languages, the literature is increasingly abundant, but frequently misleading because of the poor understanding of the role plaid by these languages.

6.9 Conclusions and Future Perspectives

In this chapter, an overview of some general problems of artificial intelligence, of some specific aspects of man-machine interaction and of computer-robotic hybrid type of systems able to perform this interaction were presented. The focus was on the missing background to create such machines, and a "human-

oriented" set of capabilities in the machine, leading to a "human-oriented" (virtual) reality. We concluded that various new concepts and techniques have to be developed to fulfill the goal, and we tried to establish the framework of such a development. We emphasized the key part plaid by the *relationship representation*, on one side, and by *parallel* (*secondary, annex, and subliminal*) languages in higher level interaction, on the other side. We have emphasized some new means (relational bases, sensorial bases etc.) needed to increase the capability of intelligent systems. Instead of more conclusions, we present a program for the future.

A program for the near future (less than 10 years.)

- Machines detect several secondary (parallel) languages.
- Machines generate secondary language messages and interact at these levels.
- Machines are able to interpret texts (as an actor, at the level of voice messages) when producing voice messages.
- Texts are interpreted at the level of secondary languages.
- Machines react spontaneously and interactively with people, for instance when someone they know comes in the room: i) recognition of users' bio-psychic states (close to human level ability); ii) recognition of users' personalities and of every user as a bio-psychic personality (not merely as a physical entity); iii) recognition of specific attitudes of the partner; iv) adapt to the partner and to the his/her specificity, and to his goals; v) are able to recognize groups of partners, their relationship, and to establish a group relationship (with that group), i.e, has ability to report itself to the group; vi) adapt to the specificity of group communication.
- A program for the far future (10-30 years):
- Development of *wholarchic groups*: groups of equal, spread responsibility. Virtual teams, virtual society. Machine as a socialization developer.
- Machines *emulate* personality (20 years): i) reflexivness: have a set of reflexes; ii) emotiveness: develop patterns of sensations; iii) behaviorality: have a specific (own) behavior; iv) variability (human like variability); v) human like variability due to time (hour, season, cyclic time).
- Machine is able to *anticipate* humans' wills and goals. (30 years)
- Machines *improve their "personality"* and accumulate more abilities to talk and interpret texts and speech, by reading and talking. (20 to 30 years)
- Computer morality: machines protect their own and others' existence. (10 to 30 years)

- The *embedded-consciousness*: (30 to 50 years) i) able to identify itself; ii) able to manifest itself in a different way that other machines; iii) able to create and structure its own "life experience"; and possibly: iv) able to develop their own goals and strategies.

Determining what computers could do in the next 10 to 30 years may help determining the directions of research for the next period. There are two main directions of research, aiming to: i) acquisition of positive capabilities, and ii) acquisition of limits.

A. Acquisition of positive capabilities

- Understand natural languages (at least one linguistic, and possibly several non-linguistic), and read manuscripts, understanding the meaning.
- Understand people intentions, their moods.
- Socialize in human groups and interact (give family advice, help growing up children, help stopping quarrels.)
- Socialize in teams with other machines and organize themselves on groups.
- Have individual evolution.
- Interpret: play and compose music, play theater (human-like actors, with gesture, mimics, and intonation), create inventions, make discoveries.
- Manage large systems, for example take care of everyone's health in a community.

B. Acquisition of limits

In what follows, we use the more general term of "limits" to denote various types of higher-level constraints, as related to legal and moral constraints. These limits are *learned* by humans and are finally reflected in their behavior and personality.

The topic is related to "laws for computer behavior". In [Heer and Lum], Isaac Assimov is quoted with his generally known "laws of robotic behavior", presented in 1940:

1. "*a robot may not injure a human being, or through inaction allow a human being to come to harm;*"
2. "*a robot must obey the orders given it by human beings except should such orders conflict with the first law;*"
3. "*a robot must protect its own existence as long as such protection does not conflict with the other two laws*".

The main idea is that computers should follow some "Ten Commands for Computers", similar to the ones for humans. Heer and Lum say: "*In the distant future, designers would do well to heed Assimov's laws; but they bear little*

relation to present technical realities." Actually, we already have to consider such laws in the near future. Moreover, we should consider rudiments of computer morality. By computer morality we understand laws to be obeyed by computers for their behavior is considerate, sociable, not harmful. Such rules and behavior patterns should be both burned into the silicon chips and imbedded, as dynamic rules, learning rules, in the software.

The acquisition of limits refers to regulatory characteristics. These are "negative" characteristics, showing what a machine should not do, as an individual machine, or in teams of machines. The following should be either forbidden or dealt with imposing strong constraints:

- make final decisions;
- act beyond the human decision horizon (a concept meaning: capability of humans to foresee and decide for the present or the future);
- organize themselves in larger groups, characterized by a certain level of hierarchy higher than allowed.

However, we find this approach – in the line opened by Assimov – a limited one. Indeed, better than forbidding actions based on a set of rules is to think in terms of "computer morality." Abilities and means to learn the moral behavior should be imbedded.

Acknowledgements

The author thanks Prof. Tadashi Kitamura and Prof. Takeshi Yamakawa for the invitation to contribute this chapter. The author acknowledges the partial support of three research Grants from Fonds National Suisse (FNS, National Swiss Funds) during the past four years.

References

[1] Agripa, C., De incertitudine et vanitate scientiarum et artium atque excellencia verbi Dei declamatio, Anvers, (1530) and Declamation sur l'incertitude, vanite et abus des sciences et des arts, Paris, (1582).

[2] Bacon, F., Cogitata et visa de interpretatione naturae sive de inventione rerum et operum, De interpretatione naturae sententiae. In: Despre Intelepciunea Anticilor.

Editura Stiintifica si Enciclopedica, Bucharest, (1976).

[3] Biberi, I., Arta de a scrie si de a vorbi în public (The art of writing and of public speaking). Ed. Enciclopedica Româna, Bucharest, p.17, (1992).

[4] Draganescu, M., Eseuri (Essays). Ed. Academiei Romane. Bucharest, (1993).

[5] Eco, U., Treatise of General Semiotics. (Romanian translation: Tratat de semiotica). Editura Stiintifica si Enciclopedica, Bucharest, (1982).

[6] Heer, E., and Lum, H., Toward Intelligent Robot Systems in Aerospace. In: Heer, E. and Lum, H. (Eds.), Machine Intelligence and Autonomy for Aerospace Systems. Progress in Astronautics and Aeronautics. Vol. 115. American Inst. of Aeronautics and Astronautics, Inc., Washington, DC, (1988).

[7] Ieremia, M., private communication. This assertion was taught to students by the renowned psychiatrist, the late Professor Petre Brânzei. I am not sure that he originated this pseudo-paradox, (1981).

[8] Malrieu, D., Les apports d'une etude differentielle de la demande bibliographique pour la modelisation des utilisateurs. Intellectica, no.15, pp.187-214, (1992/1993).

[9] Minsky, M., Future Models for mind-machines. Proc. Int. Symposium on Natural Language Understanding and AI (ISKIT'92), Iizuka, Japan, (July 12-15), pp.1-6, (1992).

[10]Suppes, P., Metafizica Probabilista. Humanitas Publishing House, Bucharest, (1990). (Translation of Probabilistic Metaphysics. Basil Blackwell Publisher Ltd., Oxford, England, 1984).

[11] Spencer, H., Principes de Psychologie. Tome second (Volume 2). Nouvelle Edition. Editeur (Publisher) Felix Alcan. Paris, (1898).

[12] Teodorescu, H.N., Artificial sensiology, artificial consciousness and the sensitive computer. Proc. International Conference on Intelligent Technologies in Human-Related Sciences, ITHURS'96. July 5-7, Leon, Spain. Vol.1, pp.XXXV-IL, (1996).

[13] Teodorescu, H.N., Computer semiotics: understanding meanings and parallel languages. Proceedings, IIZUKA'98 Int. Conference on Soft Computing, Iizuka, Japan, World Scientific, (1998).

[14] Winston, P.H., Artificial Intelligence. Addison-Wesley Publ. Co., Inc., (1977). (Romanian translation: Technical Editions, Bucharest, 1981).

This page is intentionally left blank

Chapter 7

Time Emerges from Incomplete Clock, Based on Internal Measurement

Yukio-Pegio Gunji, Hideki Higashi, and Yasuhiro Takachi
Kobe University

Abstract

There is no grounding in which a particular distinction between reality and representation or between the inside and the outside (surroundings) can be objectively verified. This perspective, however, yields us the study of the origin of the grounding or the origin of a particular universal structure by which the grounding seems to be verified. We especially focus on the origin of the representation of the time consisting of the present, past and future. Starting from the analysis of skepticism such as Chinese room and the diagonal argument, we propose a particular model of the interaction with open surroundings in which distinction of the inside and outside must be incomplete. This kind of interaction is called the internal measurement. The aspect of incomplete distinction is expressed as the distinction realized by invalidation of distinction. We propose two models. In the first one, a distinction is defined by a set of state and a set of rules, and the invalidation of the distinction is expressed as a self-similar set in a Lorentz plot. This self-similar set is used as a return map and then generates a new state. This is, therefore, regarded as a primitive model of incomplete clock. We show that the biological evolutionary pattern called punctuated equilibrium can be explained by this model with respect to an exponent of power-law distribution. In the second model, a distinction is defined by a directed graph and its free category. The directed graph at the tth step is transformed into a graph at the (t+1)th step by invalidating the free category derived from a graph at the tth step. This is also regarded as a model of incomplete clock. This incomplete clock can generate particular universal structures, limit and co-limit which can be interpreted as the time consisting of the present, past and future.

Keywords : internal measurement, recursive clock, self-similarity, Turing test

7.1 Introduction

The most fundamental problem is the origin of universal structure that seems to yield a particular grounding or encoding [1-3], or the origin of representation [4]. A universal structure is illustrated by a particular structure such as a language, mathematics and consciousness. A language is believed to be the en-

coding system connecting a word with its referent. In a broad sense, one generally talks about the code between representation and real entity, and believes that this code is based on consciousness. This encoding system is believed to be the grounding of a universe, however it is often rigorously criticized. Especially in philosophy, Wittgenstein criticized the theory of description[5], the correspondence between a symbol and its use, and between act and the foundation of act. Kripke took after Wittgenstein's idea and criticized the coding system between a word and its description by introducing a skeptic [6,7]. Notwithstanding these critiques, science and technology often accept encodingism [4], however some people re-found the fallacy resulting from encodingism (e.g., frame problem).

We also consider that the theory of description or encodingism is connected with the fallacy of science and technology regarding evolutionary process, cognition and life. In other words, because we regard the universal structure of encoding not as an ontological existence but as a particular local and temporal structure, we can think about its origin and the notion of emergence[8]. A language does not exist as an ontologically real entity, however, it is locally and temporally believed that a language really exists. Does this perspective mean a simple inversion of cause-effect relation, from the explanation based on the coding to talk the organization of a system, community or society, to the explanation based on the convention of a community? Even Wittgenstein and Kripke have been sometimes misunderstood as the men proposing such an inversion, and the notion of a language game is sometimes regarded as a particular convention of a community. If so, this proposal contains the in-principle fallacy. Once the skepticism or the argument on symbol grounding problem in the form of infinite regression is objectively verified, a man proposing this argument itself is destined to be independently separated from an interacting community, society or universe, whereas skepticism itself can point out the inseparability between them. In fact, Wittgenstein and Kripke pay much attention to skepticism of skepticism, which yields not the inversion but the essential development from the grounding of coding.

In science, there are some conflicts similar to the conflict between the notion of language game and the theory of description, when the evolution, development or organization is discussed. In physics, Rossler and Finkelstein propose endo-physics in which the open surroundings or context are included in formal description of local interaction, whereas the previous physics in which an observer is independently separated from an object is called exo-physics [9,10], and is taken after by Primas [11]. Matsuno independently proposes the idea of

internal measurement, in which an observer participates in an object because measurement process itself cannot be separated from an object [12]. The internal measurement evokes the process of poiesis expressed as recursive alternation of cause-effect relation or of producing and products. It, therefore, implies the intensive interaction between inside and outside, and denies complete separation between them. In cognitive science, Bickhard[4] proposes interactivism against encodingism, in paying attention to the difference from naïve interactivism proposed by Brooks [13]. Similar concepts are proposed in the physics of complex systems by Tsuda, Ikegami and Kaneko, and the notion of interaction, itself, is enhanced in thinking about evolutionary process, development and self-organization [14-16].

This trend might be entitled by "from outside to inside", where the inside means not the internally closed and individual perspective but the interactive perspective with an participating observer. Only the inversion via skepticism or critiques to the encodingism or realism, however, just accelerates the conflict between internal and external perspective. For example, if one simulates the mimic of a particular structure that can be society by means of local interactions among agents in a computer, there is no real entity of society in agents. Then, a researcher who simulated this might say that the notion of society is just an illusion, which can accelerate the conflict against those who believe real entity of a society, however this researcher enhances the significance of interaction against the notion of inside vs outside. We can verify the skepticism to a particular universal structure, whereas we can say "a particular structure". It leads us to the notion of the internal process without any foundation or grounding emergently generating the notion of coding as a particular universal structure. This is possible because any argument and structure is originated through interactions but without foundation. This perspective can invalidate the conflict between internal and external perspective.

In the present paper, we argue the origin of universal structure that can make the correspondence between reality and representation or encodingism possible. Especially we focus on the origin of time, because many researchers believe that time is much more fundamental rather than clock, and that time consisting of the present, past and future exists a priori. However, the notion of such a time is just a universal structure and is originated from use of a clock (i.e., After use of a clock, the representational content of a clock is invented as the notion of time). In focusing on the formal expression of the internal measurement or interaction without grounding, we need rigorous estimation of the status of skepticism, infinite regression, or an argument proving a paradox, and in

strictly speaking, the status of skepticism of skepticism. We first discuss the status of skepticism by illustrating the well-known Chinese room, and sketch a formal model of the distinction by invalidating distinction. Second, we propose the model of recursive clock based on the idea of making present by invalidating the distinction of present and future, by illustrating biological evolutionary pattern called punctuated equilibrium. Finally we extend this clock model by introducing the relationship between a graph and its free category and argue the origin of time.

7.2 Distinction by Invalidating Distinction

7.2.1 *Chinese Room Problem*

The problem regarding the origin of universal structure (e.g., time consisting of past, present and future; external observer; the notion of society) is illustrated in the problem of the origin of consciousness. It can be replaced with a particular question; can a computer have consciousness? As known very well, Turing proposes, which is called, a Turing test; given a computer behind the curtain, a subject asks some questions to it, and a computer replies, which is iterated; if a subject cannot consequently determine whether one behind the curtain is a man or a computer, it can be regarded as something with the intelligence as same as men's. This is a Turing test.

A philosopher, Searle criticized a Turing test and proposes a Chinese room [17]. A computer behind the curtain is replaced with a particular room in which a number of cards with Chinese characters, an American who cannot understand Chinese, and a particular manual. The American communicates with a Chinese out of a room, only by cards, thanks for the manual by which an American can know how to choose some cards and arrange them in order to reply to a question in the form of cards from a Chinese outside. Consequently a Chinese out of a room can believe that a man inside can understand Chinese, while the man cannot understand Chinese at al. This yields an irony to a Turing test.

Chinese room sounds like a critique to behaviorism. If so, Searle accepts the separation between machinery operation (or behaviors) and its foundation (or an entity making machinery operation possible), and he criticizes Turing who pays attention only to behaviors. Turing is not a simple behaviorist. Actually, Searle proceeds toward the approach similar with the interactivism after proposing a Chinese room, which shows Chinese room is not a critique to behav-

iorism.

We can understand Searle's critique as following; (1) Appearance of making decision by which a man regards a particular computer as a thing with intelligence is local, individual and empirical, while it cannot be separated from social context, others, or a network of speaks and acts. Local decision, even if it is saying that a computer has intelligence, is possible, and is essentially independent both of a-priori-definitions of men and computers. In this sense, decision in Turing test's situation seems to be reasonable because a thing behind the curtain can be proved neither a man nor a computer. Turing, however, proposes this situation as a standard to determine machines with intelligence from machines without it. The notion of standard or tests is destined to be a-priori-concept independent of context, others and language-acts networks. Because making decision in the sense of Turing test is not understood in local interactionism but in global idealism, the standard of making decision must require indicating a particular foundation or semantics. Consequently the notion of test requires complete separation between machinery operation and its foundation, or between syntax and semantics in principle. (2) Any operations cannot be divided into pure syntax (machinery operation) and pure semantics. If this separation is assumed in principle, it entails a paradox, which is illustrated by Chinese room.

(3) Searle, however, does not deny a particular situation in which it is possible to say that the computer has an intelligence. He just criticizes global idealism. Searle basically commits Austin's philosophy of speak=act, therefore he argues that if the condition under which the exchange of symbols between a computer and a man cannot be separated from act (i.e., it is as same as communication) is enhanced, then it is immediately possible for a man to locally decide that a computer has intelligence. Such a condition might be possible by adding physical interactions with the interaction between a computer and a man. Interactivism proposed by Bickhard is partially takes after Searle's approach. Bickhard starts from the critique of encodingism, and illustrates various aspects resulting from the fallacy of encodingism [4, 18]. This is the same as the critique of complete separation of syntax and semantics, which is shown as Chinese room. Bickhard illustrates that such a separation between real entity and representative does not exist but is originated in a universe of interactions, which is the origin of representation. This approach is, in a broad sense, close to the perspective of a language game proposed by Wittgenstein. In other words, Bickhard extends the idea of a language game and dilates it to a materialistic interactions. We agree with the basic concept of interactivism, however we

show another approach because we find another status of Chinese room, or skepticism.

7.2.2 *Skepticism of Skepticism*

The development from skepticism against encodingism is similar with the one from skepticism against the separation of a word and meaning, or of a symbol and a rule by which one can use a symbol. Interactivism is similar with the perspective of a language game. In the case of a language game, this development is manifested in Kripke's arguments regarding plus-quus problem [7]. We here compare Chinese room to plus-quus problem and manifest the structure of skepticism. Kripke asks you, "Which rule did you follow in calculating the addition?" In order to find the fallacy resulting from the separation between a particular symbol of addition, "+", and a rule designating how to use "+", Kripke assumes two rules called as plus and quus. Two rules are the same for all your previous experiences of calculations on "+", such as 1+3=4, while they are different from each other for un-experienced number that is, just for the sake of convenience, labeled by 51. It is defined that 51+1=52 in plus, and that 51+1=1 in quus. Because the difference is cited with respect only to the un-experienced number, 51, you cannot determine whether you followed plus or quus in addition. This is plus-quus skepticism.

It is clear that Chinese room can be compared to plus-quus problem, because two semantics for addition, such as plus and quus can be replaced with two semantics for the operational communication with a Chinese outside, such as the semantics with ability of understanding Chinese (a man understanding Chinese) and the semantics without this ability (an American mentioned before). In Chinese room problem, Searle points out that semantics can be proved both, as well as the indeterminacy of plus and quus. We can, however, find a trick in Searle's argument. Because Searle attempts to criticize the separation of syntax (machinery operation) and semantics (foundation of the operation) by the reductio ad absurdum, he has to keep this separation in his argument. He, however, breaks this separation. He first defines an American inside as an entity that only contributes to syntax or machinery operation, because a particular manual is assumed to be the foundation by which the room can communicate with a Chinese outside. But, an American is finally used as essential character contributing semantics in Searle's arguments. Actually the separation between syntax and semantics in Searle's argument is not enough to induce a paradox, because an American inside can read and understand the manual, which means

his contribution to semantics in advance.

This seems to be a trick, however, is it necessary to say a trick? We go back to the plus-quus problem to estimate whether plus-quus problem contains a similar trick. In Kripke's argument, the separation between a particular symbol and how to use a symbol must be kept, and the mixture of them must be prohibited because of the means of reductio ad absurdum. This separation is, however, broken by indicating un-experienced number as a particular number, 51. A word, "an un-experienced number" must be just a symbol in the argument of skepticism by the principle of separation between a word and its meaning, and the meaning of "an un-experienced number" must be defined independent of our every-day convention of addition. Yet, because the meaning of "an un-experienced number" is required to distinguish plus from quus, it must be defined dependent on addition. It consequently follows that "an un-experienced number" is designated by a particular number, 51. Once "an un-experienced number" is designated by a particular number, it means the mixture of a symbol and its meaning. So we conclude that quus-plus problem contains the same trick as Chinese room.

The next question arises whether the mixture of two notions belonging to different categories(e.g., a word and meaning), that is called a trick mentioned above, can be avoided in principle or not. We here estimate the formal reductio ad absurdum, such as a diagonal argument. In Cantor's diagonal argument cardinality of an infinite set A is compared with that of its power set, $P(A)$, that is isomorphic to 2^A,which means a set of maps from A to $2=\{0,1\}$ [19, 20]. Compared to plus-quus problem, a particular symbol "+" and a rule designating how to use "+" are replaced with A and 2^A, respectively. Therefore, these two notions must be distinguished. First Cantor defines a map f from A to 2^A, and then assumes that this map is surjective (i.e., for any p in 2^A, there exists b in A such that $f(b)=p$) because his aim is to prove a contradiction only from the assumption of surjective f. Yet Cantor indicates the unknown map g defined by

$$h(f(a)(a)) = g(a) \tag{1}$$

with arbitrary map h, and this leads to a fixed point,

$$h(f(b)(b)) = f(b)(b), \tag{2}$$

implying a contradiction. As well as plus-quus problem, the unknown map that cannot exist in an infinite set, 2^A, is indicated by concrete meaning such that

$h(f(b)(b))=f(b)(b)$ by using an arbitrary map, h. That is why we can doubt even the foundation of diagonal argument. It is generally possible to argue skepticism of skepticism. The final question arises whether our skepticism of skepticism means that the skepticism is impossible, or not. This question leads us to estimate the status of skepticism of skepticism [21].

7.2.3 *Distinction by Invalidating Distinction*

Even formal proof contains the mixture of two kind of notions belonging to different categories, whereas the mixture is prohibited. Although we call this mixture a trick mentioned before, this mixture is inevitable in principle. Any indications or symbol usage seems to be founded by existence of referent or meaning, which requires the separation of indication and referent. Indication or symbol usage, however, cannot be separated from context, and then one cannot help finding the mixture of symbol and its meaning. Here we can find perpetual succession consisting of temporal distinction between a symbol (machinery operation, pure syntax) and its meaning (the notion rules, pure semantics), and temporal mixture of them. In other words, given a distinction, invalidating previous distinction makes present distinction possible, which enables perpetual succession of operations, even for mathematical arguments [21,22].

We can finally doubt even skepticism and the foundation of the formal reductio ad absurdum. It does not mean that skepticism is impossible. If we say that skepticism is impossible, then we have to accept that any arguments and statements are also impossible. Notwithstanding logical impossibility, any arguments are possible. What we have to focus on is this aspect. In this sense, the form of skepticism itself can provide a particular model of distinction. Any distinction has neither grounding nor standard by which distinction is possible. Yet, if we, observers, constitute a model of distinction of natural material, we have to confront the expression of the grounding. If we completely neglect the grounding of distinction in a model of distinction (i.e., the definition of distinction is given without any grounding), it means that grounding of distinction for the very object is the same as that for an observer. The separation between an observer and an object, paradoxically, entails that they follow the same grounding. If we pay attention to the grounding and express it explicitly, we have to accept the intelligence (i.e., particular grounding) in a material, by which a particular distinction is possible. Local distinction cannot contain a particular grounding, while we cannot express local distinction whether we explicitly constitute a particular grounding or neglect it. Another expression is

required for local distinction.

Our argument on diagonal arguments can yield a particular model of local distinction, in the form of "distinction by invalidating distinction". Given two different categories corresponding to the notions of symbols and the notion of rules, invalidating this distinction can make new distinction. In this framework, we will propose two kind of models. The first one is directly constituted by our argument on diagonal argument. Given a distinction between a set of states, A, and a set of transition rules, A^A, invalidating this distinction is expressed as $A \simeq A^A$, this isomorphism is regarded as an equation, a solution of it is regarded as a transition rule, and a new state is obtained by this newly obtained transition rule. This procedure means a unit of a clock making "ongoing present". The second model consists of a directed graph, its free category and the transformation between them. Two different logical types are expressed as graph and free category, and we invalidate this distinction. This is also a model of the recursive clock making ongoing present as a directed graph. These two models are summarized as follows;

(i) Two different categories, called the present state labeled by time step, t, and the possible space in which the present state can be allowed, are defined. The former can be regarded as a result of the choice from the latter. They constitute duality, and there is an explicit well-defined distinction between them.

(ii) The operation by which the distinction mentioned in (i) is invalidated is defined. As a result of this operation, new present state labeled by $t+1$, is obtained.

It, therefore constitutes a clock measuring ongoing present. Because this clock proceeds by invalidating the difference of logical types, new present state can appear with newly developed structure or higher order. This clock does not mean simple iteration but unit time generated by recursion. In the following sections, we will argue these two models in detail.

7.3 Punctuated Equilibrium Resulting Asynchronous Clock

7.3.1 *Self-organized Criticality as a Master Clock*

In order to manifest a model of the recursive clock, we take the pattern of biological evolution. We first estimate a particular model of self-organized criticality, called Bak-Sneppen (B-S) model that explains patterns of biological evolution consisting of intermittent explosion of speciation and long stasis

following the explosion[23-25]. The idea of self-organizing criticality referring to the tendency of dynamical systems to organize themselves into a poised state far out of equilibrium [26-28], make physicists focus on these particular extinction patterns, that is called punctuated equilibrium. B-S model was proposed to explain the power-law distribution of avalanche size and the interval of first turn (i.e., lifetimes) [29-33]. More simpler or modified models were proposed whilst retaining the essence of the original model, and some analytic results were obtained, in terms of universality class [34], devil's staircase with dimension-dependent [35], and a Markovian continuous time random walk [36].

B-S model is defined as follows. Consider an ecosystem consisting of N species arranged in the one dimension (k=1,2,.., N). Each species is labeled by fitness as a real value in [0.0, 1.0]. Initially, the distribution of fitness over an ecosystem is randomly chosen. Time development is defined as follows; choose the least fitness; if the least fit species is the kth species, the fitness of three species of the kth, (k-1)th and (k+1)th are randomly re-chosen (flipped) in [0.0, 1.0]. This dynamics can lead to the genesis of critical value (~0.67) of fitness, above which species are almost stable because the least fit and its neighboring sites are flipped (Fig. 7.1). Under a stable critical value, flipped species can be successively flipped, because random flip can lead new fitness smaller than the critical value that is larger than 0.5. It, therefore, reveals the power-law distribution of lifetimes of species defined by the interval of flips (Fig. 7.2A).

The power-law distributions of avalanche size and lifetimes are influenced by the choice of least fitness. If the flip site is randomly chosen and this site and its neighboring sites are flipped, then the distribution of avalanche size and lifetimes reveal the exponential-law (Fig. 7.2B). The existence of stable critical value depends not only on the least fit species but its neighboring species. If only the least fit species is flipped at each time step, then the fitness of all species is gradually increased and is saturated at 1.0. The flip of the least fit species influences the increase of fitness, whereas the flips of neighboring sites influence the decrease of fitness because the neighboring sites can have larger fitness. This balance can generate and stabilize the critical value.

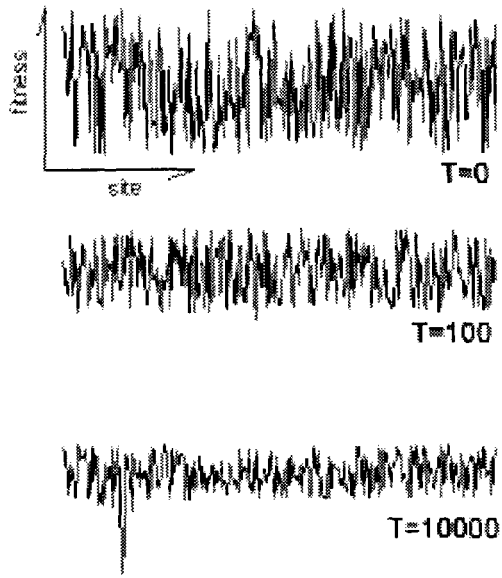


Fig. 7.1 Time Development of Bak-Sneppen Model. The symbol T represents the numbers of time step. See text.

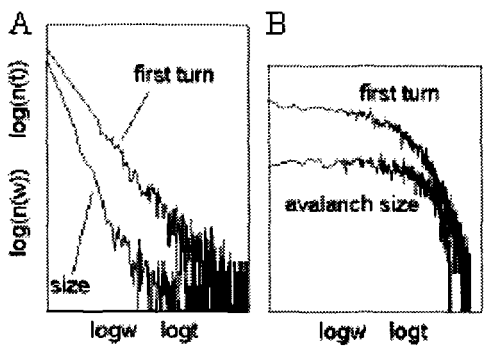


Fig. 7.2 A. Power-law Distribution of Avalanche Size and Lifetimes of B-S Model. B. Power-law Distribution of Avalanche Size and Lifetimes of Random Site Flip.

The B-S model can be regarded as a model of a unique clock. The notion of clock is divided into continuous motion and measurement of this motion. In B-S model, the continuous motion is derived from the distribution of fitness, and is measured by each local site (species). The interval between flips at each site is its local unit time, dependent on the rank of fitness in a distribution of fitness. It, therefore, reveals a unique master clock controlled by the distribution of fitness that uniquely exists in a given ecosystem, whereas each local species measures this unique motion incompletely. Why measuring incompletely? Because the flip is randomly done, and then the rank-distribution of fitness is temporally perturbed. Finally, we can say that the essence of B-S model, in generally speaking, phase transition of self-organized criticality, is one unique clock with incomplete measurement.

We claim the notion of a unique master clock, because there is no master clock in a material universe. If one accepts a unique master clock, then the notion of incomplete measurement can be expressed by deficient measurement. This is defined as a-priori distribution of error. By contrast, if one does not accept the master clock, one has to confront with a many-clocks system, in which modulations among clocks perpetually proceed. There is another weakness of B-S model claimed by Sneppen himself [37, 38], that is regarding the exponent of power-law distribution. The B-S model reveal the exponent 1.0 in a term of power-law distribution of lifetimes, while the exact value, 2.0, is observed in the fossil record [39]. The power-law distribution of avalanche size has not been obtained in the fossil record, some natural phenomena show the exponent of 2.0. The distribution of the number of branching in a lineage also reveal the power-law distribution with exponent, 2.0. One of the authors estimate the distribution of lifetimes, avalanche size, and the number of off-springs in a lineage of stallion, all of them reveal the power-law whose exponent is 2.0[40]. We finally conclude that the ubiquitous exponent, 2.0, should be test. Our problem in terms of biological evolution is "Can we explain the power-law with a particular exponent, 2.0, by the model consisting of many clocks?"

7.3.2 *Asynchronous Clocks*

Whenever we consider many-clocks system and the modulation between them, we confront with the conflict between digital and analog time. Imagine that you have to modulate your watch in accordance with my watch. In this situation, you identify the state of my watch as time that must be digital, and then you modulate your watch that is ongoing (analog). That is why you have to com-

pare digital with analog time.

The modulation between digital and analog time is contained in our model consisting of N sand-clocks in discrete time [41]. The amount of sand deposited in the bottom, and that of falling sand at kth sand-clock at tth step is represented by $c_k(t) \in [0.0, \theta)$ and $a_k(t) \in [0.0, 1.0]$ with a real number, θ. Each sand-clock proceeds interacting with its neighboring sand-clocks with radius m, by

$$c_k(t+1) = f(c_k(t) + a_k(t+1) - r) \tag{3}$$

$$a_k(t+1) = h(<a_{k-m}(t-2), a_{k-m}(t1)>,...,<a_{k+m}(t-2), a_{k+m}(t-1)>)a_k(t) \tag{4}$$

where $f(x)=x$ if $x \leq \theta$; 0.0 otherwise, r is a parameter but now is set 0.0. A sand clock flips when the amount of deposited sand exceeds θ. An operator h makes a self-similar return map [22] satisfying $A(t) \simeq U(a_j(t-2), a_j(t-1))$ with $j = k-m, k-m+1,.., k+m$. $U(x, y)$ is the neighborhood of a point $<x, y>$, and $A(t) = \{<x, y> \mid y = h(<a_{k-m}(t-2), a_{k-m}(t-1)>,...,<a_{k+m}(t-2), a_{k+m}(t-1)>)x, \ ^{\forall}x \in [0.0, 1.0]\}$. The isomorphism, $A(t) \simeq U(a_j(t-2), a_j(t-1))$, is possible to take

$$a_k(t+1) = P(y_{g(f(i)-1)}, y_{g(f(i))}) + \sum_{i=1}^{\infty} P(y_{f(i)-1}, y_{f(i)})m_i \tag{5}$$

$$m_i = \prod_{s=1}^{i-1} (x_{f(i-s)} - x_{f(i-s)-1}) \tag{6}$$

where $<x_i, y_i> \in D(t) = \{<0.0, p>, <a_{k-m}(t-2), a_{k-m}(t-1)>,...,<a_{k+m}(t-2), a_{k+m}(t-1)>, <1.0, q>\}$, and p and $q \in [0.0, 1.0]$ are chosen at random. For any i, given $<x_{f(i)}, y_{f(i)}>$, $x_{f(i)-1} < x_{f(i)}$ and given $<x_{g(f(i))}, y_{g(f(i))}>$, $y_{g(f(i))-1} < y_{g(f(i))}$. $P(e,d)$ is the random choice of e and d, and where $f(i)$ is chosen such that $m_i x_{f(i)-1} \leq a_k(t) - \sum_{s=1}^{i-1} x_{f(i-s)-1} m_{i-s} < m_i x_{f(i)}$ with $m_1 = 1.0$ and $x_{f(1)-1} \leq a_k(t) < x_{f(1)}$. The index i represents the ith contraction to generate self-similar set. Fig. 7.3 shows an example of $A(t)$ obtained by Eq-(5) and (6), that is used as a return map from $a_k(t)$ to $a_k(t+1)$ (i.e., Eq-(4)). It shows self-similarity, and the neighborhood of each point is isomorphic to a cartesian subspace.

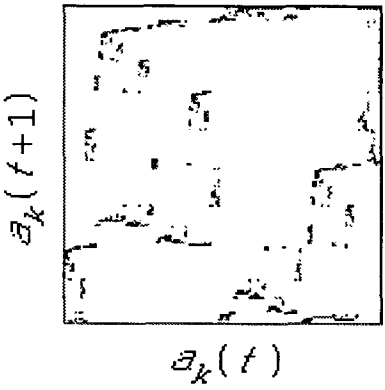


Fig. 7.3 A Typical Return Map Obtained by Eq-(5).

Taking the framework of the notion of recursive clock mentioned in 2-3, this model is sketched as follows; Two different logical types, the state and the map, are defined. The states are expressed as a set of pairs of falling sand, ($a_j(t\text{-}2)$, $a_j(t\text{-}1)$), and the map can be expressed as a subspace crossing all points of $\{(a_j(t\text{-}2), a_j(t\text{-}1))\}$. As for a local sand-clock, it has to determine a particular map crossing given all points in order to making its own new state of $a_k(t+1)$. This require a particular criterion by which a map is determined. Once we observers who make a model define a particular criterion or a potential map, it, however, implied that a local sand-clock has intelligence. But, a sand-clock has no intelligence. This consideration makes us to notice that the problem of how to determine a particular map is not a problem for a sand-clock but a problem for observers who make a model of a sand-clock. We have to consider the aspect that a sand-clock can proceed time independent of the problem of the choice of potential maps, however this problem is inevitable for observers. That is why we express this aspect by invalidating the distinction between the state and the map which can make a new distinction that makes a new state of $a_k(t+1)$ possible. Such a invalidation is expressed as self-similarity in the Lorentz plot and application of this self-similar pattern as a map to the state of $a_k(t)$.

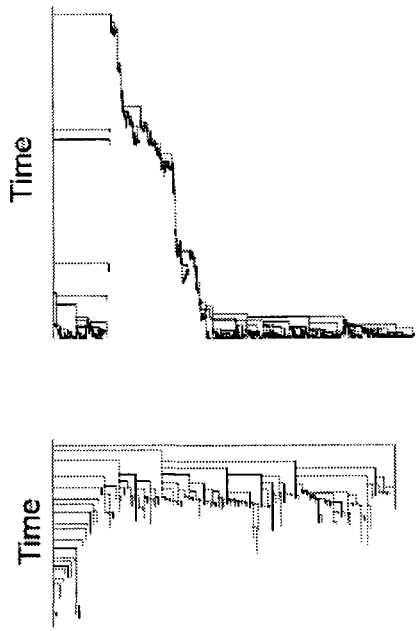


Fig. 7.4 Typical Pattern of Branching Sand-clocks. Time proceeds vertically and ranched new sand-clocks are located at the left hand of a mother sand-clock.

A sand clock model is augmented with branching dependent on the number of flips and deaths. The threshold with respect to counting the number of flips is replaced by a large value of θ of Eq-(3). In Eq-(3) threshold function f is replaced by $f(x)=x$ if $0.0<x\leq\theta$; branching if $x>\theta$; death $x\leq 0.0$, where a parameter r is a positive value and is constant as 0.5, and $\theta=3.0$. Branching is defined by accretion of sand clocks at the nearest site of k; previous sand clocks at $k+1$, $k+2$,.. are shifted to $k+2$, $k+3$,.., and then a newly generated clock is inserted at the $k+1$ site; $c_{k+1}(t+1)$ is chosen at random smaller than 0.1 and $a_{k+1}(t+1) = a_k(t+1)$. Death is defined by removal of a sand clock, whilst the site remains vacant till it can be occupied by a newly generated sand clock by branching.

Typical branching patterns of this model are shown in Fig. 7.4, where the radius of interaction is 2. They reveal a typical pattern of punctuated equilibrium in the fossil record. In our model, the lifetime between branching and death is defined as the interval between branching. The lifetime is normalized by minimal interval. The normalized lifetime is distributed as $\propto t^{2.0}$ (Fig. 7.5),

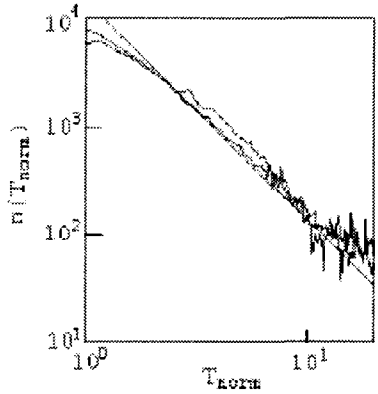


Fig. 7.5 Distribution of Normalized Lifetimes. Line represents an exponent of 2.0.

which shows a power-law. The trend of $\propto t^{2.0}$ is robust regardless of the radii of interaction. This value of the exponent is robust also regardless of parameter, r. Avalanche size distribution of extinction of sand clocks also reveals the power-law whose exponent is close to 2.0 [41]. The power law distribution of the number of offspring reveals also a particular exponent, 2.0. We finally conclude that many-clocks system can exhibit punctuated equilibrium characterized by the power-law with a particular exponent, 2.0.

This result shows that the model of a recursive clock producing an on-going present state is a powerful tool to explain the power-law derived from the intrinsic intermittent behaviors, and that the intrinsic intermittent behavior has nothing to do with the perspective of self-organized criticality whose essence is the perturbed master clock. In any formal models, there are terms contributing local or global features. We call them local and global terms, respectively. Universal scaling seems to be influenced both by local and global terms, and then many researchers attempt to constitute the model based on the balancing of them. Especially, in B-S model, there is a particular balancing that is a perturbed equilibrium between the choice of global extremal value and local flips. Although this perspective can yield us the impression that the power-law is explicitly influenced by the global feature, this impression is based upon the belief that any systemic features can be classified either into pure local or into pure global one. This is the same fallacy as the theory of description, the separation between pure syntax and pure semantics, and the encodingism, as men-

tioned in the section 2. By contrast, if we pay attention to the aspect of local interaction that is not closed in terms of semantics, then we can explain the universal structure only by local interaction.

Because the local interaction in our model is temporally changed in the form of self-similar set, it can contain infinite possible maps. This is just an expression of local distinction without grounding or foundation. It also contains the primitive idea of recursive clock counting the present state by invalidating the distinction between the previous present states and the future states expressed as the notion of maps. By extending this idea, we will argue the origin of time in the perspective based on the clock.

7.4 Origin of Time

7.4.1 *Directed Graph and Free Category*

The relationship between a grammar and a language in a formal language contains two different logical types mentioned in 2-3, (i) and (ii). In this section we use this relationship and constitute a model generating a particular universal structure, that can be compared to the structure of time consisting of present, past and future. We call grammar and language, actual state and possible space, respectively. Imagine a grammar generating strings by the concatenation of a letter, f. Because any grammars can be expressed as a graph consisting of directed edges and nodes, this grammar is expressed by a graph consisting of one node * and a directed edge f from * to *. By contrast, a language is defined as a set of all possible strings, and this case is represented by $\{1, f, ff, fff, ...\}$, where 1 is a null letter that makes concatenation null with $1f=f1=f$. When concatenation and a null letter can be replaced with composition of edges and identity, a language can be expressed as a free category derived from a given directed graph[42]. A category consists of a collection of objects; A, B, .. and arrows such as $f:A\rightarrow B$, defined between objects, where composition of arrows can be defined (e.g., given f from $A\rightarrow B$ and $g:B\rightarrow C$, $gf:A\rightarrow C$ is also an arrow) with associative law $(hgf=h(gf)=(hg)f)$ and an identity arrow 1_X from X to X can be defined for each object X such that given any arrow f from any object Y to X and g from X to Z, $1_X f=f$ and $g1_X=g$. That is why a language is a category (Fig. 7.6).

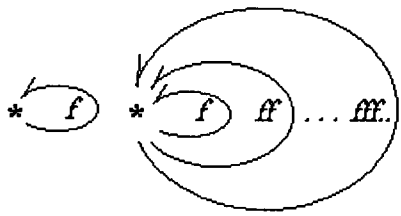


Fig. 7.6 A Grammar and Its Language.

In denoting all categories by **Cat**, a language can be expressed in **Cat**. However, generally a grammar as a directed graph is not a category. In order to express any graph as a category, a specific category **Grph** is introduced. **Grph** is a category whose object is defined by a pair of two sets D_0, D_1 and two functions *dom* and *cod*, where D_0 is a set of edges of a graph and D_1 is a set of nodes of a graph, and *dom* and *cod* are from D_0 toD_1. A function *dom* maps an edge to its source node and *cod* maps an edge to its target node. For example, given a directed edge *f*: from *A* to *B*, $dom(f)=A$ and $cod(f)=B$. An arrow of **Grph** is a specific operator (natural transformation) *t* from an object $<D_0, D_1, dom, cod>$ to an object $<E_0, E_1, dom, cod>$ such that $domt(1)=t(0)dom$ and $codt(1)=t(0)dom$, with $t(0)$: $D_0 \rightarrow E_0$ and $t(1)$: $D_1 \rightarrow E_1$. An object of **Grph** clearly has a one-to-one correspondence to a particular graph, and an arrow of **Grph** is a graph morphism. If a directed graph is expressed in **Grph**, we can describe an operator from a grammar to a language by an operator from a category to a different category; that is a functor.

A functor *F* from a category ***C*** to a category ***D*** is defined allowing conservation of identity and composition. That is; an object *C* in ***C*** is mapped to $F(C)$ in ***D***, and an arrow *f* from *C* to *C'* is mapped to $F(f)$ in ***D*** allowing $F(gf)=F(g)F(f)$ and $F(1_C)=1_{F(C)}$. Two functors from **Grph** to **Cat** and from **Cat** to **Grph** are denoted by *F* and *U* respectively. A functor *U* that forgets any categorical structure such as identity and compositions is called a forgetful category, and maps a category to an underlying graph. Given a category *C*, *UC* is an object of **Grph** corresponds to an underlying graph forgetting a categorical structure of *C*. A functor *F* is an operator mapping a graph *G* to a category *FG* that is called a free category.

It is known that exists a one-to-one correspondence between particular graph morphisms in **Grph** and functors in **Cat**. In denoting a set of all graph morphism from G to G' in **Grph** by **Grph**(G, G') and a set of all functors from C to C' in **Cat** by **Cat**(C, C'), this correspondence is expressed as

$$\mathbf{Cat}(FG, C) \simeq \mathbf{Grph}(G, UC). \tag{7}$$

F and U bridge a part (grammar) and whole (language) satisfying the construction of (7). It, therefore, reveals the formal duality between graphs and categories.

The isomorphism (7) is an essential relationship between a graph and its corresponding free category, and it also means the relationship between a grammar and a language generated by the grammar. Note that all possible forms originated from a grammar can be prescribed in the form of a language. By contrast, a grammar is regarded as an actually realized form of possible forms, and an actual realized form is one of many examples or a model.

We refer to the significance of the identity arrow. Clearly compositions represent all possible transitions derived from a directed graph, while the identity seems to be independent of the notion of possible space. However possible space is expressed as a set of all possible compositions, it does not contain "no passing". In other words, the notion of all possible compositions contains just the notion of passing an edge, and so no passing is located outside of this particular possible space. It, therefore, means incomplete possible space. In order to compensate this, the outside can be in-formed in a possible space, and then particular arrows representing no passing are defined, that are identities. Identity represents "doing nothing", whereas other arrows represent "doing something". The existence of identity replaces the negative statement of "do not do" by that of "do nothing". The notion of possible space can thus be completed by the existence of identity arrows.

Next we use the duality of graphs and free category, and define a particular graph morphism by invalidating the distinction between graphs and free category.

7.4.2 *Invalidating Possible Space*

We define the clock that proceeds by invalidating the distinction between a graph and its free category. The unit time of this clock can be expressed as the two successive procedures; the first one is operating a functor F to a given

graph at the tth step, denoted by $G(t)$. The second one is invalidating the result of the operation of F, and this operation is denoted by INV from **Cat** to **Grph**. As a result the free category obtained from the first procedure is replaced by a particular graph denoted by $G(t+1)$ such that

$$G(t+1) = INV(FG(t)), \tag{8}$$

Because the operation of F is adding compositions and identity arrows to a graph $G(t)$, the procedure of invalidating $FG(t)$ is invalidating the identities and compositions or the universality in terms of the existence of identities and compositions. In a diagram of $FG(t)$, one identity and one composition arrow is randomly chosen, and they are invalidated as follows.

First we define the invalidation of an identity arrow. In a free category $FG(t)$, every object has an identity that is independent of its surrounding objects connected by arrows. In other words, one cannot express the notion of context in terms of the notion of identity in a category theory. One can, however, find the notion of context for an object. Given a category consisting just of an arrow $f:A\rightarrow B$, an object A is connected with an object B by an arrow f. The one of these two objects yields the context of the other, or the arrow f yields the context to an object A or B. This notion of context can produce the locality of identity. If independent of the context, identity for an object B can be defined independently of an arrow f. That is why an identity of B denoted by id_B is defined to satisfy $h= \mathrm{id}_B\ h$ with any arrow $h:X\rightarrow B$ and any object B. The local identity Lid_B is, instead, defined such that $f=\mathrm{Lid}_B f$ only with an arrow $f:A\rightarrow B$. For example, given $f:\{a, b, c\}\rightarrow\{0, 1, 2\}$ with $f(a)=0$ and $f(b)=f(c)=1$, one of candidates of Lid_B is expressed as $\mathrm{Lid}_B(0)=0$ and $\mathrm{Lid}_B(1)=\mathrm{Lid}_B(2)=1$. It is shown as Fig. 7.7. Clearly, this case satisfies $f=\mathrm{Lid}_B f$, while Lid_B is not a "real" identity.

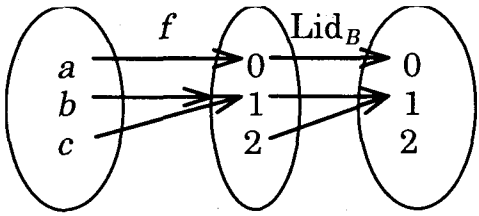


Fig. 7.7 An Example of Local Identity.

In this manner, a local identity can be defined by various forms. We expand this idea of local identity as follows. Given $f:\{a, b, c\}\rightarrow\{0, 1, 2\}$ with $f(a)=f(b)=f(c)=0$, the operation of f is the same as the operation of $f':\{a, b, c\}\rightarrow\{0, 1\}$ with $f(a)=f(b)=f(c)=0$. It can allow us the replacement of f with f', and also can allow that $f'=\mathrm{Lid}_B f'$ with Lid_B: $\{0, 1\}\rightarrow\{0, 1, 2\}$ with $\mathrm{Lid}_B(0)=0$ and $\mathrm{Lid}_B(1)=2$. This case allows a monic arrow of Lid_B. By contrast, given the same map, f, if f is replaced by f'': $\{a, b, c\}\rightarrow\{0, 1, 2, 3\}$ with $f(a)=f(b)=f(c)=0$, it can allow Lid_B: $\{0, 1, 2, 3\}\rightarrow\{0, 1, 2\}$ with $\mathrm{Lid}_B(0)=0$, $\mathrm{Lid}_B(1)=1$ and $\mathrm{Lid}_B(2)=\mathrm{Lid}_B(3)=2$, where Lid_B is an epic arrow (in Sets, it is a surjective map). It can be summarized as following. Given $f:A\rightarrow B$, if f is replaced with $f':A\rightarrow C$ with $C\supseteq B$ and $f(x)=f'(x)$ for any x in A, $\mathrm{Lid}_B:C\rightarrow B$ can be epic and can commute

$$\begin{array}{ccc} A & \xrightarrow{\quad f \quad} & B \\ \| & & \uparrow \\ \| & & \Big| \ \mathrm{Lid}_B \\ \| & & \Big| \\ A & \xrightarrow{\quad f' \quad} & C \end{array} \tag{9}$$

and if f is replaced with $f'':A\rightarrow C$ with $B\supseteq C$ and $f(x)=f'(x)$ for any x in A, $\mathrm{Lid}_B:C\rightarrow B$ can be monic.

This kind of local identity is also defined for an object A given $f:A\rightarrow B$. If f is replaced with $f':D\rightarrow B$ with $D\supseteq A$ and $f(x)=f'(x)$ for any x in A, $\mathrm{Lid}_A:A\rightarrow D$ can be monic and can commute

$$\begin{array}{ccc} A & \longrightarrow & B \\ \| & & \uparrow \\ \| & & \Big| \ f' \\ \| & & \Big| \\ A & \longrightarrow & D \\ & \mathrm{Lid}_A & \end{array} \tag{10}$$

Similarly, if f is replaced with $f':D\rightarrow B$ with $A\supseteq D$ and $f(x)=f'(x)$ for any x in D, $\mathrm{Lid}_A:A\rightarrow D$ can be epic. We define the invalidation of identity by introducing a partial order with respect to objects, such as $D\supseteq A$ or $A\supseteq D$. It, therefore, allows the introduction of D such that $A\cong D$.

The invalidation of an identity is completed after removing redundancy of commutative diagram (9) or (10). The removal of redundancy is defined as follows; given the commutative diagram of $f'\mathrm{Lid}=f$, either f' or $f'\mathrm{Lid}$ is retained

and other arrows are removed. Because Lid is regarded as an identity or null arrow, the case in which only f' is retained is possible. After Lid is expressed in a graph of $G(t+1)$, Lid is distinguished from an identity and is regarded as a normal arrow.

Second, we argue the locality of composition. In a category theory, its axiom allows a composition of arrows to be an arrow. This is a universal rule. In order to define locality of composition, it is allowed that some compositions are possible while in others the notion of composition is lost in its own right. It is defined by; given two arrows, $f:A\rightarrow B$ and $g:B\rightarrow C$, a composition gf: $A\rightarrow C$ is obtained, and after that we remove an object B and arrows f and g. It, therefore, leads to the loss of status of composition from gf. Because we finally propose a dynamical model of graph morphism via free category, we need to prevent the explosion of the number of nodes in a directed graph at each time step.

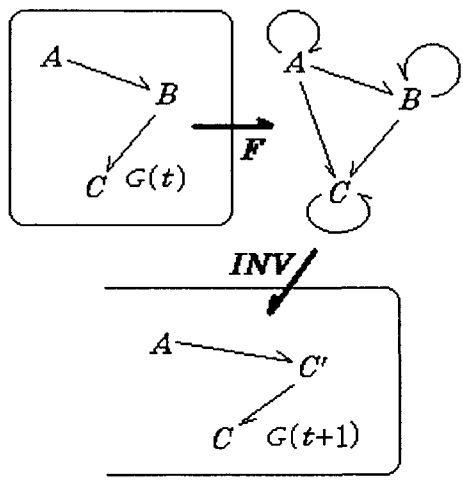


Fig. 7.8 Unit-time of Recursive Clock, Progressed by Invalidating Identities and Compositios.

Due to the invalidation of both identities and compositions, a directed graph is transformed from the tth to the $(t+1)$th, as shown in Fig. 7.8. It is clear to see that the invalidation of an identity can produce redundancy of arrows, whereas

the operation of removal of redundancy and the invalidation of composition can make the redundant component of a local identity be explicitly revealed. For example, imagine the case in which an arrow $f:\{a, b, c\}\rightarrow\{0, 1, 2\}$ with $f(a)=f(b)=f(c)=0$ is contained in a directed graph $G(t)$. This arrow cannot yield a return value of 1 and 2, while they are defined in the co-domain of f. The invalidation of an identity of $\{0, 1, 2\}$ can make the existence of Lid with Lid(0)=0, Lid(1)=2 and Lid(2)=1, possible in $G(t+1)$. The redundancy carried by Lid cannot be explicitly revealed in $G(t+1)$, however $G(t+1)$ contains the succession of arrows as Lid and f. Yet, if $f':\{a, b, c, d\}\rightarrow\{0, 1, 2\}$ with $f'(a)=f'(b)=f'(c)=0$ and $f'(d)=1$ appears resulting from the invalidation of an identity of $\{a, b, c\}$, the redundancy hidden in Lid can be exposed by $\mathrm{Lid}(f'(d))=\mathrm{Lid}(1)=2$ in $G(t+2)$. As a result, it is possible to make the case that the generated redundancy is kept hidden for a while and is suddenly exposed, which looks like a drastic change in terms of behaviors.

7.4.3 *Origin of Universality and Time*

Our model can exhibit the emergence of a particular structure carrying universality that is a limit or a co-limit, by invalidating identity and composition. The basic process is sketched as follows. If a sub-graph such as $A\rightarrow B$ is contained in $G(t)$, and the identity id_B is invalidated by $\mathrm{Lid}_B:B\rightarrow B'$ in $G(t+1)$, it can be followed by similar invalidation of $\mathrm{id}_{B'}$ in $G(t+2)$. If this $\mathrm{id}_{B'}$ is opened to recursive definition and this can be regarded as a particular functor, F, then $\mathrm{id}_{B'}$ can be invalidated by an infinite chain of arrows such as

$$B\rightarrow B'\rightarrow B''\rightarrow..\ B^{(n)}\rightarrow B^{(n+1)}\rightarrow.. \tag{11}$$

The removal of redundancy requires an equation of $B\simeq F(B)$, and the case is possible that there exists a solution denoted by E. In this case, for any object X, there is a unique arrow from X to E, which is the co-limit.

For example, in starting from the graph as

$$A \xleftarrow{\ f\ } C \xrightarrow{\ g\ } B \tag{12}$$

consider the case that the identity id_C is invalidated by a local idenity $C'\rightarrow C$ by constructing a set C' by adding some elements with C. This leads to an infinite chain of arrows and an infinite arbitrary set as X, and requires $f'=f\mathrm{Lid}$ and

g'=gLid with Lid:$X\rightarrow C$, f': $X\rightarrow A$ and g':$X\rightarrow B$. In a category of sets, a particular object or the solution of it is a cartesian product.

Start from the following diagram as a directed graph $G(0)$ at the 1st step.

$$A \xrightarrow{f} C \xleftarrow{g} B \tag{13}$$

It is also assumed that for any $a\in A$ and $b\in B$, $f(a)=c_0\neq g(b)=c_1$, for the sake of convenience. After the operation of a free functor, we obtain a free category that is added a given directed graph only with identity arrows, id_A, id_B and id_C. Next we operate the invalidation procedure. We choose a particular identity and redefine it as a local identity. We here choose an id_C and make it a local identity as;

$$\begin{array}{ccccc} A & \xrightarrow{f} & C & \xleftarrow{g} & B \\ \| & & \uparrow & & \| \\ \| & & \mid \mathrm{Lid}_C & & \| \\ \| & & \downarrow & & \| \\ A & \xrightarrow{f'} & C' & \xleftarrow{g'} & B \end{array} \tag{14}$$

where Lid_C:$C\rightarrow C'$ is defined by

$$\mathrm{Lid}_C(c_0) = \langle c_0, 0\rangle, \tag{15a}$$

$$\mathrm{Lid}_C(c_1) = \langle c_1, 1\rangle \tag{15b}$$

We consider the case in which $\mathrm{Lid}_{C'}$:$C'\rightarrow C''$ is constructed at the following step as well as Lid_C:$C\rightarrow C'$, and check whether this chain can constitute partial order consisting of embedding and projection. For any n, we can constitute an embedding, $\mathrm{Lid}_{C(n)}$:$C(n)\rightarrow C(n+1)$, as

$$\mathrm{Lid}_{C(n)}(\langle c_j, r_1, r_2, \ldots, r_{n-1}, r_n\rangle) = \langle c_j, r_1, r_2, \ldots, r_{n-1}, r_n, r_n\rangle \tag{16}$$

where j=0, 1 and $r_k\in\{0, 1\}$ with k=1,.., n. We can also constitute the corresponding projection, $\mathrm{Lid}_{C(n)}{}^R$:$C(n+1)\rightarrow C(n)$, as

$$\begin{aligned} &\mathrm{Lid}_{C(n)}{}^R(\langle c_j, r_1, r_2, \ldots, r_n, r_{n+1}\rangle) = \langle c_j, r_1, r_2, \ldots, r_{n-1}, 1\rangle, \text{ if } r_n=r_{n+1}=1; \\ &\langle c_j, r_1, r_2, \ldots, r_{n-1}, 0\rangle, \text{ otherwise.} \end{aligned} \tag{17}$$

From this definition, we obtain

$$\begin{aligned}&\mathrm{Lid}_{C(n)}\,\mathrm{Lid}_{C(n)}{}^{R}(\langle c_j, r_1, r_2, \ldots, r_n, r_{n+1}\rangle) = \mathrm{Lid}_{C(n)}(\langle c_j, r_1, r_2, \ldots, r_{n-1}, 1\rangle),\\ &\text{if } r_n=r_{n+1}=1;\ \mathrm{Lid}_{C(n)}(\langle c_j, r_1, r_2, \ldots, r_{n-1}, 0\rangle), \text{ otherwise,}\\ &= \langle c_j, r_1, r_2, \ldots, r_{n-1}, 1, 1\rangle) = \langle c_j, r_1, r_2, \ldots, r_n, r_{n+1}\rangle, \text{ if } r_n=r_{n+1}=1;\\ &\langle c_j, r_1, r_2, \ldots, r_{n-1}, 0, 0\rangle \subseteq \langle c_j, r_1, r_2, \ldots, r_n, r_{n+1}\rangle, \text{ otherwise.}\end{aligned} \tag{18}$$

It therefore leads,

$$\mathrm{Lid}_{C(n)}\,\mathrm{Lid}_{C(n)}{}^{R} \subseteq \mathrm{id}_{C(n+1)}. \tag{19}$$

We also obtain

$$\begin{aligned}&\mathrm{Lid}_{C(n)}{}^{R}\mathrm{Lid}_{C(n)}(\langle c_j, r_1, r_2, \ldots, r_n\rangle) = \mathrm{Lid}_{C(n)}(\langle c_j, r_1, r_2, \ldots, r_n, r_n\rangle)\\ &= \langle c_j, r_1, r_2, \ldots, r_n\rangle,\end{aligned} \tag{20}$$

because; if $r_n=1$, $\mathrm{Lid}_{C(n)}(\langle c_j, r_1, r_2, \ldots, 1, 1\rangle) = \langle c_j, r_1, r_2, \ldots, 1\rangle$ and if $r_n=0$, $\mathrm{Lid}_{C(n)}(\langle c_j, r_1, r_2, \ldots, 0, 0\rangle) = \langle c_j, r_1, r_2, \ldots, 0\rangle$. It means

$$\mathrm{Lid}_{C(n)}{}^{R}\mathrm{Lid}_{C(n)} = \mathrm{id}_{C(n)}. \tag{21}$$

We can define $C(n)\subseteq C(n+1)$ by (19) and (20), and then obtain $C(n)\subseteq C(n)$, and the statement; if $C(i)\subseteq C(j)$, $C(j)\subseteq C(n)$, then $C(i)\subseteq C(n)$. Therefore $\langle\{C(n)\}, \subseteq\rangle$ constitutes a partial order, and this can be regarded as a functor. Because infinite chain consisiting of $\mathrm{Lid}_{C(n)}$:$C(n)\rightarrow C(n+1)$ has the lowest upper bound as a solution of domain equation proposed by Scott[43] in the form of infinite product. In our case, such a solution is isomorphic to co-product. Thus the co-product is originated in our model.

The origin of particular universal structures of limit and co-limit implies the origin of time consisting of present, past and future. Fig. 7.9 shows a commutative diagram of limit and co-limit. Limit is an object that is a starting domain reaching other objects. When these arrows are regarded as a stream from the certainty, this diagram means that for any certain things (objects, X) there is a unique arrow toward the limit. In this statement, one regards X as certain things operating A, that is called present, however one can find the most certain thing in the limit. It, therefore, means that the representation of the certainty for present is found as the past. The past is represented by a universal structure for any present. As well as limit, co-limit yields the representation of the present that is

certain for any future. This interpretation is reasonable because arrows are inverse, compared with the diagram of limit. Finally, the time consisting of the present, past and future is generated as the origin of limit and co-limit. Note that such a time does not ontologically exist, however, it is originated in a particular local present denoted by $G(t)$. The notion of time is originated through the recursive motion of the clock measuring present state $G(t)$ by invalidating the distinction between $G(t)$ and $FG(t)$. This is our essential point.

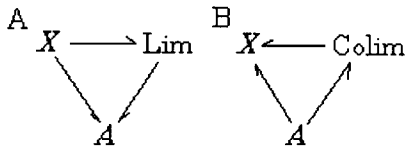


Fig. 7.9 A. A Cone Representing the Limit. B. A Cone Representing the Co-limit. The object X represents any object.

7.4.4 *Loss of Grounding*

We finally show that our model can contain the loss of grounding, and that the universal structure can be originated from losing the grounding of simulating our model. First we attempt to formalize the procedure of calculating a free category, and expand it in a more general form. If it is possible, it can be regarded as the grounding in which the simulating our model of internal measurement is possible. What is the calculation of free category? Given a graph G, it requires writing down all arrows of FG. The procedure of writing down all arrows is conducted in a set theory, and then calculation requires a specific category, **Sets.** We denote a category in which a graph G is assigned, $\boldsymbol{A}$, and a category in which a free category FG is assigned, $\boldsymbol{B}$. In this situation, when each node of a graph G is represented by A, all arrows whose domain is designated by FA is expressed as Hom(FA, -). Note that Hom(FA, -) is a Hom-set functor from $\boldsymbol{B}$ to **Sets**. Substituting B for Hom(FA, -) yields Hom(FA, B) representing a set of all arrows from FA to B. In counting each node A in G, one can write down all arrows whose domain is FA. Therefore it requires distinguishing a node from any other and assigning it, which leads to a specific functor

X:A→**Sets**. Due to this functor, each node A is assigned as a set, and is represented by XA.

Finally it requires categories, A, B and **Sets**, and two functors X: A→**Sets** and F:A→B. Because assignment of a node and writing down all arrows are separated from each other, this is necessary to take XA×Hom(FA, -), and it is a functor. Next question arises how can we complete a particular program of writing down all arrows of FA. Now the procedure of writing down all arrows is expressed as a universality of XA×Hom(FA, -) for any other operations, M:A→B, and then we take all possible operations (natural transformations) by Nat(XA×Hom(FA, -), M). By the adjunction of product and exponential, we obtain

$$\mathrm{Nat}(XA\times\mathrm{Hom}(FA, -), M) \simeq \mathrm{Nat}(\mathrm{Hom}(FA, -), M^{XA}) \tag{22}$$

where M^{XA} is a functor with M^{XA}:$\boldsymbol{B}^{XA}$→**Sets**^{XA} such that Hom(XA, B) is mapped to Hom(XA, MB). Note that B^{XA} and **Sets**^{XA} are a comma category of (XA↓$\boldsymbol{B}$) and (XA↓**Sets**), respectively. It means that an object and an arrow in (XA↓$\boldsymbol{B}$) are an arrow, v_i:XA→B_i with any object B_i in $\boldsymbol{B}$, and a cone such that v_j=$f_{ij}v_i$ with any arrow f_{ij}:B_i→B_j in a category $\boldsymbol{B}$. Therefore, M^{XA} is a functor from $\boldsymbol{B}$ to **Sets**. Because of Yoneda's lemma, Nat(Hom(C, -), F) ≃ F(C) with a functor F:$\boldsymbol{C}$→**Sets**, we obtain

$$\mathrm{Nat}(XA\times\mathrm{Hom}(FA, -), M) \simeq M^{XA}(FA). \tag{23}$$

Finally, because $M^{XA}(FA)$ = Hom(XA, MB), we obtain

$$\mathrm{Nat}(XA\times\mathrm{Hom}(FA, -), M) \simeq \mathrm{Hom}(XA, MFA). \tag{24}$$

Applying an isomorphism such that LimHom(A, B) ≃ Hom(Colim$A$, B) ≃ Hom(A, Lim$B$) to the isomophism, (24), it is expressed as

$$\mathrm{Nat}(\mathrm{Colim}(XA\times\mathrm{Hom}(FA, -)), M) \simeq \mathrm{Hom}(XA, \mathrm{Lim}MFA). \tag{25}$$

Because there is no arrow except for identity in a particular category consisting of XA×Hom(FA, B) with each A in $\boldsymbol{A}$, Colim represents a co-product of this particular category, **Sets**. Then the isomorphism (25) is replaced by

$$\mathrm{Nat}(\sum_{A\in A}(XA\times\mathrm{Hom}(FA,-)),M)\simeq\mathrm{Hom}(XA,\mathrm{Lim}MFA). \tag{26}$$

Now first we take $\mathrm{Hom}(XA, \mathrm{Lim}MFA)$, and a function y,

$$y\colon \mathrm{Hom}(XA,\mathrm{Lim}MFA)\to\mathrm{Nat}(X,MF) \tag{27}$$

is defined by $y(\varepsilon_A)=\mu\in\mathrm{Nat}(X,MF)$ for any $\varepsilon_A\in\mathrm{Hom}(XA,\mathrm{Lim}MFA)$. It is easy to see that y is a bijection, and

$$\mathrm{Hom}(XA,\mathrm{Lim}MFA)\simeq\mathrm{Nat}(X,MF). \tag{28}$$

It is also easy to see that

$$\mathrm{Nat}(\sum_{A\in A}(XA\times\mathrm{Hom}(FA,-)),M)\simeq$$

$$\mathrm{Nat}((\sum_{A\in A}(XA\times\mathrm{Hom}(FA,-)))/\sim,M). \tag{29}$$

where an equivalent relation is defined by the following: First we take a functor $X\times\mathrm{Hom}(F(-), B)$ for $f\colon A\to A'$, and obtain $Xf\times\mathrm{Hom}(Ff, B)\colon XA\times\mathrm{Hom}(FA, B)\to XA'\times\mathrm{Hom}(FA', B)$. Then, $\mathrm{Hom}(Ff, B)g = gFf$. Therefore, $Xf\times\mathrm{Hom}(Ff, B)(x, gFf) = (Xf(x), g)$. Equivalent relation is defined so as to coincide the state before and after operating $X\times\mathrm{Hom}(F(-), B)$ with each other. Therefore,

$$(x, gFf)\sim(Xf(x), g) \tag{30}$$

is defined. It, therefore, leads to

$$\mathrm{Nat}((\sum_{A\in A}(XA\times\mathrm{Hom}(FA,-)))/\sim,M)\simeq\mathrm{Nat}(X,MF). \tag{31}$$

When we generalize ($\sum$ $(XA\times\mathrm{Hom}(FA, -))$)/~ by $\mathrm{Lan}_F X$, we have an expression as

$$\mathrm{Nat}(\mathrm{Lan}_F X, M) \simeq \mathrm{Nat}(X, MF). \tag{32}$$

where **Sets** is replaced by arbitrary category C and $\mathrm{Lan}_F X$ is called a left-Kan extension [42, 44].

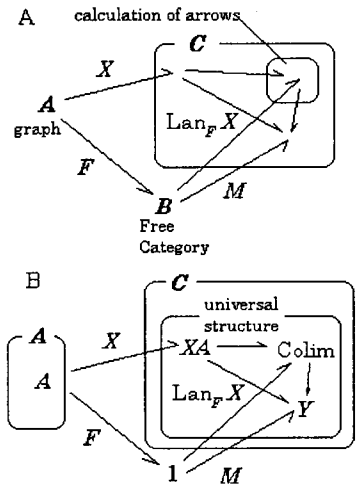


Fig. 7.10. A. Calculating All Possible Arrows in Left-Kan Extension. B. Co-limit in Left-Kan Extension.

Finally, we can say that a Left-Kan extension is a particular grounding in which the free category for a given directed graph can be calculated. In this sense, the grounding is located outside our model, and it seems to yield the separation of a model and the grounding. This is shown as Fig. 7.10A. A free category is substituted into a category B and then all arrows of free category are calculated in C. Our model, however, shows the origination of limit and co-limit. What is the relationship between these universal structures and the grounding of the left-

Kan extension? Fig. 7.10B shows co-limit in the left-Kan extension. In a diagram of the left-Kan extension, substituting a singleton set into a category B, whole diagram of the left-Kan extension itself reveals co-limit (Fig. 7.10B). It yields that the left-Kan extension has no longer lost the status of the grounding because it is not located outside our model.

The status of the grounding loses its meaning. In this section we first constitute the grounding assuming that the simulation requires a particular grounding. It, therefore, means that the grounding does not exist a priori and can be just constituted. After that we find that the universal structure in a directed graph, $G(t)$, is originated by the loss of the status of the grounding, or by the mixture of the model and its grounding. In this sense, our model can invalidate the grounding of the model itself and this invalidation can generate a particular local universal structure. The notion of the time consisting of the present, past and future is one example of this structure.

7.5 Conclusion

We start from the critique of theory of description based upon the complete separation of reality and representation, or the encodingism, and focus on the origin of universal structure that is regarded as the grounding of coding between reality and representation. In this approach, how to use the critique, skepticism, infinite regression or paradox plays an essential role in formalizing the model of internal measurement or the interaction with open surroundings. Skepticism of the separation of reality and representation starts from the assumption of such a separation and ends by leading to a paradox. We, however, argue that skepticism of skepticism is also possible because skepticism itself inevitably breaks the assumption of this separation that must be kept in the argument. We estimate the Chinese room, plus-quus problem, and Cantor's diagonal argument as skepticism, and find examples of this aspect in skepticism.

The analysis of skepticism of skepticism can yield a particular model of the interaction with open surroundings that is distinction without its grounding. Note that the surroundings or environments are regarded as semantics in an abstract sense. Therefore the notion of open surroundings means "arbitrary semantics", and we have to pay attention to the notion of arbitrariness. If we arbitrarily define a particular semantics, the interaction, however, is expressed with a particular closed semantics. The expression of the interaction loses the notion of arbitrariness or the notion of open surroundings. Finally, the notion of

open-interaction has to be expressed by the interaction without its grounding or semantics in a formal model. Whenever we use the notion of state or symbols, we express a particular distinction. The model of open interaction has to be expressed by distinction without its grounding.

The distinction in an argument by skepticism proceeds by invalidating distinction between the previous state and its grounding. This can be used for a general model of distinction or the open interaction. We propose two kind of models in this framework in which given a distinction or duality consisting of the previous state and its grounding, invalidating this distinction can generate a new distinction. In the first model, the duality is expressed as a set of states and a set of rules, and the invalidation of this duality is expressed as a self-similar set in a Lorentz plot. Because this self-similar set is used as a return map, the new state is obtained by such an invalidation. We apply this idea on the model of incomplete clock that consists of many sand clocks interacting with each other, and explain the biological evolutionary pattern called punctuated equilibrium. This illustrates another possible approach to the power-law phenomena or intermittent behaviors that has been explained as the self-organized criticality. It also suggests that incomplete clock proceeding by the invalidation of distinction between the present and future might be essential rather than the notion of time. The second model is augmented by extending the idea of incomplete clock. The duality consists of a directed graph corresponding to the realized state (the present) and its free category corresponding to the possible space. The invalidation of this distinction is defined by introducing local identity and removal of the redundancy of compositions, which yields another new directed graph. In this manner, the unit time from a graph at the tth step to a graph at the $(t+1)$th step is defined, which yields the progression of the incomplete clock. This model can generate limit and co-limit in a directed graph at a particular time step, which is interpreted as the origin of the time consisting of the present, past and future. Finally we can say that the notion of time emerges from the incomplete clock is a formal model of the internal measurement or interaction with open surroundings. In other words, the origin of external measurement or representation can be explained from the internal perspective.

Acknowledgements

We are greatly acknowledge Professor Dick Bird for careful reading, correcting English and some suggestions.

References

[1] Gunji, Y-P., in Actes du Symposium ECHO, edited by A.C. Ehresmann et al. Univerite de Picardie, Amiens, p.94, (1996).
[2] Gunji, Y-P., World Future 49, 483, (1997).
[3] Gunji Y-P., and Ito, G., PhysicaD 126, 261, (1999).
[4] Bickhard, M.H. and Terveen, L., Foundational Issues in Artificial Intelligence and Cognitive Science, Impasse and Solution, Elsevier Sc. Pub., New York, (1996).
[5] Wittgenstein, L., Philosophical Investigations, Basil Blackwell and Mott, Ltd., Oxford, (1953).
[6] Kripke, S., Naming and Necessity, Basil Blackwell, Harvard University Press, New York, (1980).
[7] Kripke, S., Wittgenstein on Rules and Private Language, Blackwell (Basil), New York, (1982).
[8] Bird, D., Acta Polytechnica Scandinavica Ma91, 44, (1998).
[9] Rossler, O.E., Chaos Solitons and Fractals 4, 415, (1994).
[10]Finkelstein, D.R., Quantum Relativity, A Synthesis of the Idea of Einstein and Heisenberg, Springer-Verlag, Berlin, (1996).
[11]Primas, H., Acta Polytechnica Scandinavica Ma91, 83, (1998).
[12]Matsuno, K., Protobiology: Physical Basis of Biology, CRC Press, Boca Raton, MI, (1989).
[13]Brooks, R.A., Science 253-5025, 1227, (1991).
[14]Tsuda, I., International journal of Neural Systems 7, 451, (1996).
[15]Ikegami, T. and Kanoko, K., Physica D 42, 235, (1990).
[16]Kaneko, K., Chaos 2, 3, (1992).
[17]Searle, J.R., Behavioral and Brain Sciences 13, 585, (1990).
[18]Bickhard, M.H., in Social and Functional Approaches to Language and Thought, Academic Press, (1987).
[19]Lawvere, F.W., Lecture Notes in Mathematics 92, (1969).
[20]Gunji, Y-P., Appl. Math. Comput. 61, 231, (1994).
[21]Gunji, Y-P., and Toyoda, S., Physica D 101, 27, (1997).
[22]Gunji, Y-P., Ito, K. and Toyoda, S., Physica D110, 289, (1997).
[23]Eldredge, N. and Gould, S.J., in Models in Paleobiology edited by T.J.M. Schopf, Freeman San Francisco, p.82, (1972).
[24]Gould, S.J. and Eldredge, N., Paleobiology 3, 114, (1997); Nature 366, 223, (1993).
[25]Gould, S.J., Phil. Trans. R. Soc. Lon. B353, 307, (1998).
[26]Bak, P., Tang, C., and Wissenfeld, K., Phys. Rev. Lett. 59, 381, (1987).
[27]Tang, C. and Bak, P., Phys. Rev. Lett. 60, 2347, (1988).

[28]Sneppen, K. and Jensen, M.H., Phys. Rev. Lett. 71, 101, (1993); Phys. Rev. Lett. 70, 3833, (1993).

[29]Bak, P. and Sneppen, K., Phys. Rev. Lett. 71, 4083, (1993).

[30]Bak, P., Sneppen, K. and Flivbjerg, H., Phys. Rev. Lett. 71, 4087, (1993).

[31]Paczuski, M., Maslov, S., and Bak, P., Phy Rev E53, 414, (1996).

[32]Bak, P. and Boettcher, S., Physica D107, 143, (1997).

[33]Bak, P., How Nature Works?, Springer-Verlag, New York, (1996).

[34]Kauffman, S.A., Origins of order: Self-Organization and Selection in Evolution, Oxford University Press, Oxford, (1992).

[35]Sornette, D. and Dornic, I., Phys. Rev. E54, 3334, (1996).

[36]Boettcher, S. and Paczuski, M., Phys. Rev. Lett. 76, 348, (1996).

[37]Bornholdt, S. and Sneppen, K., Phys. Rev. Lett. 81, 236, (1998).

[38]Kristensen, K., Donangelo, R., Koiller, B. and Sneppen, K., Phys. Rev. Lett. 81, 2380, (1998).

[39]Raup, D., Bad Genes or Bad Luck?, W.N. Norton & Company, New York, (1991).

[40]Takachi, Y., Kobe Univ. Master Thesis, (1999).

[41]Gunji, Y-P., and Takachi, Y., Phys. Rev. E (submitted).

[42]Walters, W.F.C., Category and Computation Theory, Cambridge University Press, Cambridge, (1991).

[43]Scott, D., Lecture Notes in Mathematics 188, Springer, New York, (1971).

[44]McLane, S., Categories for Working Mathematician, Springer-Verlag, New York, (1976=1998(2d ed.)).

This page is intentionally left blank

Chapter 8

The Logical Jump in Shell Changing in Hermit Crab and Tool Experiment in the Ants

Nobuhide Kitabayashi, Yoshiyuki Kusunoki, and Yukio-P. Gunji
Kobe University

Abstract

We propose and sketch a novel approach toward the study of behavioral plasticity. When one encounters a new animal behavior, one formally describes this behavioral pattern, however, one confronts with more new behavioral patterns as the observation proceeds. As a result, the question arises how the hierarchy in behavioral pattern is originated and/or is changed. Focusing on the relationship between before and after the emergence of new behavioral pattern, we explain emergency a newly behavior and the origin of hierarchy of behavior. In particular, we illustrate the changing shell in experiments by hermit crabs, and the usage of a cart in experiments on the transportation of foods by ants. We observe a particular time-series sequence of hermit crabs and ants behaviors following a 1/f or Zipf's law. The behaviors following the 1/f or Zipf's law manifest emergence as logical jump on the part of object.

Keywords : 1/f noise, Zipf's law, hermit crabs, changing shell, ants, usage of tool, food-retrieving behavior, emergent property, anthropomorphism

8.1 Introduction

In natural condition it is said that hermit crabs move only forward, and do backward only in changing a shell [1]. Hermit crabs have to carry on the back shell in order to avoid counter-pest and cannibalism. It can be said that the shell is a boundary between the inside and outside. Moving forward and backward can be compared to the behavior toward outside and inside, respectively. In this study, we examined the behavior of hermit crabs without a shell in order to expose hermit crabs to the situation in which hermit crabs have to determine where is inside or outside. In preliminary experiments, we observed a behavior

as follows: first, a hermit crab moved backward into a tube, and subsequently did forward to the entrance of a tube from its point, and it sometime repeated. It can be said that backward behavior to enter into a tube reveals that the inside part from the entrance of a tube is the inside for the individual. Subsequently forward behavior toward the entrance of tube can reveals that the outside part of a tube from a turning point is reconsidered as the outside for a individual. So one can see that a hierarchy about inside and outside appear in repeat of forward and backward behaviors. In shell changing experiment, this behavior were roughly estimated by the distance of locomotion because the speed of locomotion decrease at turning points.

In order to describe the estimation to a new situation on the part of the object, one of the authors proposed a mathematical model [2], and applied it to the schooling behavior of fish [3]. We illustrated the usage of a cart in the food-retrieving behavior of ants, and proposed a weak definition for usage of a tool connected a 1/f or Zipf's law [4]. Mizukami and Gunji demonstrated an exemplification of Zipf's law applied to the final stage of the learning process of goldfish [5]. Mochizuki also found a 1/f noise characteristic in the territorial behavior of three-spined sticklebacks when they are subjected to a paradoxical situation [6]. We apply the idea of inevitable discovery [2] to the behavior of hermit crabs in a new condition.

In investigating animal behaviors, one can often notice the human-like behavior, and relevant reports on animal behavior frequently contain implicit anthropomorphic assumptions [7, 8]. For example, it has been observed that wild chimpanzees use sticks to eat termites from mounds [9, 10, 11]. This process is often called 'fishing termites'. Generally, in the case of monkeys and apes, the notion of tool-using is utilized for describing their behavior [12, 13]. In contrast, for case of lower animals, one may regard the expression of tool-using as a misinterpretation, because of apparent absence of intelligence. Anthropomorphism is generally considered to play an important role in sentimental attitudes towards animals. Of course, scientists can not guard against immediate attribution of a human motivation or emotion to the animal, when animals are observed to behave in human-like way. Such kinds of behavior can appear even among neighbors in terms of an observer (you) and an object (a neighbor) because a mechanism, which refuses the idea of anthropomorphism, can be applied also to a human being. This kind of contradiction cannot be removed in the framework of the philosophical investigation made by Wittgenstein and Kripke [14]. As far as it goes, one cannot help avoiding the term "anthropomorphism". The anthropomorphic nature of a human leads to inevitable contra-

dictions in investigating animal behavior. So, if anthropomorphism appears unavoidable, is it possible to take it into account as an important aspect of animal behavior?

In tool experiment, we concentrate on food-retrieving behavior of ants. Generally, the ants are known to change the way they carry food, depending upon the size of food. Ants carry small food items singly. When they encounter large food, they carry it to the nest in a cooperative action [15, 16, 17, 18]. In our own experience, we often construct a piled food on a dish, when fine grained food (the upper one) is put over the saucer-like food (the lower one). It is demonstrated by our experiments that ants prefer fine grained foods to saucer-like ones. We observed that ants transported saucer-like food with fine grained foods on the top of it. This behavior can be interpreted as usage of cart as a tool. It is interesting to note that some digger wasps were also reported to use pebbles as a tool to tamp down soil to seal their burrows [19]. We measure the instances when an observer detects the start and finish of the transportation process, and it is performed in anthropomorphic way. First, it appears that an observer cannot avoid anthropomorphism regardless of his intention, as will be discussed below. Second, an inevitable discovery follows, and anthropomorphism becomes performative. Third, we can perform a measurement of Zipf's law through this inevitable discovery just before the transportation process, which one cannot help anthropomorphizing. In tool experiment, we describe an experiment on ants carrying food, estimate whether they use a tool or not, and verify the relationship between the usage of a tool and Zipf's law.

8.2 A Relationship between a Self-similarity and a Paradox

Let us consider how the usage of a tool can result in a paradox. Typically, when one observes that a chimpanzee pokes a stick into the nest of ants and eats ants by slicking the branch, it is said that a chimpanzee uses a stick for fishing. Is this not a naive anthropomorphism? Can one reach the idea of fishing or of the tool used by a chimpanzee without anthropomorphism? In order to clarify this point, imagine that an observer believes that chimpanzee has no ability to use tools, and the observer attempts to study chimpanzee's feeding habits. He has to determine whether a material is food or non-food without ambiguity, and that requires the definition of food. Imagine that he defines chimpanzee's food as a substance that is carried toward a chimpanzee's mouth by his hand and subsequently vanishes in his mouth. Most substances can be proved to be either food or non-food, though an observer can find a particular material that can be

proved neither food nor non-food. This material is carried toward a chimpanzee's mouth, but it cannot vanish in it. As far as an observer does not give up his definition of chimpanzee's food, the existence of such a material leads to a logical contradiction. Due to this inconsistency one can resolve the contradiction by introducing a new notion in addition to "food" and "non-food". It suggests a kind of logical jump, and the underlying paradox is similar to the Russel's paradox in a set theory. Russel's paradox results from mixing the notion of indicating an element with indicating a set consisting of elements [20]. It is expressed as follows; the first type set is defined as a set involving itself as an element, and the second type set is defined as a set not involving itself. A specific set of 'whole2' is defined as a set consisting of all second type sets. Now supposing that whole2 is the first type, whole2 involves whole2 due to the definition of the first type set. Therefore, from the definition of whole2, whole2 does not involve whole2, and it means that whole2 is the second type set. That is a inconsistent with the assumption. By contraries, supposing that whole2 is the second type, whole2 must be involved in a set of all second type sets, that is nothing but whole2. It means that whole2 is the first type. Finally both assumptions entail to a contradiction, and that is a paradox. In other words, one indicates an element inclusively on one hand, and one addresses the symbol indicating wholeness of elements on the other hand. Compared with the situation in which an observer reaches the third meaning via a contradiction of whether food nor non-food, it can be considered that the third meaning is acquired by addressing the symbol indicating the aspect of a contradiction. If one attempts to formally describe this logical jump, one is destined to confront Russel's paradox. We can see double inconsistency in this situation. The first one is a contradiction in terms of the definition of food, and the second one appears in resolving a paradox. It implies that there is no formal way to resolve a paradox.

We can see a logical jump in an informal way, as well as the procedure of constructing denotational semantics. The first contradiction of neither food nor non-food can be expressed as a fixed point of the operation of negation (because food and non-food are connected by negation, *Not*, such as *Not*(food) = non-food). If the operation of negation is denoted by a symbol F, and a substance of whether food or non-food is denoted by x, we obtain $x = F(x)$, and x is called a fixed point of F. Note that the third meaning should satisfy this formula or be a solution of this equation [21]. If one introduces this type of novel meaning, the new x is both food and non-food. This implies a logical jump from a fixed point of neither food nor non-food to a fixed point of both food

and non-food. Note that even the latter statement is not the resolution of a paradox because an observer assumes first that any material is proved to be either food or non-food, and, hence, the statement is still inconsistent. This situation implies double paradox as mentioned above, though it shows the positive implication of a fixed point or a contradiction [22, 23]. The logical jump is not the formal resolution in its turn. It suggests the situation in which the new notion is introduced. This new notion is introduced by the observer and cannot be interpreted either as food or as non-food. Note, that there is still no grounds for an observer to call this new meaning a tool. The logic after introducing new notion is still anthropomorphic. Without anthropomorphism, we could not use the term "tool". However, its existence paradoxically introduces the impression of using a tool, or provides a way to give a 'weak' definition of a tool used in feeding. Weak definition does not mean that a particular tool must correspond to this definition, but that one can use the term "tool" whenever one encounters a similar situation. It is worth also noting that the notion of weak definition is weaker than necessary condition, because it is not necessary. In this paper, a weak definition of a tool, that uses the property of self-similarity, is given. First, we focus on a fixed point x of the operation of determination of food or non-food, denoted by F. This fixed point can be also expressed as an infinite recursion, $x = F(F(F(...F(x)...)))$ by infinite substitution of $x = F(x)$ to $x = F(x)$. Let's consider this fixed point x as a two dimensional material. Then, this infinite recursion can imply infinite precision in terms of determinant of food or non-food. In other words, we define the operation of F as the contraction in a two dimensional domain, indicating either food or non-food. Because $F(x) \cong x$ implies a self-similar set, we can find s-arm branching trees as x. In this x, the length of k th branch, which is denoted by x, is n/sk, where n is the length of the first branch, and the number of branch or the frequency of branch x, which is denoted by $f(x)$, is sk. Therefore we get $x \bullet f(x) = (n/sk) \bullet sk = n$ (constant). Roughly speaking, this idea can be generalized as one of solution of $F(x) \cong x$. A fixed point-based definition of tool implies that the notion of a material used as a tool shows self-similarity, i.e. it is invariant in terms of faithfulness of the notion of food however it is contracted. Faithfulness of the notion of food can be estimated by a particular definition, as discussed in later section. If faithfulness of food is denoted by m, invariance of faithfulness with respect to contraction is expressed as $f(m)*m$ = constant, where m is the value of faithfulness and $f(m)$ is probability of m. If distribution of $f(m)$ does not have off-set peak, m directly means rank. Then $f(m)*m = c$ represents, which is called, Zipf's law, and i.e., $log(f(m)) = -m + c$. Therefore, we propose a weak definition for the

notion of tool related to Zipf's law calculated for a particular food substance. Under this definition, an observer cannot detect Zipf's law until chimpanzee or another particular animal uses the tool. We apply this idea to the behavior of ants. In the next section we discuss an experiment on ants transporting food, estimate whether they use a tool or not, and verify the Zipf's law for the usage of tool.

8.3 Methods

8.3.1 *Shell Changing Experiment*

We collected the terrestrial hermit crab *Coenobita purpureus Stimpson* at Sesoko-jima of Okinawa Islands, Japan. All the experiments were conducted in laboratory, September of 1997. Hermit crabs without a shell were provide by roasting tips of shell of hermit crab over a little fire.

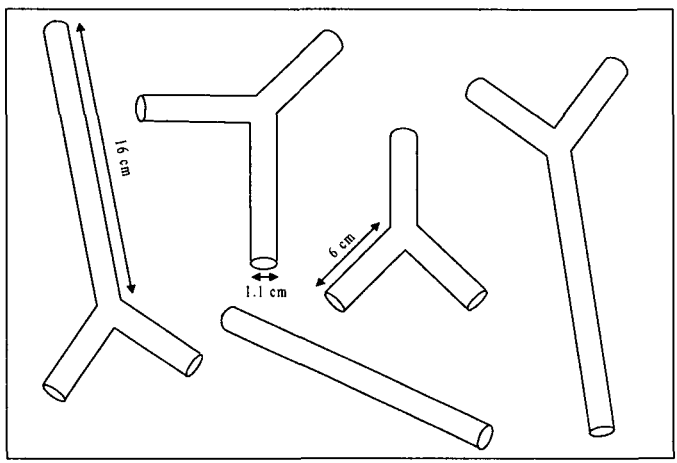


Fig. 8.1 Experimental Setup on the Bottom of the Aquarium. There is no sand and water.

Behavior of 10 hermit crabs in the experimental setup (Fig. 8.1) were recorded by a digital video camera for 1 hours per a trail. 23 independent runs of experiment were conducted and the sum of using individual is 230 included 34 one with shell. In order to estimate the attitude of hermit crabs toward inside

and/or outside as explained above, the time series of the distance of individual's locomotion per a 1/2 second were referred to and analyzed in terms of its power spectrum. The image processor (EASY, Library co., Japan) was used to measure the distance of individual's locomotion. When hermit crabs were out of tube one by one, we called this behavior 'singular behavior'. When hermit crabs were out of tube with another or more individual within a life-size, we called it 'plural behavior'. We called hermit crabs with a shell 'shell behavior'. When hermit crabs were in a tube, we called this behavior 'tube behavior'. Especially in terms of tube behavior, it is divided into 'moving behavior' and 'staying behavior' dependent on the sum of distance of locomotion. The distance of locomotion includes the locomotion carried a tube on the back like a shell. In each state of an individual, the start of analysis is a point of time of the first locomotion in flame to measure by image processor.

8.3.2.1 *Food on Food*

In order to demonstrate some details of ants' usage of a tool, we focus on whether ants use a handcart in food transportation or not. Given a particular food placed over a dish-like material, we observed whether ants concentrated around the dish-like material, and carried the material with another food on it, or not. If it was observed, it looked like ants using the dish-like material as a handcart to carry another food. As mentioned above, we can expect that Zipf's law and/or 1/f power spectra can be observed before ants start using a handcart. Of course, in natural conditions, one cannot observe ants using a handcart. Therefore, if a particular ants' food is put on a dish-like material that can not be eaten by ants (e.g. plastic plate), one cannot expect the ants to use a handcart. So, we conducted experiments with 'food on food', where the upper food was preferred to the lower one. If ants concentrate around the lower food and carry it, no matter what food is on its top, we conclude that the usage of a handcart is observed. First, we conducted preliminary experiments aimed to evaluate the food preferences of ants. Second, we experimented with food-retrieving process, for various combinations of food constructed as "food on food". The experiments of the first and second type were conducted under specific sets of conditions called Exp.1 and Exp.2.

8.3.2.2 *Preliminary Experiments 1*

All the field experiments were conducted at the campus of Kobe University, Japan, in the period from July to October of 1996 in order to investigate wheth-

er *Formica japonica Motschulsky* change the way of food transportation depending on the size of food pieces (solid dise-like, or fine grained). Preliminary experiment 1 was conducted to elucidate which food, sardine or bacon, yolk or white of an boiled egg, and slightly roasted salted spawn of pollack or Bologna sausage, was preferred by ants. Rough observation was made for the each pair of foods in preliminary experiments. The foods were put on a creamy white wooden board 31×31 cm^2 and 1cm thick, separated by 10 cm, at the distance of about 1 m from the nest. Behavior of ants was recorded by 8mm video camera. After the analyzing of records, the preferred food was defined and used as an upper food in the next tool experiment1. In all experiments conducted in this paper, if ants from another nest intervened, the recording process was interrupted or the corresponding time series discarded. The ants coming from other nests were detected by different direction of their appearance. To quantify ants' preferences, taste index is introduced as

$$t = Tm/(Tm + Tl),$$

where Tm is the number of ants transporting the preferred food, and Tl is the number of individuals transporting the other one.

8.3.2.3 *Tool Experiment 1*

To investigate whether the way how Formica japonica Motschulsky carry food depends on the size of food only, we introduced the piled food. On the creamy white wooden board, the upper food was placed over the lower food. The preferred kinds of food, like sardine, yolk and pollack, were chosen as the upper food, whereas bacon, white and sausage were taken as the lower one. The upper food was prepared fine grained, and the lower one was cut in 2 cm squares of 1 mm thick. The kinds of food, salted spawn of pollack, yolk, bacon, sardine, Bologna sausage, white of an egg, are denoted as P, Y, B, Sr, S, W, respectively. A piled food is denoted as SrB if the upper food is sardine (Sr) and the lower one is bacon (B). Three kinds of experiments in terms of food combinations (SrB, YB, PS) have been conducted. Food-retrieving behavior was recorded by an 8mm video camera. It was expected that the upper, more preferred, food would be transported first, while the lower one would be left on the wooden board. To avoid accidental transportation, special control experiments have been also conducted for all kinds of the upper food. In these experiments, the upper food was put on cloths, and no transportation was ob-

served.

We have analyzed position of individual ants with respect to the foods they touched. As a result, food index is introduced as

$$f = Pu/(Pu + Pl),$$

where Pu is the number of ants touching the upper food and Pl is the corresponding number for the lower food. The numbers Pl and Pu were measured once per second. Pl was estimated as the number of ants whose head was in contact with the edge of the lower food at the moment the measurement was performed. Pu was estimated as the number of ants on the top surface of the lower food (covered by the upper food) at the moment of measurement. In order to estimate ants' attitude to each kind of food, the food index was measured and analyzed in terms of its power spectrum.

8.3.2.4 *Preliminary Experiments 2*

Preliminary experiments 2 were conducted similar to preliminary experiments 1, but this time the food preferences were estimated for more food combinations of pollack, yolk, sausage, white, sardine, bacon. The experiments were conducted on a board 75×173 cm^2 and 1cm thick. In order to estimate the preference between two kinds of foods, the fine grained food of two sorts was put on a board separated by the distance of 20 centimeters. Recruitment and food transportation processes were recorded for 10 minutes by a digital video camera installed in the field . Preferred food was used as the upper one, and one and the same kind of food was used during a day. As a result of these experiments, the taste indexes has been arranged in descending order, and the taste ranks has been assigned to each kind of food. All the field experiments were conducted at the campus of Kobe University, Japan, from 15th of August to 11th of September, 1997.

8.3.2.5 *Tool Experiment 2*

The upper food (total amount about 1.0g) was fine grained to small particles and the lower one was cut in 2 centimeter squares of 2 millimeters thick. Recruitment and food transportation were recorded by a digital video camera installed in the field from about 14:00 until the sunset. When the number of ants coming from the nest and concentrated around the target food were large enough, the recruitment and food transportation process were recorded for

5×50 minutes in each of the experiments with a particular food combination. After almost all upper food had been transported, the upper food particles were supplied, thus defining the end of a time series. Each 50 min record can consist of several consecutive time series. Table 8.1 shows all the used combination of food. The number in parentheses shows the taste index. Columns and rows represent upper food and lower food, respectively. Blank cells correspond to the absence of experiments because of reverse preferences of ants to corresponding foods (as follows from preliminary experiments). 88 experiments, including 5 control experiments (with the same upper and lower foods) were conducted. 5 experiments per a combination of upper and lower foods were as a rule carried out, except for control experiments.

upper food / lower food	P	Y	B	S	Sr	W
P	1					
Y	5(.87)	1				
B	5(.92)	5(.66)	1	5(.61)	2(.80)	
S	5(.94)	5(.83)	6(.68)	1		5(.57)
Sr	5(.80)	5(.76)	5(.75)	5(.54)	1	
W	5(.92)	5(.92)	5(.74)		5(.55)	1

Table 8.1 The Number of Control Experiments for Different Food Combinations.

In our experiments, there were so many ants distributed over the target piled food that we gave up direct counting of their number by eye. Instead, we used the image processor (EASY, Library co., Japan), measuring the areas occupied by ants being in contact with food in two dimensional space. The number of ants touching food was estimated by the total area they occupy. The area was sampled every 4/15 seconds. In order to estimate Zipf's law in terms of the food rate, we introduce a time window of 125 time steps. Then, moving this window along the food rate time series, at each time step we divide it to subintervals of length

$$Interval\ [i] = t_{i+1} - t_i,$$

where $|f(t_{i+1}) - f(t_i)| = 100$, and t^0 corresponds to the beginning of the window. At each time step we order the sub intervals with the window by descending order and analyze the histogram H(R) for their ranks R. Then we suppose that the following relation holds and extract the slope s(t) of the regression line in logarithmic coordinates by least square method. The points of the time series corresponding to the occasion when no ant touches the lower food have been removed from the analysis of the slope[t]. Note, that the distribution with the s(t)1.0 is called Zipf's law. Finally, we investigate the temporal change of the slope[t], which is defined by the type of the piled food carried by ants. In the tool-experiment 1, a ratio between the number of ants touching the upper food and lower food before the start of transportation was analyzed by means of Fourier transform. In tool-experiment 2, when the number of ants in contact with upper food was not possible to measure directly, the ratio was roughly estimated by analyzing the dynamics of residuals in the time series. It can be reasonable because of time lag of individual motion in touching lower food before approaching upper food and touching lower food after leaving upper food. If ants carry the lower food with the upper food on the top of it, we called this behavior 'tool transportation'. If ants carry the upper food only, we call it 'no transportation'. The time before and around the start of tool transportation is called 'before transportation'. The tool transportation, longer than 200 time steps, is called 'long transportation'. In the case of tool transportation, a time series was analyzed from the start of transportation till its finish, or until the instance of interruption by the supply of the upper food. The measured values during this time interval are called a single time series. Transportation index is defined as

$$CR = c/l,$$

where c is the duration of transportation in a single time series and l represents the total duration of the time series.

8.4 Results

8.4.1 *Shell Changing Experiment*

We analyzed 27 cases in tube behavior; 29 ones in shell behavior, 8 ones in

moving behavior, 8 ones in staying behavior, 13 ones in singular behavior, and 15 ones in plural behavior. There is a typical pattern of the exponent of power spectra in natural phenomena which the exponent around low frequency are different from the slope around high frequency. So, in this study, we analyzed in terms of its power spectrum from 1 or 2 to 60 frequency.

Fig. 8.2 shows a relationship between the slope and frequency in the cases of tube behavior and shell behavior. In the cases of tube behavior, which was nude individuals behaving in a tube, a peak in (-1.0, -1.1] was seen. In the cases of shell behavior, which was natural condition in term of a shell, three peaks in (-0.8, -0.9], (-1.2, -1.3] and (-1.4, -1.5] were seen. The distribution of the slope showed statistically significant difference between the cases of tube behavior and shell behavior. We used the both side t-testing procedure for estimating the significance of the difference (P=0.003).

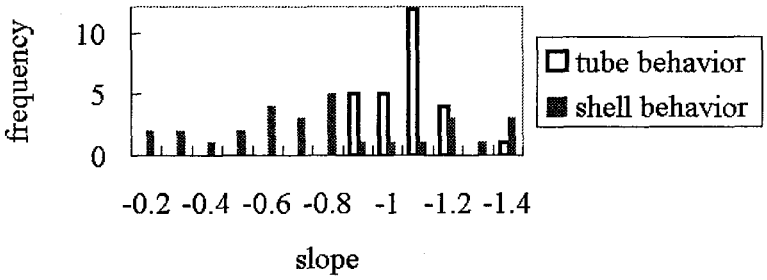


Fig. 8.2 Relationship between Tube Behavior and Shell Behavior. It is the histogram of the slope.

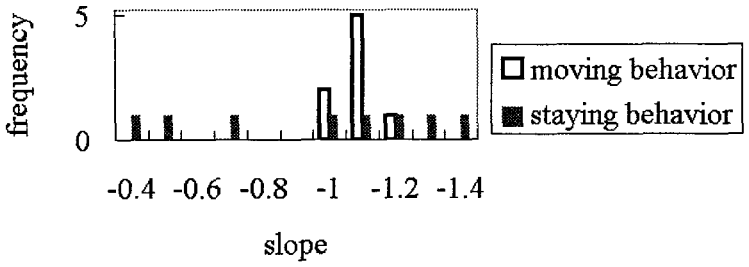


Fig. 8.3 Relationship between Moving Behavior and Staying Behavior. It is the histogram of the slope.

Fig. 8.3 shows the relationship between the cases of moving behavior and its staying behavior, which are two state of the same individual. In the cases of moving behavior, highest peak in (-1.1, -1.2] was seen. In the cases of staying behavior, no peak were seen.

Fig. 8.4 shows the relationship between the cases of singular behavior and plural behavior. In the cases of singular behavior, highest peak in (-1.0, -1.1] was seen. In case of plural behavior, highest peaks in (-0.7, -0.8] was seen.

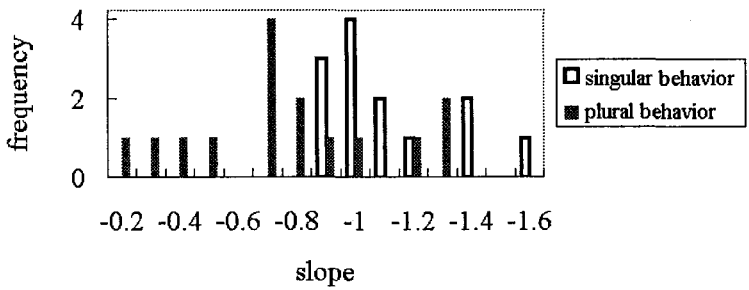


Fig. 8.4 Relationship between Singular Behavior and Plural Behavior. It is the histogram of the slope.

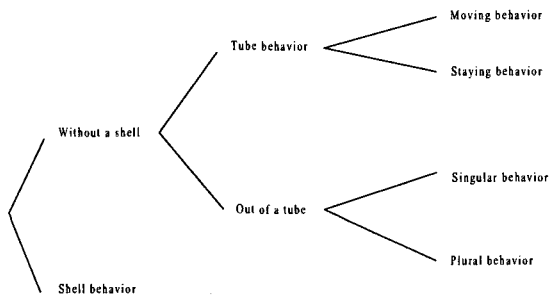


Fig. 8.5 As a Result, a Hierarchy was Found.

8.4.2 *Tool-Experiment*

Fig. 8.6 shows the power spectrum of a time series of food index before the start of tool transportation, and that in the corresponding control experiment. In logarithmic coordinates the spectra show near linear behavior.

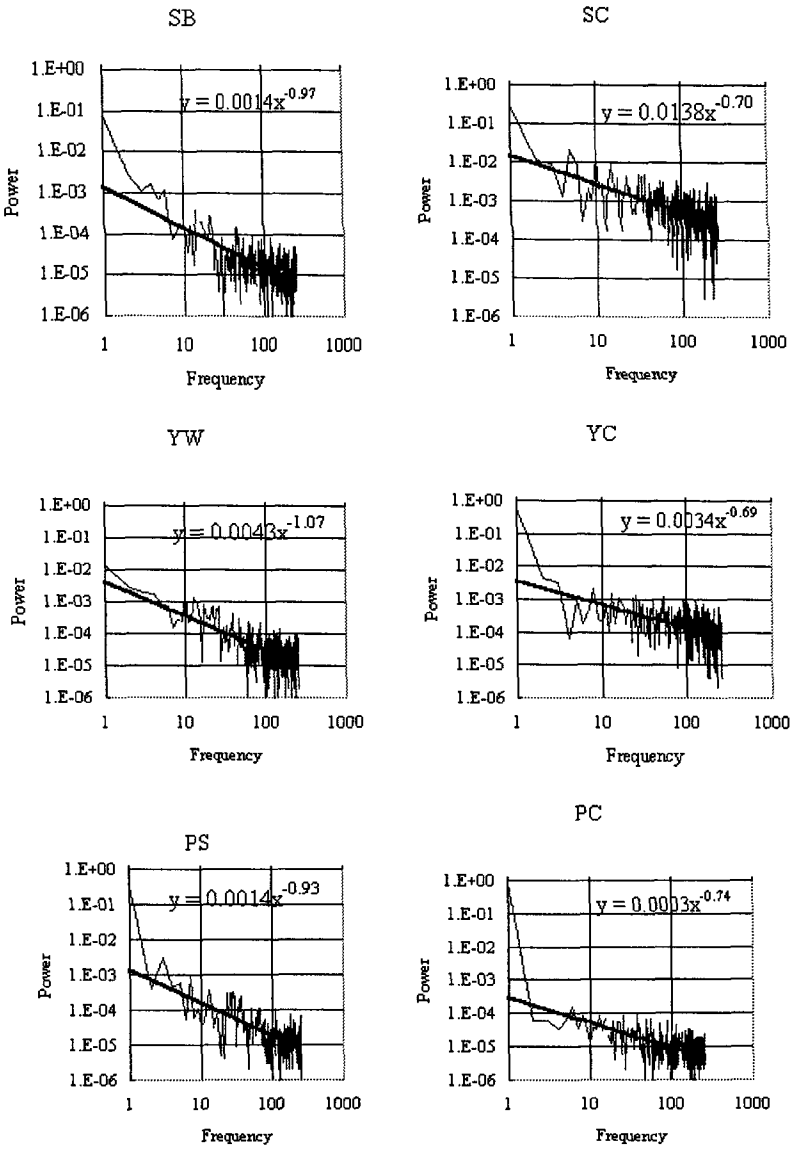


Fig. 8.6 Power Spectra of Time Series of Food Rate: around the beginning of tool transportation (left graphs), control experiments (right graphs).

At the beginning of food transportation, the slopes are nearly -1.0 for either food combination, SB, YW or PS, as shown in Fig. 8.6. In the control experiments, slopes are always far from -1.0, and for any of SC, YC or PC, power spectra have no correlation in frequency, thus demonstrating the behavior similar to that of thermal noise. If lower part is not food, ants need not to make an estimate of it. However, if both upper and lower parts are food, ants estimate food prior to transporting it. The slope -1.0 in power spectra could suggest ants' own estimation to food [21, 24]. The estimate of the slope as -1.0 just before transportation may be not accurate enough for a quantitative analysis. However, it can be used as a qualitative characteristic to distinguish the slopes in before transportation and control experiments.

8.4.3 *Tool-Experiment 2*

In 25 out of 88 experiments, the behavior corresponding to the transportation of lower food together with the upper one was observed. Fig. 8.7-a shows a typical time series of the slope[t] for the tool transportation process. The black parts of the bar in the upper part of the plot show the time intervals when well defined tool transportation occurred, shaded parts correspond to indefinite small motions of the food, and white parts indicate no motion at all. Ants begin to transport food at about the 500th step. After a while ants stop the transportation, and from about the 700th step they resume the food transportation. The transportation process lasts from the 750th step to the 1100th step. After the 1250th step, the transportation continues in an intermittent manner. It is important to note, that sharp decrease in slope[t] takes place around the start of tool transportation, and at the 887th step the slope is very close to -1.0, as shown in Fig. 8.8-a. Such decreases occur in some pauses of tool transportation, near the 1500th, 2400th, and 2750th step. The minimal values of the slope corresponding to each of these drops gradually increase in the long transportation. Fig. 8.8-b shows an example of the slope behavior in a long transportation process. In the cases of no transportation or control experiments, the drops do not occur and the slope is far from - 1.0, as shown in Fig. 8.7-b and 8.7-c. So we can claim that there is a qualitative difference between before transportation and long/no transportation. Minimum values of slope[t] in a time series which follow Zipf's law occur before or at the beginning of tool transportation in all the analyzed tool transportation cases. Mean values of the amplitude variation (difference between minimum and maximum values), maximal and minimal values of a time series, calculated by averaging over all our experiments, showed sta-

tistically significant difference between the cases of tool transportation and no transportation. We used both side Mann-Whitney testing procedure for estimating the significance of the difference (P(U=25)=0.05, P(U=41.5)=0.05, P(U=90)=0.10). Thus, the results of statistical analysis show that the drops in the slope time series of tool transportation are not contingent.

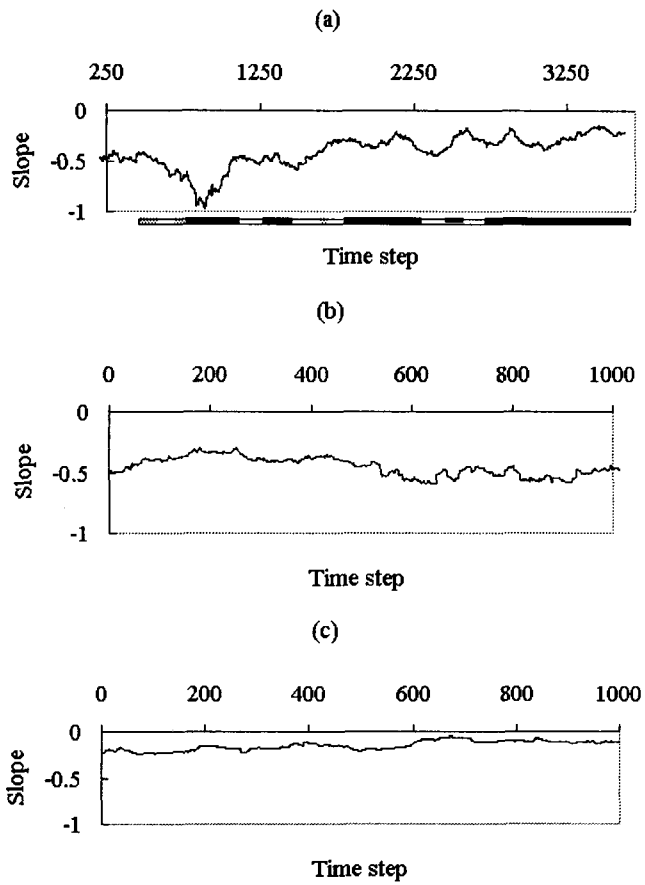


Fig. 8.7 (a) A Time Series of Tool Transportation. Gray parts of the bar in the upper part correspond to small motions of food at about the same place, black parts correspond to apparent motions. (b) A Time Series of No Transportation. (c) A Time Series of a Control.

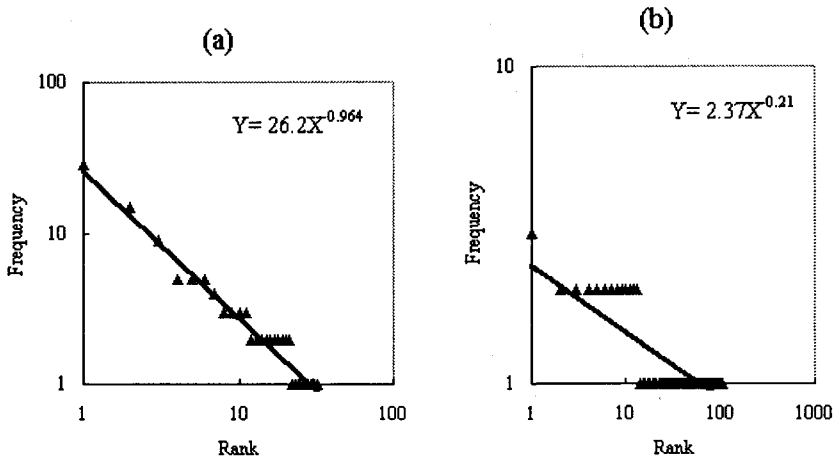


Fig. 8.8 The Histogram of the Rank of Interval [i]. (a) Before Transportation; (b) Long Transportation.

Fig. 8.9 shows the least slope[t] before tool transportation vs. the time duration of the interval between the step with the least slope and the step with the first subsequent maximal slope[t]. It corresponds to 16 before transportation time series, 13 long transportation ones (with 3 time series out of 16, corresponding to before transportation ones, discarded, as not containing long transportation), 18 no transportation time series. Most of no transportation cases (circles) lie to the left of the dashed line. On the other hand, most of before transportation cases (squares) are located to the right of it. The correlation coefficients to corresponding circles and squares are -0.826 and -0.739 respectively ($\alpha < 0.001$ in the both side testing), and there is a statistically significant difference between them ($z = 0.602$, $\alpha < 0.01$, in the both side testing). Therefore, it provides an additional evidence that the sharp decrease of slope[t] before transportation is not contingent. Most of long transportation cases (triangles) are situated around or to the left of no transportation ones. One cannot find a significant difference between these two cases, because the correlation coefficient for the long transportation data is not significant. However, its position on the plot indicates that the cases of before- and long transportation are different. As follows from our analysis, one can safely distinguish between tool transportation and no transportation. Zipf's law is not a universal characteristic of the motion of ants. It is only relevant for the usage of a cart, and can be used to character-

ize the start of transportation process. Note, that at the start of transportation an observer always encounters a logical contradiction in the description of it, and has to change the notion of a "tool" to resolve it.

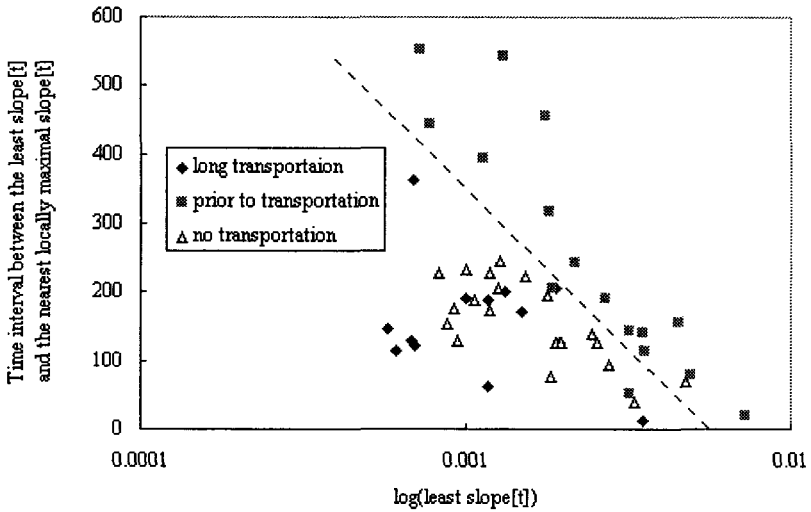


Fig. 8.9 Relaxation Time after the Drop in the Amplitude of Slope for Different Types of Transportation.

Fig. 8.10-a shows transportation index vs. taste ranks sorted by descending order. Possible upper foods for a given lower food were defined as those having the food index > 0.5. So, the taste ranks are arranged in the following order P > Y > B > S > Sr > W, as shown in Table 8.1. The squares along the diagonal on the bottom plane correspond to control experiments, that is the same food as the upper and lower one. Despite the expectation following from a machinery model of ants' behavior, a combination between near tasting ranks (PP, PY, YY, BB, BS, SS, SB, SSr, SrSr, SrW, WW) does not give rise to any tool transportation at all. Most preferable tool transportation occurs for a combination of moderate taste ranks. Fig. 8.10-b shows transportation index vs. taste index. The upper foods are sorted and arranged by descending order of food index, denoted by 1, 2,.., 6. These results show that the greater is the difference in preference between the upper and lower food, the more often the tool transportation takes place. On the other hand, tool transportation does not happen for

the upper food with the lowest taste index in combination with either of lower foods. Note, that the order of taste indexes is different from that of taste ranks. However, whether we take the taste rank or taste index, the result of Fig. 8.10-a and Fig. 8.10-b are in contrast to a machinery model of ants' behavior. It can be said that these results demonstrate a paradox demanding a logical jump, and suggest to use the notion of tool as this jump. One can also note that it would be difficult to explain the relationship between two kinds of food without ants' estimation.

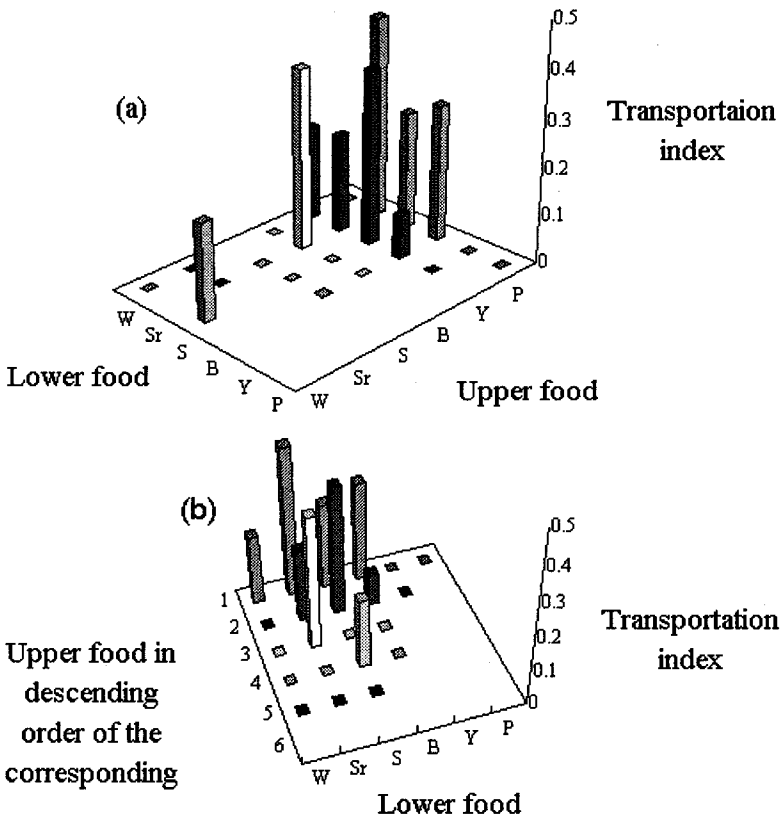


Fig. 8.10 Transportation Index According to (a) Taste Ranks and (b) Ranks of Taste Inndex. 0 value of transportation index corresponds to no transportation.

8.5 Conclusion and Future Perspective

As a result, in shell changing experiment a hierarchy was found as shown in Fig. 8.5. Even if one tries to identify this hierarchy, 1/f like-power spectrum suggests the level of hierarchy. In each level of hierarchy, the distribution of the slope in the power spectrum could suggest an active participation on the part of hermit crabs for locomotion as mentioned below. Firstly, compared tube behavior with shell behavior, individuals of shell behavior need not determine a new boundary between inside and outside in a shell because shell is a accustomed and firm boundary. It is said that a boundary is determine by the end of shell. On the other hand, individuals of tube behavior, in which most of case were near the slope -1.0, need to determine where the end of a shell is, because an artificial tube is opened at both sides. Secondly, compared moving behavior with staying behavior, which are two states in a tube, one can regard moving behavior as trying to determine a boundary in a tube with no end. On the other hand, one can regard staying behavior, in which the slope were always far from -1.0, as compromising an indefinite boundary. It is important to note that the difference of distribution of the slope between the cases of moving behavior and staying behavior cannot be influenced by a characteristics of an individual because the same individuals behave two antipodal states. Thirdly, compared singular behavior with plural behavior, one can regard singular behavior, in which a peak was in (1.0, 1.1], as trying to search a shell. On the other hand, one can regard plural behavior, in which peaks were in (-0.7, -0.8] and (-1.3, -1.4], as compromising the situation without a shell irrespective of the danger of its life. These results illustrates the relationship between the behavior following the 1/f and its positiveness toward outside. We showed the emergence of a hierarchy of behavior confronting a paradox such as a situation without a shell and no shell in an experimental setup, and the relationship between behavior and the distribution of the slope in terms of power spectrum. Any attempt for describing animal behaviors encompassing the notion of autonomy may bring us into a confrontation to a contradiction between animism on one hand and closed description [4]. We proposed 1/f as an exemplification against such a dualism.

In tool experiments we observed ants transporting food. The observations have been performed within a short time interval in order to reduce the possibility of foraging behavior of ants belonging to a particular colony. To avoid food accumulation in a single nest, plural colonies were provided to our experiments, that made their foraging behavior approximately the same. Ants

could change their preference in food, depending on the condition of colonies and/or on seasons. However, we did not encounter these problems because we carried out the experiments during one and the same season. We propose a novel approach to regarding the problem of anthropomorphism by illustrating some aspects of tool usage. Anthropomorphism is generally precluded in modern science, including biology, and is regarded sometimes just as an anachronism in ethology. However, this notion underlies the observed behavior of animals. In describing the process of making decision, definition and/or usage of a tool, one cannot say that anthropomorphism has nothing to do with the description of the behavior of animals. On one hand, naive anthropomorphism cannot be used as a scientific term or notion, and, on the other hand, trivial description of the behavior of animals could entail a logical contradiction [25]. Especially in ethology, any attempt to express an aspect of animals' act that may encompass the notion of communication can confront us with the contradiction between naive anthropomorphism and trivial description. In our perspective based upon inseparability of an object and an observer, there is no break-through idea in a closed dualism of naive anthropomorphism and description. One has to look over the relationship between the antagonistic moments, and to talk about observer's inevitable logical jump from machinery-like description to the acceptance of anthropomorphism. If an observer manages to preserve the machinery-like description of the behavior of animals, he can confront a logical contradiction and can conclude that machinery-like description has less potential to explain the behavior of animals. This inference can make the one accepting the notion of anthropomorphism. In other words, an inevitable acceptance of the notion of animal's own usage of a tool enables an observer to get rid of this logical contradiction and/or make a logical jump. Therefore, a notion of the tool usage that trivially appears in ethological texts cannot be used without an inconsistency. At a glance, it sounds as if the appearance of inconsistency could reduce a scientific significance of a particular research. However, one can replace the inconsistency by a particular measurable self-similar pattern in a structure exhibiting a contradiction.

In this paper, we propose how to detect the instance when the inconsistency appears in a time series of the behavior of animals, and show that 1/f noise-like behavior and/or Zipf's law can be obtained in a particular time interval, when an observer cannot avoid referring to animal's own usage of a tool. The instance when an observer has to accept the notion of ants' own usage of a tool occurs before ants begin to use a tool. There is a weak definition of the notion of usage of a tool relevant for the inevitable appearance of anthropomorphism.

It can be shown that Zipf's law may be strongly relevant for the description of this property in a particular system, and the notion of an emergent property cannot be separated from the observer's viewpoint [21]. The proposed weak definition of tool usage and autonomy can be used in various problems concerning the notion of origin.

Acknowledgements

We would like to thank Professor Ito, K. for extensive support during this research, Asistant professor Nagahama, T., Technical Advisor Nakano, Y and Sangen, S. for support for experiments, Dr. Rabov,V. B. corrected for English.

References

[1] Gunji, Y.P., Biology Forum 89, pp.69-78, (1996).
[2] Gunji, Y.P., & Toyoda, S., Physica D 101, pp.27-54, (1997).
[3] Gunji, Y.P., Kusunoki, Y., Chaos, Solitons & Fractals 8, 10, pp.1623-1630, (1997).
[4] Kitabayashi, N. & Gunji, Y.P., Biology Forum 90, pp.393-422, (1997).
[5] Mizukami, E. & Gunji, Y.P., (submitted).
[6] Mochizuki, T., Master Thesis in Kobe University, (1996).
[7] Gould, J.L. and Gould, C.G., The Animal Mind. Scientific American Library, New York, (1994).
[8] Griffin, D.R., Animal Thinking. Harvard Univ. Press, Cambridge, MA, (1984).
[9] Suzuki, A., Short communications. Primates 7, 4, pp.481-487, (1966).
[10]Lawick-Goodall, J., van, Animal Behaviour Monographs 1, 3, pp.161-311, (1968).
[11]Lawick-Goodall, J., Tool-using in Primates and other Vertebrates. In Lehrman, D.S., Hinde, R.A., and Shaw, E. (eds.) Advance in the study of Behavior. New York and London, Academic Press, pp.195-249, (1970).
[12]Kohler, W., Intelligensprufungen an Menschenaffen, Berlin, (1921).
[13]Kohts, N., Sci. Mem. Mus. Darwinianum, 3, (1935).
[14]Kripke, S., Wittgenstein on rules and private language, Blackwell, Basil, New York, (1982)
[15]Sudd, J.H., Discovery 6, pp.15-19, (1963).
[16]Holldobler, B., Stanton, R.C., Markl, H., Behav. Ecol. Sociobiol. 4, pp.163-181, (1978).

[17]Markl, H. & Holldobler, B., Behav. Ecol. Sociobiol. 4, pp.183-216, (1978).
[18]Franks, N.R., Behav. Ecol. Sociobiol. 18, pp.425-429, (1986).
[19]Aerends, G.P., Jur. Tijd. Voor Entomol. 84, pp.72-275, (1941).
[20]Russel, B., Principles of mathematics, Cambridge University Press, 1903, 2d ed., (1937).
[21]Gunji, Y.P. & Toyoda, S., Physica D 101, pp.27-54, (1997).
[22]Gunji, Y.P., Applied Math. and Comp. 47, pp.267-288, (1992).
[23]Gunji, Y.P., Biosystem. 35, pp.33-62, (1995).
[24]Gunji, Y.P., Kusunoki, Y., Chaos, Solitons & Fractals 8, 10, pp.1623-1630, (1997).
[25]Kitabayashi, N. & Gunji, Y.P., Biology Forum 90, pp.393-422, (1997).

This page is intentionally left blank

Chapter 9

The Neurobiology of Semantics: How Can Machines be Designed to Have Meanings?

Walter J Freeman
University of California at Berkeley

Abstract

That branch of semiotics called semantics deals with the relationships between meanings and representations. In my view meanings exist only in brains, which have no representations in them. A meaning is the focus of an activity pattern that may occupy the entire available brain. It is constructed by intentional action, that is followed by learning from the consequences of that action. Communication between brains requires that meanings be represented by construction of words, gestures, symbolse, etc., which elicit meanings in other brains. A representation, is a material object or process, that has no meaning in itself. EEG data indicate that meaning is carried by spatiotemporal patterns of neural activity in frames like a motion picture. The discrete steps occur by cortical phase transitions in the 2-D arrays of neurons interacting synaptically to form wave packets. The rapid exchanges of discrete wave packets between interactive cortical domains generate self-organized dynamics controlling behavior including making representations of meaning. The dynamics of neural arrays is described by sets of differential equations, leading to the possibility of constructing intelligent machines that have the capacity to generate and represent meanings that are comparable to those existing in small animals in machines currently under study in situated robotics.

Keywords : brain dynamics, communication of meaning, intentional action, meaning, phase transitions, representation, semantics (semiotics), situated robotics, wave packets

9.1 Introduction

Biologists studying vocalizations of animals ask not only what the properties are, but what do they mean to other animals? They infer that it represents some central state of intent in the caller that is designed to elicit actions in others. No one can know precisely what the central states are in the transmitter and receivers, but the behaviors that depend on vocal communication suffice to describe

the dynamics of the social system in which the animals are embedded, and the role that is played by the communication of meaning by representations. The lesson is that the calls in themselves contain no meaning, even though they represent meaning with the intent to communicate meaning by evoking formation of comparable meanings in recipients.

Engineers who want to make semantic machines are faced with the task of defining meaning, which at present exists only in brains, and then with the task of learning how to build machines that make meaning [24], [12], [3]. The requirements on network models to simulate the chaotic dynamics of brains include global though sparse connectivity, continuous time dynamics, and distributed spatial functions in two-dimensional arrays of nonlinear integrators. Analog hardware may suffice to emulate the biological functions of sensory cortex in brains by use of nonlinear differential equations [4]. Digital computers serve for parameter optimization [2], [21], [8], but numerical instabilities are only partially overcome by use of noise [9]. A step toward machine intelligence may be to use a model of a sensory cortex as an interface between the unconstrained real world, which is infinitely complex, and the finite state automata that constitute the main support for most artificial intelligence. That is, models from brain dynamics can provide eyes and ears for conventional computers.

However, this step will require that a major problem be addressed: the relation between representation and meaning in brain function. Shannon-Weaver information theory, which is representational, has divorced meaning from information and therefore does not apply to brains. The aim of my presentation is to sketch some of the principal elements of the problem, as a basis for discussing some possible pathways toward solutions through a better understanding of the biological basis of meaning as relations between brain states and behavioral actions, not between symbols in syntactical systems.

9.2 Communication by Representations

Operational discreteness is essential for communication in dialogue. A pair of brains can act, sense, and construct in alternation with respect to each other, not merely as dogs sniff, but as two humans speak, listen, and hear. Consider brains **A** and **B** interacting (Fig. 9.1), where **A**-**B** are parent-child, wife-husband, rabbit-dog, philosopher-biologist, neuroscientist-rabbit, etc. **A** has a thought that constitutes some meaning M(a). In accordance with this meaning **A** acts to shape a bit of matter in the world (a trace of ink on paper, a vibration of air, a

set of keystrokes on e-mail, movements of the face, etc.) to create a representation (a sign or symbol for humans, merely a sign for animals) directed at **B**, R(ab). **B** is impacted by this shaped matter and is induced by thought to create a meaning M(b). So **B** acts to shape a bit of matter in accordance with M(b) in a representation R(ba), which impacts on **A** to induce M(a+1).

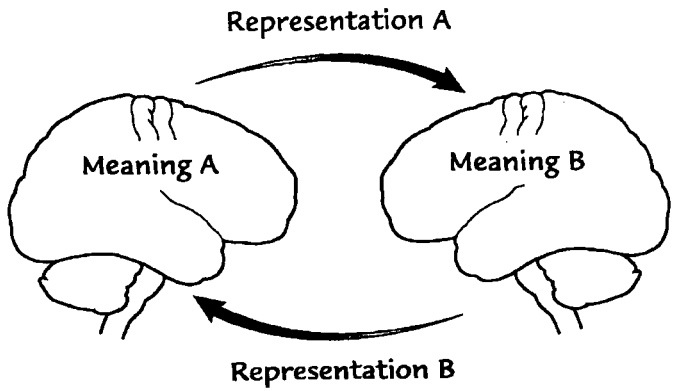


Fig. 9.1 Meaning and Representation.

And so on. Already by this description there is implicit recognition of a discrete ebb and flow of conversation like recurrence of tides, so that meanings M(i)'s as constructions of thoughts become the internal active states, and the R(ij)'s as attributes of matter become the external representations. By its nature an external "re"presentation can be used over and over. It cannot be said to contain or carry meaning, since the meanings are located uniquely inside **A** and **B** and not between them. Moreover, the same R's induces different meanings M(i) in other subjects **C** who intercept the representations. The objects that are used to communicate are shaped by meanings that are constructed in **A** and **B** iteratively and induce the constructions of meaning in **B** and **A** alternately. If communication is successful, then the internal meanings will come transiently into harmony, as manifested by cooperative behavior such as dancing, walking in step, shaking hands, exchanging bread, etc. Symbols can persist like books and stone tablets, while minds fluctuate and evolve until they die.

9.3 Observations of Electroencephalograms

A biological approach to the problem of meaning is to study the evolution of minds and brains, on the premiss that animals have minds that are prototypic of our own, and that their brains and behaviors tell us what essential properties are common to their minds and to our own minds.

Experimental measurements of brain activity (EEG) that follows sensory stimulation of animals show that sensory cortices engage in construction of activity patterns in response to stimuli [5]. The operations are not those of filter, storage, retrieval, or correlation mechanisms. Each construction is by a state transition, in which a sensory cortex switches abruptly from one basin of attraction to another, thereby changing one spatial pattern instantly to another like frames in a cinema. The transitions in the primary sensory cortices, visual, auditory, somatic and olfactory [1], are shaped by interactions with the limbic system, which establish multimodal unity, selective attention, and the intentional nature of percepts. The interactions of the several sensory cortices and the limbic system occur in conjunction with goal-directed actions in time and space. Each cortical state transition involves synaptic changes constituting learning throughout the forebrain, so that cumulatively a unified and global trajectory is formed by each brain over its lifetime. Each spatial pattern appears to reflect the entire content of past and present experience [22], that is, a meaning.

The most important experimental finding is that the neuroactivity patterns in sensory cortex, which are correctly classified on perception of conditioned stimuli by the animals, are not invariant with respect to the unchanging physicochemical stimuli. The brain activity patterns are found to change slightly but significantly with any change in the significance of the stimuli, such as by changing the reinforcement, or adding new stimuli [6]. From numerous tests of this kind the conclusion is that brain patterns reflect the value and significance of the stimuli for the animals, not a fixed memory store. Each pattern formed in response to the presentation of a stimulus is freshly constructed by chaotic dynamics in the sensory cortex, in cooperation with input from the limbic system enacting processes of attention and intention, and it expresses the history and existing state of the animal as much as or more than the actual incident stimulus. The patterns cannot be representations of stimuli or of meanings of stimuli. They are active states induced by stimuli, constituting evolution of the brains in growth of experience [18].

9.4 The Neural Basis for Intentional Action

The making of a representation is an intentional action. All intentional actions begin with the construction of patterns of neural activity in the limbic system, which has been shown by use of lesions and by comparative neuroanatomy and behavior to be a product of the limbic system [13], [20], [7]. In mammals all sensory input is delivered to the entorhinal cortex, which is the main source of input to the hippocampus, and the main target of hippocampal output (Fig. 9.2). Goal-directed action must take place in time and space, and the requisite organ for these matrices is the hippocampus with its 'short term memory' and 'cognitive map' [17].

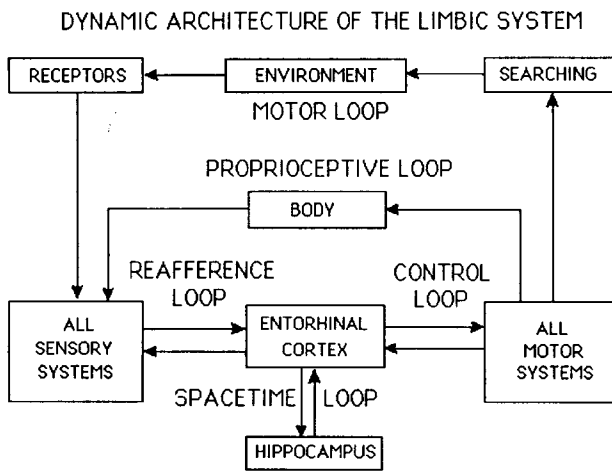


Fig. 9.2 Dynamic Architecture of the Limbic System.

Emergent patterns impact the brain stem and spinal cord, leading to searching movements adapted to the immediately surrounding world. Feedback from the muscles and joints provides confirmation that intended actions are taking place. The impact of movements of the body on sensory input is conveyed to the visual, auditory and olfactory systems. All of these perceptual constructs, that are triggered by sensory stimuli and are dependent on prior learning, are sent to the

entorhinal cortex, where they are combined. When an animal detects an odor of food, it must hold it, move,take another sniff, and decide which way to move next. The difference in strength has no meaning, unless the animal records which way it moved, when the samples were taken,and maintains records for determining distance and direction in its environment. These basic operations of intentional behavior are properties of the limbic system. The same requirements hold for all distance receptors, so it is clearly understandable that evolution has provided for multimodal sensory convergence in order to perform space-time integration on the Gestalt, not on its components.

In the description thus far the flow of neural activity is counterclockwise through proprioceptive and exteroceptive loops outside the brain. Within the brain there is a clockwise flow of activity constituting reafference. When a motor act is initiated by activity descending into the brainstem and spinal cord, the same or a similar activity pattern is sent to all of the sensory systems by the entorhinal cortex, which prepares them for the impact of the movements of the body and, most importantly, sensitizes them by shaping their attractor landscapes to respond quite selectively to stimuli that are appropriate for the goal toward which the action has been directed. These reafferent patterns have been denoted as the senseof effort [11], reafferent signals [25], efference copies [23], and preafference [14], [15].

9.5 Linear versus Circular Causality

Reafference holds the key to attention. The conventional view of sensory cortical function holds that stimuli activate receptors, which transmit to sensory cortex through a linear causal chain, with the eventual outcome of a motor response to the initiating stimulus. Modeling with nonlinear dynamics shows that the stimulus is typically not the initiating event. Rather it is the search for the stimulus that arises in the limbic system in a recurrent manner from prior search and its results. This is circular causality at the level of intentional behavior [16].

Much lower in the hierarchy of brain organization is the event in the primary sensory cortex, which consists of the destabilization of a macroscopic state by the introduction of microscopic sensory input. In this case the transition from a prior basin of attraction to a new one, which has been facilitated by limbic modulation, is guided by the sensory input that activates a learned nerve cell assembly comprising a small subset of cortical neurons. The transition to a new state is global, so that this causal chain is also circular. The stimulus-dependent

neural activity of a few neurons triggers the state transition, and then the entire domain of the primary sensory cortex transits to another pattern, which in the words of [10] "enslaves" the whole set of cortical neurons by acting as an "order parameter". This new active state has been characterized by [19] as a "dissipative structure", that constitutes the "emergence of order out of chaos".

The similarity of the properties of neural activity in the various parts of the limbic system to those in the primary sensory cortices [14], [15] indicates that populations of neurons there also maintain global attractors, which are accessed by nonlinear state transitions, and which are responsible for the genesis of motor patterns controlling goal-directed actions and of reafference patterns that prepare the sensory cortices for the consequences of those actions.

9.6 A Hypothesis on the Causal Relations of Meanings and Representations

The idea is proposed that representations are formed by the forward, counterclockwise flow of neural activity, which emerges from a microscopic level by the interactions of neurons and neuronal populations, and which places the motor systems of the brainstem and spinal cord into appropriate basins of attraction, thereby changing the sensory inflow in a goal-directed manner. The making of a representation is an ordering of the neural control systems of the musculoskeletal apparatus, that is aimed to elicit sensory feedback of a certain kind, namely the patterns of receptor discharge from representational stimuli transmitted by other beings, that place the sensory cortices into the expected basins of attraction. The meaning of the representation is implicit in the form that is given to the representation by the limbic system.

The clockwise backflow of neural activity serves as an order parameter to modulate and shape the neural activity patterns of the sensory cortices, which transmit the states of their neural populations before and after the expected inputs have occurred, and also if they do not occur as expected, or at all. It comprises not only the exteroceptive input but the proprioceptive feedback as well. This global active state, enslaving alike the limbic system and the primary sensory cortices, shapes the meaning not only of the unified sensory input consequent to the transmitted representation, but also of the emitted representation. The implication here is that the agent who is constructing and transmitting the representation cannot fully know its meaning until after the immediate consequences have been delivered through his or her own sensory systems. More generally, a poet, painter, or scientist cannot know the meaning of his or her

creation until after the act has been registered as an act of the self, nor even until the the listeners and viewers have responded with reciprocal representations of their own, each with meaning unique to the recipients.

9.7 Conclusion

Why do brains work this way? Animals and humans survive and flourish in an infinitely complex world despite having finite brains. Their mode of coping is to construct hypotheses in the form of neural activity patterns and test them by movements into the environment. All that can be known is that which has been constructed, tested, and either accepted or rejected [18], [16]. The same limitation is currently encountered in the failure of machines to function in environments that are not circumscribed and reduced in complexity from the real world. Truly flexible and adaptive intelligence operating in realistic environments cannot flourish without meaning.

This global state variable may be regarded as comparable to the operator in a thermostat, that instantiates the difference between the sensed temperature and a set point. The machine state variable has little history and no capacities for learning or determining its own set point, but the principle is the same: the internal state is a form of energy, an operator, a predictor of the future, and a carrier of information that is available to the system as a whole. The feedback device is a prototype, an evolutionary precursor.

Acknowledgements

This work was supported by research grants from the National Institute of Mental Health MH06686 and the Office of Naval Research N00014-90-J-4054.

References

[1] Barrie, J.M., Freeman, W.J., Lenhart M, Modulation by discriminative training of spatial patterns of gamma EEG amplitude and phase in neocortex of rabbits. Journal of Neurophysiology 76, pp.520-539, (1996).

[2] Chang, H.J., Freeman, W.J., Parameter optimization in models of the olfactory

system. Neural Networks 9, pp.1-14, (1996).

[3] Clark, A., Being There. Putting Brain, Body, and World Together Again. Cambridge MA: MIT Press, (1996).

[4] Eisenberg, J., Freeman, W.J., Burke, B., Hardware architecture of a neural network model simulating pattern recognition by the olfactory bulb. Neural Networks 2, pp.315-325, (1989).

[5] Freeman, W.J., Mass Action in the Nervous System. New York: Academic, (1975).

[6] Freeman, W.J., Tutorial in Neurobiology. International Journal of Bifurcation & Chaos 2, pp.451-482, (1992).

[7] Freeman, W.J., Societies of Brains. Hillsdale NJ, Lawrence Erlbaum Assoc, (1995).

[8] Freeman, W.J., Random activity at the microscopic neural level in cortex ("noise") sustains and is regulated by low-dimensional dynamics of macroscopic cortical activity ("chaos"). International Journal of Neural Systems 7, pp.473-480, (1996).

[9] Freeman, W.J., Chang, H-J., Burke, B.C., Rose PA, Badler J, Taming chaos: Stabilization of aperiodic attractors by noise. IEEE Transactions on Circuits and Systems 44, pp.989-996, (1997).

[10]Haken, H., Synergetics: An Introduction. Berlin: Springer, (1983).

[11]Helmholtz, H. von, Treatise on Physiological Optics: Vol. 3. The Perceptions of Vision (JPC Southall, Trans.). Rochester NY: Optical Society of America, (1879/1925).

[12]Hendricks-Jansen, H., Catching Ourselves in the Act. Situated Activity, Interactive Emergence, and Human Thought. Cambridge MA: MIT Press, (1996).

[13]Herrick, C.J., The Brain of the Tiger Salamander. Chicago IL: University of Chicago Press, (1948).

[14]Kay, L.M., Freeman, W.J., Lancaster, L., Simultaneous EEG recordings from olfactory and limbic brain structures: Limbic markers during olfactory perception. Ch. 6 in: Gath, I., Inbar, G.F. (eds.) Information Processing and Pattern Analysis of Biological Signals. New York: Plenum Press, pp.71-84, (1996).

[15]Kay, L.M., Lancaster, L., Freeman, W.J., Reafference and attractors in the olfactory system during odor recognition. Intern. Journal of Neural Systems 7, pp.489-495, (1996).

[16]Merleau-Ponty, M., The Structure of Behavior [AL Fischer, Trans.]. Boston: Beacon Press, (1942/1963).

[17]O'Keefe, J. & Nadel, L., The Hippocampus as a Cognitive Map. Oxford UK: Clarendon, (1978).

[18]Piaget, J., The child's conception of physical causality. New York: Harcourt, Brace, (1930).

[19]Prigogine, I., From Being to Becoming: Time and Complexity in the Physical Sci-

ences. San Francisco: Freeman, (1980).
[20]Roth, G., Visual Behavior in Salamanders. Berlin: Springer-Verlag, (1987).
[21]Shimoide, K., Freeman, W.J., Dynamic neural network derived from the olfactory system with examples of applications. IEICE Transaction Fundamentals E-78A, pp.869-884, (1995).
[22]Skarda, C.A., Freeman, W.J., How brains make chaos in order to make sense of the world. Behav. & Brain Sci. 10, pp.161-195, (1987).
[23]Sperry, R.W., Neural basis of the spontaneous optokinetic response. Journal of Comparative Physiology 43, pp.482-489, (1950).
[24]Tani, J., Model-based learning for mobile robot navigation from the dynamical systems perspective. IEEE Transactions on Systems, Man & Cybernetics 26, pp.421-436, (1996).
[25]Holst, E. von & Mittelstaedt, H., Das Reafferenzprinzip (Wechselwirkung zwischen Zentralnervensystem und Peripherie). Naturwissenschaften 37, pp.464-476, (1950).

Chapter 10

The Emergence of Contentful Experience

Mark H. Bickhard
Lehigh University

Abstract

There are many facets to mental life and mental experience. In this chapter, I attempt to account for some central characteristics among those facets. I argue that normative function and representation are emergent in particular forms of the self-maintenance of far from thermodynamic equilibrium systems in their essential far-from-equilibrium conditions. The nature of representation that is thereby modeled — an interactive, pragmatic form — in turn, forces a number of additional properties of mental process, such as consciousness being inherently contentful and from a situated and embodied point of view. In addition, other properties of interactive representation make strong connections with the central nervous system properties that are found to realize mental experience, such as a field organization of oscillatory and mutually modulatory neural processes.

Keywords : representation, cognition, representational content, consciousness, normative function, Dretske, Fodor, Millikan, mental experience, situated cognition, embodiment, interactivism, far from equilibrium systems, Piaget, goal directedness, pragmatics, information semantics, encodingism, asymmetric dependence, timing, Turing machines, central nervous system, volume transmitters, silent neurons, modulatory processes

10.1 Introduction

There are many facets to mental life and mental experience. Ultimately all of them must be addressed in the overall task of the naturalization of mind. Here I will focus primarily on three aspects of basic consciousness. In particular, basic conscious experience:

1. is a process,
2. that is contentful,
3. from a point of view.

Additional characteristics of mind, such as embodiment and a convergence with functional properties of the central nervous system, emerge in the course of the

main line of discussion.[1]

Organisms are inherently far from thermodynamic equilibrium; to go to equilibrium is to die. Work must be done in order to maintain the essential far-from-equilibrium conditions, and it must be done in ways and at times that are appropriate to the relevant environmental conditions. Even very simple living systems can exhibit this function of selecting "what to do next": some bacteria, for example, can swim if they are swimming up a sugar gradient, but tumble if they are swimming down a sugar gradient [22, 23]. Together, these interactions with the environment tend to increase the sugar supply available to the system.

I will argue that representation has emerged in the evolutionary answers to such problems of selecting "what to do next", and that several aspects of both mental experience and central nervous system processing are accounted for by that answer.

10.2 Function

The first step in the discussion is a model of the nature and emergence of normative function — function as distinguished from dysfunction. For current purposes, a brief outline of this model will suffice.

Some far-from-equilibrium systems, insofar as they are stable through time at all, depend on external support to maintain that stability. A chemical bath, for example, may be maintained in some far-from-equilibrium condition by pumping various solutions into it, and the maintenance of this activity, in turn, depends on the pumps continuing to work and receive power, and the reservoirs of those solutions remaining full. Some far-from-equilibrium systems, on the other hand, make contributions to their own stability. A candle flame, for example, maintains above combustion threshold temperatures, and, in standard atmospheric and gravitational conditions, induces convection, which brings in fresh oxygen and removes combustion wastes. Far-from-equilibrium systems that make such contributions are, in that sense, *self maintenant* systems [7].

Such contributions to the maintenance of relevant far-from-equilibrium conditions are *functional* for that system [7]. Conversely, to fail in making such contributions is *dysfunctional* for that system. Functionality, in this sense, is relative to a particular system as reference point: a heart in a parasite may be

[1] Several other aspects — such as perception, motivation, language, development, rationality, sociality, personality, and so on — have been addressed elsewhere [2, 4, 5, 6, 8, 9, 14, 16, 17, 18, 24, 25, 26].

functional for the parasite but dysfunctional for the parasitized host.

10.2.1 *Etiological Approaches to Function*

This model of function is in contrast to standard etiological approaches [43,59,60]. The central notion in these approaches is that the heart has a function of pumping blood, instead of, say, making heart beat sounds, because it is the evolutionary descendant of prior hearts that were selected for pumping blood, not for making heart beat sounds. A kidney, then, that does not filter blood is not serving the function that it has — is being dysfunctional — since kidneys in general have the function of filtering blood.

Etiological approaches to function model the having of a function as being constituted in having the right kind of evolutionary history. This has a sometimes counterintuitive consequence: if, for example, a lion were to miraculously pop into existence that was molecule for molecule identical to some lion in the zoo, the science fiction example lion would have no functions for any of its organs, because none of them would have the right kind of evolutionary history.[2] They have, in fact, no evolutionary history at all. Millikan is willing to accept this consequence [59], but although such counterintuitive consequences for purely science fiction thought experiments may be worth accepting if other successes of the model warrant, this example points to a far deeper problem — one that is, I argue, fatal to all such approaches.

In particular, the lion example exemplifies that function, on the etiological account, cannot be defined in terms of the current state of the system. Two systems can be in identical states, such as the two lions, but one of them will have organs with functions and the other not, depending on their histories. But, physics tells us, only the current state of a system can have causal efficacy. Etiological accounts, then, at best provide an epiphenomenal account of function — an account with no causal importance in the world. That is not a successful naturalization of the notion of function.

Note, in contrast, that function understood in terms of contributions to maintaining relevant far-from-equilibrium conditions is a current state definition. It does make a causal difference whether or not this flame or that organism

[2] This is from a discussion by Millikan [59]. The idea would be, for example, if the atoms in the air were to suddenly converge in such a way that they formed a lion. This, of course, is statistically impossible, even though logically possible. Millikan uses the example simply to demonstrate what she claims is a counter-intuitive, but nevertheless acceptable, consequence of the historical approach to function. I argue that there is a deeper and more important issue at stake here.

remains in far-from-equilibrium conditions. Function, then, emerges in self maintenant far-from-equilibrium systems, including in particular living systems.

10.3 Representation

Self maintenant systems make contributions to their own maintenance, but those contributions are fixed. There is no ability to change to making different kinds of contributions if the environment were to change so that some such change in self-maintenant contributions would be appropriate. Candle flames, for example, cannot shift into a "hunt for fuel" mode when the candle is getting low.

The bacterium, however, *can* make such shifts. Swimming if moving up a sugar gradient but tumbling if moving down a sugar gradient is precisely to do different things in different circumstances so as to contribute to far-from-equilibrium maintenance in ways appropriate to those changing conditions. Such systems tend to maintain their condition of being self maintenant — they are, in that sense, *recursively self maintenant* [7].

The key point to note is that such selections on the part of a recursively self maintenant system are anticipatory in nature, and that, as such, they can be in error. They are anticipatory in that they anticipate that the consequences of engaging in the selected activity, under these conditions, will in fact serve the function of self-maintenance. They can be in error because such anticipations depend on, among other things, the environment, and the environment may not cooperate. The bacterium will swim up a saccharin gradient just as readily as it will swim up a sugar gradient.

This is not a standard usage of "anticipate" because it is not meant in any necessary sense of deliberate or explicit anticipation. It is, instead, a functional sense of anticipate. Some functions — contributions to self maintenance — depend for the success of their functional contributions on particular things working out, or being the case, in the future as the functional process proceeds. To indicate that such a functional process will be appropriate, or to initiate such a functional process, then, functionally or implicitly anticipates that those necessary supporting conditions will obtain.

Such anticipations constitute the most primitive emergence of representational truth value: There is, first of all, a truth value in the anticipation itself — it is either correct that the activity will be self maintaining or it is not. Second, that truth value is about the environment: the anticipation constitutes an im-

plicit predication about the environment, viz., this is an environment in which the selected activity is appropriate. And third, it has representational content: the anticipation implicitly defines whatever those environmental properties are that would support the selected activity being successful toward self-maintenance. This is implicit definition in a dynamic generalization of the sense in which a set of axioms implicitly defines the class of models for those axioms [12, 48, 55]. So, there emerges content, which is about the environment, and which has truth value; this is representation, however primitive.[3]

10.3.1 *Evolutionary Elaborations*

Such representation, however, is quite primitive. It fits, perhaps, bacteria or paramecia, but what about more complex representation, such as in human beings? I will turn to several ways in which primitive representation can be elaborated, each such elaboration improving the adaptability of the organism.

First, notice that the "selection" of what to do next in the bacterium is a kind of triggering. Under specified conditions — conditions that normally detect sugar gradients — do X, swim perhaps, or do Y, tumble perhaps. Under more complex conditions, there may be more than one potentially appropriate next interaction, and a selection *within* some set of possibilities must be made.

One basic manner in which this more complex kind of selection can be accomplished involves three interrelated innovations beyond the triggering model. First, the relationship to potentially appropriate next interactions must be some sort of *indicative* or *pointer* relationship, not a simple triggering. Second, there must be some basis for making a selection within a set of such indicated potentialities. In general, that basis will involve information about the anticipated *outcomes* of the indicated interactions, should they be selected. That is, choose the interaction on the basis of its expectable outcomes.

But those outcomes, at least in the logically primitive sense, cannot be *represented* outcomes, on pain of a circularity in the basic model of representation. If, however, they are internal outcomes, internal states, perhaps, that are indicated in association with various interaction possibilities, then that information is functionally available inside the system, and does not require a circularity of modeling representation in terms of representation.[4]

---

[3] This is a pragmatist model of representation, rather than the standard encoding or empiricist models (e.g., [50, 65, 69]).

[4] Such a circularity will yield an infinite regress if the circle is followed in an attempt to find some foundational level that breaks out of the circle. Since there is no such level, the unboundedness of

Third, there must be some process for using such outcome indication information in the service of selecting next interactions from among those indicated. The basic process architecture within which this can take place is a goal directed system, that selects interactions from among those indicated on the basis of their fit to a current goal. (Goal directedness, however, can also involve architectures that are much less explicit: [18].)

Again, however, a potential circularity threatens. If goal conditions must themselves be represented, then again the model of representation has made necessary use of representation. Goal conditions, however, do not have to be represented in order to be functional (though clearly they can be so represented once representation as a function is already available). Goal conditions need only be *detected.* A goal of raising blood sugar, for example, need only yield a continuation of potentially appropriate activities so long as blood sugar is in fact below some threshold. No representation of blood sugar level is necessary. The bimetallic strip in the classic thermostat example does not represent temperature, but it does detect it; and the set point in the thermostat similarly does not represent temperature, but it does detect when the actual temperature has reached the set point temperature. Such functional relationships of detection are all that are necessary for goal directedness, so this potential circularity too is avoided.

So, the first evolutionary elaboration beyond simple triggerings of activities is the evolution of the ability to make use of information about interaction outcomes in the selection of next interactions. Note that such indications of anticipated outcomes not only make possible the selection of next interactions in a way much more sophisticated than simple triggering, they also permit the system to detect whether or not those indicated outcomes are in fact obtained — they permit the system to detect the truth value of its (still primitive) representations. Such system detectable error, in turn, can be quite useful in guiding further behavior, and is essential for error guided learning [15, 18].[5] Indicated outcomes, then, ground the task solutions for both interaction selection and interaction evaluation.

Goal directed processes are an important elaboration of basic triggered system activities. Another important development occurs with respect to the con-

---

the regress follows. Even prior to generating such a regress, however, such a definitional circularity is unacceptable because defining representation in terms of representation does not contribute to the task of understanding representation.

[5] One aspect of the emergence of such primitive representation is the concomitant emergence of equally primitive motivation [9].

ditions under which various interactive potentialities are indicated — the processes of detection. The most general manner in which such detections can occur is by *interactive differentiation*. If a subsystem engages in interaction with the environment, the internal course of that interaction — and, therefore, the internal outcome of that interaction — will depend in part on the environment being interacted with. Some environments will yield the same internal outcome, while other environments will yield some different internal outcome. The set of possible internal outcomes serves to differentiate the class of possible environments into those that yield outcome A, say, versus those that yield outcome, or final state, B. Such environment differentiations, in turn, can serve as the conditions for further indications of potentiality. Arriving at outcome A, for example, might indicated that interaction Q is possible, while arriving at outcome B might indicate that interaction R and interaction S are both possible.

The set of environments that would yield final state A as outcome are *implicitly defined* by the interaction subsystem that engages in the relevant interaction. As before, this is a dynamic generalization of the sense in which a set of formal sentences implicitly defines its class of models [11, 48, 55]. Differentiation and implicit definition, then, are duals of each other. Final state A of some subsystem implicitly defines A-type environments, and arriving at A differentiates the current environment as being of type A.

An interactive subsystem with possible final states, therefore, is the basic manner in which conditions for indications of potentiality are set up. But the interactive potentialities that are indicated as possible are themselves interactive subsystems with associated possible final states: the two are the same kinds of system organization. Any interactive subsystem, then, will differentiate environments in accordance with its possible final states — actually engaging in the interaction and arriving at one of the final states differentiates the environment as being of the type implicitly defined by that final state — and any interactive subsystem can be indicated as possible if appropriate prior differentiations have occurred.

This suggests the next important elaboration: indications of interactive potentiality can branch and can iterate. A given differentiation can evoke indications of potentiality of multiple further possibilities: final state A might indicate the potentialities of both P and Q. So the indicative relationships can branch. And if P is engaged, arriving, say, at final state D, that might serve to indicate the potentialities of R, S, and T. Such branched and iterated organizations of indications of interactive potentialities can, in more cognitively sophisticated organisms, be quite complex, forming vast webs of potentiality indications.

It is such webs that constitute the basis for more familiar forms of representation, such as of objects, and do so in a generally Piagetian manner (e.g., [2, 62]). The representation of abstractions, such as of electron or the number six, requires still further architectural machinery, but will not be pursued here [23, 24]. The most important properties of interactive representation that I will develop for current purposes are those of temporal and functional continuities, which underlie aspects of both phenomenology and central nervous system functioning.

10.3.2 *Information Semantics*

First, however, a detour to compare the interactive model with the approach to representation that is dominant in contemporary cognitive science: information semantics. Consider an interactive differentiation that takes place with no outputs. This is no longer a full *inter*action, but a passive processing of inputs. When differentiations can be performed in this manner, they are less costly of time and energy, and such forms of differentiation are ubiquitous in complex organisms. One major class of examples is the sensory tracts and associated "information processing" as neural activity progresses along those tracks [27]: the outcomes of such processing, at any level, implicitly define the environments that would yield those outcomes if encountered.

The important point for current purposes is that such passive differentiation processes are the paradigm of what information semantics approaches to modeling representation submit as examples of *representation.* A differentiation, passive or not, does create an informational — and, perhaps, a nomological and causal — relationship with various properties in the environment: those properties that support arriving at that internal state. Information semantics would have those properties be the content of the representation that is constituted by that final state. The states involved in the sensory information processing are said to "encode" the environmental properties that they differentiate. The interactive model, in contrast, does not attribute content to such differentiations. Instead, the differentiations are the content*less* differentiations upon which content*ful* indications of further potentialities may be based.[6]

The comparison being made here is with standard models which attribute representational content to "mere" differentiations, especially passive differentiations, such as in so called "sensory encodings". My claim, in contrast, is that

[6] Note that it is not the interactive *model* that makes contentful indications, but, rather, the organism.

such differentiations, passive or not, do not have any content — they are contentless differentiations. But, such differentiations may serve as the basis for setting up indications of further interactive potentiality, and those indications *can* have content — the content that is implicitly defined in the supporting conditions for those further potentialities. That is, such differentiations may differentiate in fact those kinds of environments in which the indicated interactive potentialities will work. But such a differentiation is not and need not be a representation of whatever the conditions are that will support those indicated interaction potentialities: a detector need not be a representation of what is detected — a differentiator need not be a representation of what is differentiated.

What's wrong with modeling the differentiations themselves as possessing content? This stance is of millennia-long standing. It is a current version of assuming that representation is constituted by correspondences between the representation and what it represents [33, 34, 36, 37, 38, 39, 46, 57, 67]. *External* examples of representation do seem to fit this approach: Morse code, blueprints, maps, ciphers, and so on. They form the basis for the never ending appeal of modeling purported *mental* representation in the same mold. But such external representations require an interpreter to know and interpret the correspondences involved, while mental representation cannot require such an interpreter on pain of a classic infinite regress of interpreters interpreting the results of previous interpretations.

This regress problem is just one of a great many fatal flaws in correspondence approaches to mental representation. I will touch upon only a few of them here (see [7, 18]). One derives from the fact that correspondence, informational, nomological, isomorphic, and causal relationships exist profusely throughout the universe, while at best an extremely small fraction of them might constitute representational relationships. So something further must be specified to attempt to pick out the special such relationships that are supposed to be representational. There is no consensus about what that additional special qualification might be, and I argue that none of them on offer works, and that none *can* work [7, 18]. One perspective on why this is so is to note that, even if some special additional property did succeed in extensionally picking out only those correspondences that are genuinely representational, that would still not constitute a naturalistic model of the nature of the representational content involved for the organism itself. For example, there is one finer differentiation in the class of correspondences that does pick out representational correspondences: those that are genuine encodings, such as Morse code. But genuine

encodings require an interpreter in order to provide those encodings with content. This is not a problem for many purposes, but for the purpose of modeling representation and representational content, it merely pushes the problem off onto understanding and modeling the interpreter, and that was the original task in modeling *mental* representation in the first place.

Another fundamental problem has to do with being able to model the possibility of representational error. The problem arises because, if the special "representational" correspondence — or informational relationship, or lawful relationship, or whatever special kind is picked out by a model — exists, then the representation exists, and it is correct. On the other hand, if that special correspondence does not exist, then the representation does not exist, and therefore it cannot be incorrect. There are multiple attempts to solve this problem, but none that succeeds, and none that even addresses the basic problem of not just the possibility of representational error, but that of *system detectable* representational error.

One such attempt regarding the "simple" possibility of error is that of Jerry Fodor [36, 37, 39, 57]. The central notion of relevance here is that of *asymmetric dependence*. The idea is that the possibilities of false evocations of a representation are asymmetrically dependent on true evocations of that representation, and this asymmetry in the dependence relationships distinguishes true from false possibilities. If, for example, a horse dimly seen on a dark night happens to evoke a representation of a cow, that evocation should somehow be modeled as being false. Fodor's point is that such evocations by horses on dark nights are dependent on evocations by cows in the sense that if cows did not evoke the representation, then horses on dark nights would not either. But the dependency is not reciprocated: if horses on dark nights never evoked the cow representation, that has no bearing on cows evoking the cow representation. The dependency between the two possibilities is asymmetric.

There are a number of problems with this kind of an account. Here is one of them: a counterexample. Consider the docking of a neural transmitter molecule, dopamine, perhaps, in a receptor on a cell surface, triggering internal activities in the cell. This constitutes a causal, nomological, informational correspondence between the transmitter molecule and the cell activities, but there is no representation involved. Still further, consider a poison molecule, crank, perhaps, that can dock on the same receptors and trigger the same internal activities. Again, there are all the kinds of correspondence relationships anyone could want, and, furthermore, there is an asymmetric dependence of the crank possibility on the dopamine possibility, but there is still no representation [7,

56].

Here is another: there is no way for any organism to know about, to be able to determine, what the various asymmetric dependency relations are among its potential evocations of representational elements. Therefore, there is no way for an organism to possess in any relevant sense what the contents are of its own representations — to know what they are supposed to represent. Still further, to detect error in its representations, an organism would have to compare such content (which it does not possess) with the actual entity or property currently being represented — the current contact with the environment [7,18], or target of representation [30] — to determine that they do not fit each other. But representing the current contact, or target, is precisely the original problem of representation all over again. So, system detectable error is simply impossible on this account. Not all representations are in error; not all that are in error are detected as being in error; not all organisms are capable of detecting such error. But system detection of representational error does occur — it underlies error guided behavior and learning — and Fodor's model (along with virtually all others) renders it impossible [7, 13, 18]. They are thereby falsified. Fodor wishes to set aside such issues of the epistemology of representation until the metaphysics of representation is clear [39]. In itself, that is an acceptable strategic move, but Fodor's metaphysics not only does not address the basic problem of representational epistemology, it makes representational epistemology impossible. Fodor's metaphysics is thereby refuted [56].

Representation as some special form of correspondence has an ancient provenance, and many different kinds of issues concerning such approaches and elaborations of such approaches have been addressed over the millennia (e.g., [44, 66, 72]). I will truncate this discussion at this point, however, with having shown that such approaches suffer foundational flaws. The interactive model, note, models the possibility of error and of system detectable error with ease. It requires no interpreter. It is a viable candidate as a model of representation and representational content. I return, then, to the main discussion of elaborating further properties of interactive representational systems.

10.3.3 *Continuities*

I will develop two kinds of continuity involved in interactive representation: a functional continuity and a temporal continuity. Consider again the set of possible final states for a differentiating interactive subsystem. I have provided examples of such sets above, always with only two possible final states —

usually A and B — for simplicity of presentation. But there is nothing that precludes such a differentiating set from being large in cardinality, or even infinite. In fact, differentiating sets with the size of the real numbers should be expected to be common. Such differentiating sets could be realized as, for example, levels of activation of some neural process, or wavelength of some oscillatory process, and so on. Infinite differentiating sets will not set up discrete indications of potentiality for each element, but, instead, will function more as the setting of parameters for further activity in the system that might be engaged in by that system.[7] System activity and control flow in such an architecture will involve a generally smooth process of engaging in current interactions as one aspect of an overall process, of which another aspect will be the exploration and following of smooth manifolds of parameterized indications of further potentiality.

I turn now to another form of continuity in the model. The interactive model is of representation emerging naturally out of action systems: representation offers a joint solution to the problems of action selection and action evaluation. Action and interaction, however, require correct timing in order to be successful. Mere speed is not sufficient: an interaction can fail from being too fast just as easily as from being too slow. Interaction has to be appropriately coordinated, and that includes temporal coordination.

Computationalist models, in contrast, are based on computer models, and, ultimately, on Turing machines. But Turing machines cannot model temporal coordination. They cannot model timing. Turing machines function with respect to a *sequence* of actions, but the timing involved in the sequence is arbitrary. Timing per se makes no difference to the Turing machine properties and is invisible to any possible Turing machine processes. If the first step required ten seconds, the second ten centuries, the third ten nanoseconds, and so on, nothing about the Turing machine per se would be different from any other timing [17].

Actual computers, of course, do involve timing, and, in that sense, go beyond Turing machines per se. But they do so with a central clock driving myriads of lock step processes. This is a viable design architecture, but an impossible evolutionary architecture: every evolutionary change in the central nervous

[7] Note that setting parameters does not, in general, in itself suffice to specify a system process or interaction. Parameters blend with each other in influencing further activity; they do not build together like bricks. Parameters are not interaction units out of which more complex such units might be constructed. Instead, they join and blend like themes of interaction [17]. This suggests that themes should constitute a major aspect of functional processing in a complex interactive system.

system would have to involve simultaneous well-coordinated changes in the processing architecture and in the timing architecture. This is vanishingly improbable even once; it is not possible (in any but a strictly logical sense) for evolutionary time spans of change.

So, the brain does it differently. Put clocks everywhere, and render all functional relationships as relationships among the clocks. This sounds odd when put in terms of clocks, but, if it is recognized that clocks are "just" oscillators, it becomes: make all processes oscillatory and render functional relationships as modulatory relationships among those oscillatory processes. In such an architecture, timing is ubiquitous. It is available anywhere that it is useful, and can be ignored if not. Note that such a framework for an architecture is at least as powerful as a Turing machine: a limit case of one process modulating another is for one process to turn the other on and off, that is, to switch the other on and off. But switches are sufficient for building a Turing machine, so oscillatory and modulatory principles have at least the power of Turing machines. They have, in fact, greater power in that they intrinsically capture timing while Turing machines cannot [17, 18].

Brain processes are commonly modeled in terms of the current technological models available. From switch boards to symbol manipulations to connectionist nets, studies of the central nervous system have tended to follow the technological lead. This yields currently, for example, a dominant model of neurons as threshold elements that fire or not depending on incoming activations and inhibitions. The paradigmatic neuron is the classic dendritic arborization leading to the extended axon, with the cell body as an appendage [27]. Of course, there are other kinds of neurons, but they are left out of the general functional picture of by what principles the brain might work.

Much of what we know about how neurons function, however, is not easily accommodated by such models. A large population of neurons never fires — the so called "silent" neurons [20, 64]. Neurons and neural circuits can exhibit base line oscillatory, or firing, rates, independent of incoming influences [32, 42, 51, 52, 68]. Some neurotransmitters are not restricted to a synaptic cleft, but diffuse throughout a local population of neurons — they are "volume" transmitters [1, 40, 47, 71]. Some neurotransmitter release is not all or none, but is "graded" in accordance with the "not all or none" oscillatory ionic waves reaching the terminal buds [20, 41]. Some neurons influence others via "gap junctions" that involve no neurotransmitter at all [32, 45, 61]. Even the glia seem to be involved in influencing neural activity [47, 73]. And so on. All of this deviation from paradigm must be construed as merely implementational on

standard accounts, though it is not at all clear why evolution would have crafted so many modes of influence if all that was functionally relevant were threshold switches.

But such a tool box of modulatory relationships among oscillatory processes is precisely what would be expected if the functional principles by which central nervous system operated were those of oscillatory processes modulating each others' activity. Gap junctions provide an extremely fast and spatially localized influence. Traditional synapses are slower and less localized. Volume transmitters are much slower and affect significant local populations. Silent neurons don't have to fire in order to modulate other activity. And so on. The interactive model puts timing at the center of any interactive system's functioning, and timing puts oscillatory and modulatory relationships at the center of the processing architecture of such an interactive system. And the central nervous system manifests multiple properties that are perplexing and at best superfluous on standard views, but are simply an evolutionary toolbox for modulatory relationships from the perspective of the interactive model.

Processes in a complex interactive system, then, can be expected to manifest at least two forms of continuity: functional and temporal. Mental processes that might be emergent in such processes, therefore, should be expected to manifest similar continuities.

10.4 Brain and Mind: Some Relations

Mental life is a process. It is a process that is inherently contentful: it involves intentionality or "aboutness". The interactive model generates a model of that process as having an ongoing execution of interaction as one aspect and an ongoing consideration of further potentialities as another aspect.[8] But the "consideration" of further process potentialities *is* the consideration of representational content. It is the consideration of the contents involved in those anticipations of further potentialities. The interactive model, then, captures mentality as a contentful process.

Mental process involves continuity in both functional and temporal aspects. Oscillatory processes continuously distributed throughout the central nervous system will manifest the properties of an oscillatory *field*. Mental process, then,

[8] This aspect is elsewhere called microgenesis. Microgenesis itself offers a powerful model both for characteristics of central nervous system functioning and for important cognitive capabilities, such as metaphor and heuristic problem solving [15].

should be emergent in fields of processes in the brain. That is, consciousness, at least in its most basic form, should be emergent in central nervous system processes organized as fields [53, 54]. Mental life manifests properties of this field organization in levels of activity of the field, fineness of differentiations engaged in, coherence (or lack thereof) of the contents being processed, and truncations of experience corresponding to truncations of field processes, such as in cases of neglect [53, 54].

Content in this model is always grounded in differentiation processes and possibilities. Differentiations are inherently indexical and deictic. They are relative to the organism making those differentiations in several senses:

1. They are differentiations that, insofar as they are spatial, are spatial in body centered coordinates — they are differentiations *produced* by interactions that that body engages in, and for the subsequent potential use in the interactions that that body engages in. For example, the toy block is just in front of me. Less indexical location representation requires more sophisticated elaborations of invariance representations. The toy block is behind me, or in my room.
2. They are differentiations only as fine as the organism is capable of making and has found to be useful in further processing. Frogs, for example, typically do not differentiate narrowly enough to distinguish flies from small pebbles tossed in front of them. Frogs have not much needed finer differentiation in their evolutionary history. On the basis of such differentiations, frogs will process the potentiality of tongue flicking and eating[9], along with other relevant possibilities should they exist, such as mating or the potentialities indicated by differentiating the shadow of a hawk overhead.

Mental life, then, is from *a point of view*, both spatially and functionally. Mental life arises in the framework of the view of the organism on all of its further potentialities, spatial, interactive, goals, values, and so on.[10] Mental life is from a point of view most fundamentally because content is from a point of view. The context independent notion of encoded content is a myth. It is impossible because mental representation cannot fundamentally be constituted as encodings. Achievement of relative context independence, of greater scope of invariance, *is* an achievement, on both an individual as well as a cultural level — in science, for example [49].

---

[9] Note that the frog's content is that of tongue flicking and eating, not that of "fly" or "pebble" or "fly or pebble" [15].

[10] See Campbell & Bickhard [24] for a model of the emergence of values within interactive systems.

That is, mental life is inherently *situated.* It is relative to the situation of the organism, again most fundamentally because content is situated. Similarly, mentality is *embodied.* Interaction cannot take place except by some body or another. Mentality is not possible in an inherently passive system — such as a computer that only processes inputs. Mental point of view, then, is situated in the entire representable realm of its further interactive potentialities; it is situated spatially and functionally and relative to the embodiment in which that mental process is taking place.

10.5 Conclusions

Mental life is a process that is inherently contentful, inherently embodied, and inherently from a situated point of view. The interactive model accounts for these properties as intrinsic aspects of interactive processes. In fact, once the relevant aspects of the interactive model are elaborated, the emergence of these corresponding aspects of mentality is automatic and completely natural.

The interactive model also accounts for otherwise puzzling characteristics of the central nervous system processes in which mind is emergent. In particular, the field characteristics of functional and temporal continuity, and the underlying biochemical level of oscillatory processes engaged in mutual modulations, together with the elaborate neural modulatory tool kit, are also automatic and completely natural from the interactive perspective.

The interactive model, thus, accounts in a very natural way for multiple properties of both mind and brain. There are, of course, important characteristics not addressed here, such as those of qualia, emotions, reflexivity, and others,[11],[12] but the naturalness with which the interactive model connects with

---

[11] The vast and rapidly growing recent literature addressing the phenomena of consciousness includes: Block, Flanagan, Güzeldere [19], Chalmers [28], Cohen & Schooler [29], Dennett [31], Flanagan [35], Marcel & Bisiach [58], Revonso & Kamppinen [63], and Tye [70].

[12] Mind is not emergent in all of its properties at once from underlying functional and physicochemical processes. This is evident, for example, from a consideration of evolution and non-human animals: not all animals are capable of reflective consciousness; not all are capable of emotions; not all are capable of learning. Necessarily, then, at least these properties must be differentiable from mind in its simplest form. Nevertheless, there is still a strong vestige on the contemporary scene of Cartesian dualism, not in an explicit dualism per se, but in the presupposition that mind differs from the non-mental in some kind of singular gulf [12]. Instead, mind seems to have evolved through a complex trajectory, involving learning, perception, emotions, reflective consciousness, and so on. If so, then these mental phenomena must be modeled as emergent in evolutionary elaborations of simple mental awareness [3, 24].

multifarious properties of both phenomenology and brain processes encourages exploration of further mental characteristics within the interactive framework.

References

[1] Agnati, L.F., Fuxe, K., Pich, E.M., Zoli, M., Zini, I., Benfenati, F., Härfstrand, A. and Goldstein, M., Aspects on the Integrative Capabilities of the Central Nervous System: Evidence for 'Volume Transmission' and its Possible Relevance for Receptor-Receptor Interactions. In Receptor-Receptor Interactions, ed. by K. Fuxe and L. F. Agnati, Plenum, New York, pp.236-249, (1987).

[2] Bickhard, M.H., Cognition, Convention, and Communication, Praeger Publishers, New York, (1980).

[3] Bickhard, M.H., A Model of Developmental and Psychological Processes. Genetic Psychology Monographs 102, pp.61-116, (1980).

[4] Bickhard, M.H., The Social Nature of the Functional Nature of Language. In Social and Functional Approaches to Language and Thought, ed. by Maya Hickmann, Academic, New York, (1987).

[5] Bickhard, M.H., A Pre-Logical Model of Rationality. In Epistemological Foundations of Mathematical Experience, ed. by Les Steffe, Springer-Verlag, New York, pp.68-77, (1991).

[6] Bickhard, M.H., How Does the Environment Affect the Person? In Children's Development within Social Contexts: Metatheory and Theory, ed. by L. T. Winegar and J. Valsiner, Erlbaum, Mahwah, NJ, pp.63-92, (1992).

[7] Bickhard, M.H., Representational Content in Humans and Machines. Journal of Experimental and Theoretical Artificial Intelligence, 5, pp.285-333, (1993).

[8] Bickhard, M.H., Intrinsic Constraints on Language: Grammar and Hermeneutics. Journal of Pragmatics, 23, pp.541-554, (1995).

[9] Bickhard, M.H., Is Cognition an Autonomous Subsystem? In Two Sciences of Mind. ed. by S. O'Nuallain, P. McKevitt and E. MacAogain, John Benjamins, Amsterdam, pp.115-131, (1997).

[10]Bickhard, M.H., Cognitive Representation in the Brain. In Encyclopedia of Human Biology. 2nd Ed. ed. by Dulbecco, Academic Press, New York, pp.865-876, (1997).

[11]Bickhard, M.H., A Process Model of the Emergence of Representation. In Emergence, Complexity, Hierarchy, Organization, Selected and Edited Papers from the ECHO III Conference, Espoo, Finland, August 3-7. Acta Polytechnica Scandinavica,

Mathematics, Computing and Management in Engineering Series No. 91, ed. by G. L. Farre and T. Oksala, pp.263-270, (1998).

[12]Bickhard, M.H., Levels of Representationality. Journal of Experimental and Theoretical Artificial Intelligence, 10, pp.179-215, (1998).

[13]Bickhard, M.H., Interaction and Representation. Theory and Psychology, in press.

[14]Bickhard, M.H. and Campbell, R. L., Some Foundational Questions Concerning Language Studies: With a Focus on Categorial Grammars and Model Theoretic Possible Worlds Semantics. Journal of Pragmatics, 17(5/6), pp.401-433, (1992).

[15]Bickhard, M.H. and Campbell, R.L., Topologies of Learning and Development. New Ideas in Psychology, 14, 2, pp.111-156, (1996).

[16]Bickhard, M.H. and Christopher, J.C., The Influence of Early Experience on Personality Development. New Ideas in Psychology, 12, 3, pp.229-252, (1994).

[17]Bickhard, M.H. and Richie, D.M., On the Nature of Representation: A Case Study of James Gibson's Theory of Perception, Praeger Publishers, New York, (1983).

[18]Bickhard, M.H. and Terveen, L., Foundational Issues in Artificial Intelligence and Cognitive Science: Impasse and Solution, Elsevier Scientific, Amsterdam, (1995).

[19]Block, N., Flanagan, O. and Güzeldere, G., The Nature of Consciousness, MIT, Cambridge, MA, (1997).

[20]Bullock, T.H., Spikeless Neurones: Where do we go from here? In Neurones without Impulses. ed. by A. Roberts and B. M. H. Bush, Cambridge University Press, Cambridge, pp.269-284, (1981).

[21]Campbell, D.T., Evolutionary Epistemology. In The Philosophy of Karl Popper ed. P. A. Schilpp, Open Court, LaSalle, IL, pp.413-463, (1974).

[22]Campbell, D.T., Levels of Organization, Downward Causation, and the Selection-Theory Approach to Evolutionary Epistemology. In Theories of the Evolution of Knowing. ed. by G. Greenberg and E. Tobach, Erlbaum, Hillsdale, NJ, pp.1-17, (1990).

[23]Campbell, R.L., A Shift in the Development of Natural-Kind Categories. Human Development, 35, 3, pp.156-164, (1992).

[24]Campbell, R.L. and Bickhard, M. H., Knowing Levels and Developmental Stages, Karger, Basel, Switzerland, (1986).

[25]Campbell, R.L. and Bickhard, M.H., Clearing the Ground: Foundational Questions Once Again. Journal of Pragmatics, 17(5/6), pp.557-602, (1992).

[26]Campbell, R.L. and Bickhard, M.H., Types of Constraints on Development: An Interactivist Approach. Developmental Review, 12, 3, pp.311-338, (1992).

[27]Carlson, N.R., Physiology of Behavior, Allyn and Bacon, Boston, (1986).

[28]Chalmers, D.J., The Conscious Mind, Oxford University Press, Oxford, (1996).

[29]Cohen, J.D. and Schooler, J.W., Scientific Approaches to Consciousness, Erlbaum,

Mahwah, NJ, (1997).
[30]Cummins, R., Representations, Targets, and Attitudes, MIT, Cambridge, MA, (1996).
[31]Dennett, D.C., Consciousness Explained, Little, Brown, Boston, (1991).
[32]Dowling, J.E., Neurons and networks, Harvard University Press, Cambridge, MA, (1992).
[33]Dretske, F.I., Knowledge and the Flow of Information, MIT Press, Cambridge, MA, (1981).
[34]Dretske, F.I., Explaining Behavior, MIT Press, Cambridge, MA, (1988).
[35]Flanagan, O., Consciousness Reconsidered, MIT, Cambridge, MA, (1992).
[36]Fodor, J.A., Psychosemantics, MIT Press, Cambridge, MA, (1987).
[37]Fodor, J.A., A Theory of Content, MIT Press, Cambridge, MA, (1990).
[38]Fodor, J.A., Information and Representation. In Information, Language, and Cognition. ed. by P. P. Hanson, University of British Columbia Press, Vancouver, pp.175-190, (1990).
[39]Fodor, J.A., Concepts: Where Cognitive Science went wrong, Oxford University Press, Oxford, (1998).
[40]Fuxe, K. and Agnati, L.F., Volume Transmission in the Brain: Novel Mechanisms for Neural Transmission, Raven, New York, (1991).
[41]Fuxe, K. and Agnati, L.F., Two Principal Modes of Electrochemical Communication in the Brain: Volume versus Wiring Transmission. In Volume Transmission in the Brain: Novel Mechanisms for Neural Transmission. ed. by K. Fuxe and L.F. Agnati Raven, New York, pp.1-9, (1991).
[42]Gallistel, C.R., The Organization of Action: A New Synthesis, Lawrence Erlbaum, Hillsdale, NJ, (1980).
[43]Godfrey-Smith, P., A Modern History Theory of Functions. Nous, 28, 3, pp.344-362, (1994).
[44]Graeser, A., The Stoic theory of meaning. In The Stoics. ed. by J. M. Rist, University of California Press, Berkeley, CA, (1978).
[45]Hall, Z.W., Molecular Neurobiology, Sinauer, Sunderland, MA, (1992).
[46]Hanson, P.P., Information, Language, and Cognition, Oxford University Press, Oxford, (1990).
[47]Hansson, E., Transmitter receptors on astroglial cells. In Volume transmission in the brain: Novel mechanisms for neural transmission. ed. by K. Fuxe and L. F. Agnati, Raven, New York, pp.257-265, (1991).
[48]Hilbert, D., The Foundations of Geometry, Open Court, LaSalle, IL, (1971).
[49]Hooker, C.A., Physical Intelligibility, Projection, Objectivity and Completeness: The divergent ideals of Bohr and Einstein. British Journal for the Philosophy of Sci-

ence, 42, pp.491-511, (1992).

[50]Hookway, C., Peirce, Routledge, London, (1985).

[51]Kalat, J.W., Biological Psychology. 2nd Edition, Wadsworth, Belmont, CA, (1984).

[52]Kandel, E.R. and Schwartz, J.H., Principles of Neural Science. 2nd ed., Elsevier, New York, (1985).

[53]Kinsbourne, M., Integrated Field Theory of Consciousness. In Consciousness in Contemporary Science. ed. by A.J. Marcel and E. Bisiach, Oxford University Press, Oxford, pp.239-256, (1988).

[54]Kinsbourne, M., What Qualifies a Representation for a Role in Consciousness? In Scientific Approaches to Consciousness. ed. by J.D. Cohen and J.W. Schooler, Erlbaum, Mahwah, pp.335-355, (1997).

[55]Kneale, W. and Kneale, M., The Development of Logic, Clarendon, Oxford, (1986).

[56]Levine, A. and Bickhard, M.H., Concepts: Where Fodor Went Wrong. Philosophical Psychology. (in press).

[57]Loewer, B. and Rey, G., Meaning in Mind: Fodor and his critics, Blackwell, Oxford, (1991).

[58]Marcel, A.J. and Bisiach, E., Consciousness in Contemporary Science, Oxford University Press, Oxford, (1988).

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

[60]Millikan, R.G., White Queen Psychology and Other Essays for Alice, MIT Press, Cambridge, MA, (1993).

[61]Nauta, W.J.H. and Feirtag, M., Fundamental Neuroanatomy, Freeman, San Francisco, (1986).

[62]Piaget, J., The Construction of Reality in the Child, Basic, New York, (1954).

[63]Revonsuo, A. and Kamppinen, M., Consciousness in Philosophy and Cognitive Neuroscience, Erlbaum, Mahwah, NJ, (1994).

[64]Roberts, A. and Bush, B.M.H., Neurones without Impulses, Cambridge University Press, Cambridge, (1981).

[65]Rosenthal, S.B., Meaning as Habit: Some Systematic Implications of Peirce's Pragmatism. In The Relevance of Charles Peirce. ed. by E. Freeman La Salle, IL: Monist, LaSalle, IL, pp.312-327, (1983).

[66]Sanches, F., That Nothing is Known, Cambridge University Press, Cambridge, (1988/1581).

[67]Stich, S. and Warfield, T.A., Mental representation : a reader, Blackwell, Oxford, UK, (1994).

[68]Thatcher, R.W. and John, E.R., Functional Neuroscience Vol. 1 Foundations of Cognitive Processes, Erlbaum, Hillsdale, NJ, (1977).

[69]Tiles, J.E., Dewey, Routledge, London, (1990).

[70]Tye, M., Ten Problems of Consciousness, MIT Press, Cambridge, MA, (1995).

[71]Vizi, E.S., Non-synaptic Transmission Between Neurons: Modulation of Neurochemical Transmission, Wiley, New York, (1984).

[72]Wittgenstein, L., Tractatus Logico-Philosophicus, Routledge, New York, (1961).

[73]Yuan, L. and Ganetzky, B., A Glial-Neuronal Signaling Pathway Revealed by Mutations in a Neurexin-Related Protein. Science, 283, pp.1343-1345, (1999).

This page is intentionally left blank

Chapter 11

Intentionality and Foundations of Logic: a New Approach to Neurocomputation

Gianfranco Basti
Pontifical Lateran University

Abstract

In this work we start from the idea that intentionality is the chief characteristic of intelligent behavior, both cognitive and deliberative. Investigating the "originality of intelligent life" from this standpoint means investigating "intentional behavior" in living organisms. In this work, we ask epistemological questions involved in making the intentional behavior the object of physical and mathematical inquiry. We show that the subjective component of intentionality can never become object of scientific inquiry, as related to self–consciousness. On the other hand, the inquiry on objective physical and logical components of intentional acts is central to scientific inquiry. Such inquiry concerns logical and semantic questions, like reference and truth of logical symbols constituted as such, as well as their relationship to the "complexity" of brain networking. These suggestions concern cognitive neuroscience and computability theory, so to constitute one of the most intriguing intellectual challenges of our age. Such metalogical inquiry suggests indeed some hypotheses about the amazing "parallelism", "plasticity" and "storing capacity" that mammalian and ever human brains might exhibit. Such properties, despite neurons are over five orders of magnitude slower than microchips, make biological neural nets much more efficient than artificial ones even in execution of simple cognitive and behavioral tasks.

Keywords : intentionality, cognitive science, artificial intelligence, connectionism, neural networks, foundations of logic, diagonalization

11.1 Introduction

In this work, we limit ourselves to "the originality of intelligent life". We begin with the hypothesis that such originality depends in logic and psychology on *intentionality.* We work from *cognitive neurosciences,* because this approach allows us to deal with intentionality from a more rigorous theoretical perspective than from classical ones, such cognitive psychology or phenomenological analysis. This methodology allows us to deal with our problem in *an objective* and, at the same time, *non–redutionistic way.* In the study of mental processes,

it links the *neurophysiological* component with the *logical* (semantic) and thus the *psychological* component — from the objective standpoint of the information processing, not from the subjective one of the introspection on consciousness states. The theoretical character of this analysis allows us to attain the *ontological* level of the analysis. I.e., allows us to discuss the *metaphysical* question of the originality of the intelligent life (traditionally defined as the problem of the immaterial character of intelligence) by using the "picklock" of *metalogic.* In other terms, it becomes possible to deal with the metaphysical question of the originality of the intelligent life, starting from the foundations of semantic "objects" such as "truth", "reference", "meaningfulness" of statements in a given language. Particularly, we start from the hypothesis that the process of logical constitution of these semantic relations and operations requires to be "implemented" in physical structures provided with given properties.

Our work is divided into two main sections. In the *First Section*, we deal with the study of intentionality, as characteristic of intelligent life, in the framework of *Artificial Intelligence* (AI) and of *connectionism* (Neural Networks, NN) research programs. We show the logical and meta–logical limitations of these two approaches to the problem of intentionality. In the *Second Section,* we discuss the relevance of a particular approach to the problem of logical foundations after the *Gödel incompleteness theorems* and its relevance for the problem of intentionality. This approach constitutes the *logical* counterpart of the well–known *epistemological* theory of true knowledge as *self–conforming* (*adaequatio*) of the mind to reality. This foundational theory consists in a particular application to the constitution of the logical objects of Thomas Aquinas's general ontology. This ontology is founded on the real distinction between *being as essence* and *being as existence,* considered as two metaphysical and/or metalogical constituents of each thing (either physical or logical). We emphasize particularly the relevance of this approach for dealing with characteristic problems related to the Gödel incompleteness theorems for formal systems. The only way to avoid such limitation theorems is to allow a change of axioms in the formal system concerned, so to make it "dynamic" or "recursive" in a deeply new sense. We suggest the relevance of such an approach for a logically consistent theory of intentionality, as well as for the solution of cognitive neuroscience problems related to neural dynamics and neural computations relationships — e.g., the true question of "parallelism" in brain computations, the "plasticity" and the "memorization capabilities" of brain computations, with respect to their artificial simulations. They are all questions for which neither AI, nor NN approaches to cognitive neuroscience, have satisfying solutions.

11.2 Intentionality and Cognitive Neuroscience

11.2.1 *The Functionalist Approach in Cognitive Neuroscience*

11.2.1.1 *The Origins of the Functionalist Approach*

When the AI research program was born in late 50's, it was generally held that a new age in psychological and neurophysiological studies was starting: the age of *cognitive sciences* [1]. Effectively, this approach seemed to constitute an escape from the old dichotomy in scientific psychology between:

1. the *subjectivism* of the introspective method of phenomenological psychology, typical of the *cognitivism* of *Gestalttheorie*; and
2. the *objectivism* of the mechanistic method of associative psychology, typical of *behaviorism.*

By way of difference, the *functionalist theory of mind* introduced by Hilary Putnam [2], argued that the objective correlate of a subjective state of consciousness is double. It is constituted by the *information flow of the logical operations* in the brain, considered as a logical (computational) machinery, and not by the simple *energy flow of its physical operations.* Philosophically, the problem of the mind–body relationship could be reduced to the problem of the relationship between the *software* and the *hardware* of the computational architecture of the brain.

The functionalist approach in the study of mind is the final chapter of a long history in the modern theory of mind that has the following main steps:

1. The first step was the development of a *rationalist theory of mind* by modern philosophers such as Descartes, Leibniz and Kant. This theory identifies the thought processes with *formal inferences,* with logical procedures manipulation of symbols according to formal rules. For Kant these rules and procedures are determined *a priori* in human minds and largely *unconsciously.* They can become aware only after a long study, so that only when they are thought *in abstracto* they become objects of a particular science such as the *formal logic* [3]. Particularly, the core of the perception is for Kant an act of *productive fantasy.* It consists in the development of a particular *schemes* or "rules for the fantasy synthesis" for each abstract concept. By this scheme, a given sensible intuition can be organized according to a given formal concept for producing a determinate perception. In short, in our mind we do not have the image of a dog. We have a rule for the constitution of different images of the singular dogs

that our sensibility presents to us in different contexts. By this deductive scheme constitution, for each abstract concept there exists a formal scheme for its application on a domain of sensible objects. This "deductive schematism" is thus defined by Kant as "an art concealed in the depth of the human soul, whose real modes of activity is hardly likely ever to allow us to discover" [4].

2. The second step toward the functionalist theory of mind is the development during the last century and the first half of our century of *symbolic* or *mathematical logic*. The aim of this research program, started since the seventeenth century with Leibniz's *characteristica universalis,* was the rigorous construction of the formal logic as a *logical calculus*. This construction reached its apogee at the end of the last century with G. Frege's work. Both the notion of *propositional function,* as a formal scheme with free variables for proposition construction, and the notions of *logical quantifiers,* for the construction of the class logic in the form of a predicate calculus were essential. This improvement made possible the rigorous systematization of the logical calculus into its main three branches of the *predicate* (*class*) *calculus,* of the *propositional calculus,* and of the *relation calculus.*
3. The third step toward the functionalist theory of mind was the demonstration of a *fundamental theorem of computability theory* by the English mathematician A.M. Turing [5]. According to this theorem, each *computable function* of the mathematical and/or of logical calculus can be *recursively* calculated through a *finite procedure* by an appropriate elementary computational architecture called *Turing Machine* (TM). Of course, the behavior of each TM can be simulated by another TM, on condition that, onto the "ribbon" (memory) of the second one, all the instructions to execute the calculations of the first one are explicitly written in the language of the second TM. Fundamental consequence of this theorem is that the *universality* in computations can be granted *iff* each singular TM in turn can be simulated by a *Universal Turing Machine* (UTM) with an infinite "ribbon". In other words, it is supposed that the universality of the *codes* or "alphabets" used by each single TM for executing its recursive computations can be founded only through the isomorphism (biunivocal correspondence) between these alphabets and the *universal fixed alphabet* of the UTM

   A formal consequence of this theory and of the notion of "λ–calculus" developed by A. Church is the so called *Church's-Turing's thesis* accord-

ing to which the class of *all the computable functions* is equivalent to the class of the recursively computable functions and this class, in turn, is equivalent to the class of functions computable by a TM. This thesis, because of Gödel incompleteness theorems [6], cannot be formally demonstrated so to remain only a hypothesis. Immediately related with such a limitation theorem is the other one according to which it cannot be formally demonstrated that a UTM can calculate through an ending procedure. This is the famous *halting problem* demonstrated by Turing himself. However, the anthropological consequence of this theory is that, if we accept the rationalist theory of mind, that is, if we reduce the human thought to a logical calculus, each individual human mind has to be considered as logically equivalent to a TM. Hence each singular mind has to be considered as a function of some "universal mind", defined in the rigorous terms of a UTM. Such a consequence, that constitutes the metaphysically *monist* core of any functionalist theory of mind (see, for instance, [7]), became effective when the final step toward this theory was available in modern scientific psychology.

4. The fourth and ultimate step toward the cognitive sciences was the main hypothesis underlying the so–called "genetic" approach to the study of intelligence. This approach was developed by the Swiss psychologist J. Piaget within the classical approach of cognitive psychology [8]. According to this hypothesis, the development of abstract intelligence in human individuals corresponds to the acquisition of the *operative schemes* of four fundamental logical operations (identical, inverse, reciprocal and correlative). These operations constitute the so called "group of the four transformations" granting the relations of *reflexivity, transitivity* and *symmetry* (and hence of *equivalence* and *(extensional) identity*) of the logical reasoning. These schemes are owned by the subject at the unconscious level, so to recover to modern cognitive psychology the notion of the *cognitive unconscious* (see above p. 241) of the Kantian theory of mind schematism [9].

   An essential difference with the Kantian schematism has, however, to be soon emphasized. It is essential indeed for our aims of a theoretical treatment of the perception problem within the framework of the cognitive sciences. While the Kantian schematism is essentially *deductive*, Piaget's schematism would be *inductive.* What is essential for Piaget's theory of perception is in fact that the perceptual schemes of the operative intelligence are submitted to a procedure of continuous redefinition with respect

> to changing reality. It becomes possible by supposing a mechanism of *assimilation – accommodation* of the schemes. That is, the new sensible knowledge, as far as it cannot be assimilated to the old *a priori* schemes, determines an accommodation of these schemes to the new occurrences. In this way it grants the development in time of the intelligent capabilities of the subject. This "evolutionary" idea of the scheme constitution recovers thus to modern cognitive psychology the core of the Scholastic theory of *an inductive schematism* typical of its theory of perceptual intentionality [10].

Until now, Piaget's idea has not found a proper operational correlate in the modern theory of computability. It relies on reasons we illustrate in the next paragraph (See pp. 245ff.), ultimately depending on the same foundations of modern logic and mathematics (See pp. 263ff.). Our systematic effort is thus related to a re-consideration of the foundations of logic and mathematics to overcome these essential limitations. They involve not only the psychology of perception and the cognitive science, but also the modern theory of computability in its many applications in all the fields of modern science.
Finally, and more deeply, these limitations involve the same destiny of realism in modern epistemology (See pp. 252ff. and pp. 263ff.).

However, as far as we do not consider this essential point of the *inductive* versus a *deductive* procedure of scheme constitution, and we uncritically accept an a priori constitution of the schemes in the *cognitive unconscious* of human mind, the following conclusion is not hazardous. The functionalist approach to cognitive sciences is a sort of operational translation of the Kantian transcendental philosophy of mind [11]. Namely, just as the very same *software* can be implemented into different *hardware's,* so, in the framework of the functionalist approach, it could be possible to intend a computer simulating a formal operational scheme like a transcendental counterpart of what individual minds do at the empirical level.

More precisely, this fundamental statement of the functionalist approach to the study of mind can be synthesized in D. R. Hofstadter's terms by the principle of the "AI dogma". Every time the computer simulated successfully a human intelligent behavior, the software of this computation *must necessarily* imply some essential isomorphism with the "software" running in the human brain [7].

This "dogma" exemplifies in one only statement the core of the famous *Turing test* [12], because it is a direct consequence of the computability theory for TM's. Let us suppose that we have to test whether is it a human individual or a

computer the mysterious individual "who" is giving us the "intelligent" responses to our questions we are setting "him". "His" mystery is that we cannot see "him" because "he" is in another room and we can communicate with "him" only through a teletypewriter. If a computer effectively gives these intelligent responses, but they are indistinguishable from those normally given by a human individual, the intelligent human behavior has been perfectly simulated by the computer. Hence, according to "AI dogma", some fundamental isomorphism *must* exist between the software running in the machine and the software running in the human mind. The possibility that each TM can be perfectly simulated by another TM "instructed", "programmed" in a suitable way, implies this consequence for the cognitive sciences.

11.2.1.2 *Formal Semantics and the Problem of Schematism*

This possibility exemplifies also the response (see, for instance [11], [13-14]) that the *functionalist approach* tried to give to the problem of *conscious intentionality* in terms of A. Tarski's [15-16] and R. Carnap's [17] *formal semantics.* Indeed, what the Scholastic philosophy enhanced since Middle Age and modern phenomenological and cognitive psychology rediscovered since the pioneering work of F. Brentano [18], is the *intentional character of any psychical act as such.* In other words, what characterizes any psychical act, as far as it is distinguished from a physical act, is its intrinsic *reference to a content,* or "aboutness". This content has to be considered both in its *extensional* sense (that is, as a given object either physical or ideal) and in its *intensional* sense (with "s", i.e., the intended meaning we associate to that object). In short, *intending* an intensional content and *referring to* some extensional content is what constitutes a psychical act as *intentional.* Hence, considering the act of thought in a purely formalistic way without any reference to a content, in the sense of Descartes' *cogito* or of Kant's *Ich denke überhaupt,* is rightly considered by the phenomenology as a misleading abstraction from the real situation of human psychology. When I think, I desire, I will, I feel, I perceive, etc., I think, desire, will, feel or perceive always *something*! In this way, the problem of logical *truth* of a given proposition has, from the psychological standpoint, an intentional character and from the formal logic standpoint a *semantic* character. Semantics is indeed the logical discipline which "deals with certain relations between expressions of a language and the objects (or "state of affairs") 'referred to' by those expressions" [16]. A. Tarski indeed, for the first time in the history of modern logic, defined in a rigorous way for formal languages this

semantic relationship to a content (*reference*) and the semantic relationship of *truth* within a purely *extensional* and *formalistic* approach to this problem.

Tarski solves the problem of a formal definition of semantic concepts like truth by affirming the necessary *semantically open* character of any formal language whose truthfulness has to be rigorously defined and hence (recursively) proved. That is, in discussing the problem of the formal, *consistent* (i.e., that does not imply contradictions) definition of semantic concepts, we have *always* to distinguish between two different languages. The first, the *object–language,* is the language to be checked. The definition of truth we are seeking applies to propositions of this language. The second, the *meta–language,* is the language in which we "talk about" the first one and in terms of which we can construct a consistent definition of truth for the first language propositions. Of course, the two notions are *relative* and not absolute. Indeed, if we want to check the truth of the proposition of the meta–language, we have to consider it the object–language of another meta–language, and so on. The conclusion that no formal language can be the meta–language of itself is directly related with Gödel's demonstration of incompleteness of formal arithmetic (Peano's axiomatic arithmetic), against original Hilbert's formalistic program [16]. This is because formal semantics must use a *recursive* procedure of *satisfaction* for defining formally the notions of truth and reference[1]. Now, for our aims, three reflections about Tarski's semantic theory of truth are to be made:

1. *From the computational standpoint,* Tarski emphasizes that the recursive procedure of satisfaction of a given propositional function with n free

[1] The satisfaction is a particular semantic relation between arbitrary objects and propositional functions. Generally an object (e.g., *snow*) satisfies a propositional function (e.g., *x is white*) if the latter becomes a true proposition when the name of the object is used to replace the free variables in it (e.g., *snow is white*). Of course, in our case we cannot use this definition of satisfaction for defining truth, since it supposes the definition of truth. Tarski must use thus a recursive procedure for the definition of satisfaction. Starting from objects satisfying the simplest propositional functions (e.g., for natural numbers, all the numbers x and y satisfying the functions "*x is greater than y*", or "*x is equal to y*"), we can define the conditions under which compound functions are satisfied too (e.g., the logical disjunction "*x is greater than y*, or *x is equal to y*" is satisfied for all x and y satisfying at least one of the above simplest functions). In this way, we can construct the formal definition of truth in terms of satisfaction: *A proposition is true if it is satisfied by all the objects and it is false otherwise* [16]. Where, of course, the totality of the objects of which we are speaking about are to be interpreted as the totality of the objects to which the propositions of the object–language refer. With similar recursive procedures, it is possible for formal semantics to give rigorous definitions – in the same framework of the distinction between two different formal languages: the object - language and the meta – language – also of other semantic terms, such as the notion of *reference and/or designation* (e.g., "Columbus designates (denotes) the discoverer of America") as well as the notion of *definition* (e.g., "$x \cdot 2 = 1$ defines (uniquely determines) the number ½").

variables is a relation with $n + 1$ terms (e.g., for unary functions is a binary relation, for binary functions is a ternary relation, and so on). So, in defining the notion of satisfaction for formal languages with propositional functions of an arbitrary number of free variables, we are not faced with only one notion of satisfaction, but with infinitely many notions that must be introduced simultaneously because they cannot be defined independently. In this way, the core of a recursive procedure of satisfaction, is to define a recursive procedure of substitution of a many–termed relation between propositional functions and an indefinite number of objects, with a binary relation between functions and finite *sequences* of objects with an arbitrary number of terms [16]. The relationship of this theory of truth with Gödel theorems is thus immediate. Indeed, such a recursive substitution implies to have only one unary function for enumerating recursively a collection of objects, so to have only one ordered sequence of them. In this way, a recursive procedure of substitution is identical with that recursive procedure of *coding* called Gödel *numbering.* The essential result of Gödel theorems is indeed that such a coding function cannot be written in the same formal language (arithmetic) in which the objects and/or the functions to be enumerated are written. That is, it is not possible to conceive such a substitution procedure as a *diagonalization procedure.* A diagonalization procedure can be defined as the iterative procedure of substitution of an *n*–ary function with an unary function. For instance, given a binary function of the type $h(x, z)$ or $f_z(x)$, the diagonalization would consist in its iterative substitution with the unary functions $h(x, x)$ or $f_x(x)$. This last way of writing a unary function, $f_x(x)$, is notable because in it the same x plays the double role of argument and of index of the same function. This suggests that the diagonalization procedure is effectively a procedure of *class closure by diagonalization,* that is, the computational counterpart of the logical notion of *complete induction* [19]. This suggestion is much more than a suspect in the case of the substitution procedure relative to the notion of satisfaction in Tarski's theory of truth. Is not Tarski's definition of truth identified with the satisfaction of a propositional function *simultaneously* for *all* the objects of a given linguistic domain (see note 1)? If Tarski poses the distinction between an object–language and a higher order meta–language as necessary and sufficient condition for his formal (recursive) definition of truth, is thus precisely because such a class closure by diagonalization cannot be performed without contradiction inside the same formal language. This demonstration is indeed

the main result of Gödel theorems of incompleteness of the formal arithmetic. This result, precisely through the work of Tarski and Turing, can be thus extended to any formal language.

2. *From the logical standpoint*, another consequence derives from the precedent discussion. As Tarski himself and Gödel rightly emphasized, from such a semantic approach *no absolute* notion of truth becomes possible, even in a *local sense,* i.e., for a *finite* domain of objects. In this regard, it is important to avoid a possible misunderstanding. It is really true that Gödel theorems hold only for *general recursive functions.* That is, they properly hold only for functions defined on all their *infinite* domain of application. On the contrary, it is impossible to exclude the convergence of the recursive procedure for *partial recursive functions* [20]. Namely, it is impossible to demonstrate Gödel results for recursive functions defined on only a finite subset of their infinite domain of application. In this case, indeed, the recursive procedure could converge within the domain of application even though out of the (sub–)domain of definition. In this sense, it is formally correct to invoke with Kleene a healthy *finitism* to avoid the more destructive effects of Gödel theorems in computability theory [20]. On the contrary, in the case of the semantic notion of truth such a finitism has no effect. Either finite or infinite a domain of objects is, to meet Tarski's criterion of satisfaction, it is necessary to migrate outside a formal language for judging from a higher logical order the truthfulness of its propositions. For this unavoidable necessity of a higher level meta–language, such a formal definition of truth implies that truth notion cannot be absolute at all, but always *relative.* Commenting on this evidence, Gödel in his philosophical reflections rightly quotes Plato's theory of truth. Especially, this result is consistent with the truth theory expressed in Plato's famous *Letter VII.* According to this text, any true knowledge necessarily exceeds any procedure of demonstration as well as any "fixed form" of language. That is to say, truth exceeds any "formal language" that pretends to assert *forever* its primitives and its rules. In short, for Plato as well as for Gödel, the logical universals, either exist ultimately by themselves, or no consistent procedure of construction (i.e., of formal definition and/or of formal demonstration) could ever pretend to constitute them.
3. Finally, *from the epistemological standpoint*, another consequence must be drawn from the previous discussion that is essential for our aims. Owing to its pretension of meeting the Aristotelian notion and hence the

common sense notion of truth, namely the notion of truth as "correspondence to reality", it seems that the semantic theory of truth implies by itself an epistemological position of *realism.* Effectively, K. R. Popper tried to interpret it as a theory of truth as *correspondence to facts* [21], as if Tarski's theory of truth was able to give to Popper's biology–inspired epistemology its rigorous formal, and hence scientific foundation. This interpretation of Tarski's results is absolutely inconsistent and the possibility of interpreting the semantic theory of truth as supporting a position of epistemological realism was always explicitly rejected by Tarski [16]. The semantic theory of truth has nothing to say about the conditions under which a given simple ("atomic" in L. Wittengstein's terms) proposition (and overall an empirical proposition) like *snow is white* can be asserted. As he correctly affirms, his theory implies only that whenever we assert or reject this proposition, we must be ready to accept or reject the correlated meta–proposition: *the sentence "snow is white" is true.* In other terms, the semantic theory of truth has nothing to do with the problem of *the formal constitution of true propositions* but only with the problem of *the formal justification of true propositions.* In other words, it is completely immersed within the *axiomatic method* identifying logic with the "logic of justification" of proposition already constituted, and not within the *analytic method,* identifying logic with the "logic of discovery" (See pp. 263ff.). For this reason, in the approaches of the formal semantics and of the functionalist theory to the problem of *reference* there is no room for the treatment of the problem of the *real reference* (See pp. 269ff.). This problem is methodologically excluded in them. So, J. A. Fodor, quoting R. Carnap, rightly emphasized that the treatment of the intentionality problem within the functionalist theory of mind has to be conjugated with a rigorous principle of *methodological solipsism* [13]. In fact, the functionalist theory has nothing to do with the problem of the *reference to reality* of some mental state. Better yet, if we accept the use of Tarski's and Carnap's formal semantics within the functionalist theory of mind as the only possible *scientific* counterpart of the *naive* notion of intentionality in the "folk psychology", the epistemological realism can be only negated in the name of the above remembered methodological solipsism [11.13-14]. Any mind–state that we might characterize as a "propositional attitude" (= the psychological counterpart of a propositional function in the functionalist theory of mind) can refers only to another mind–state or "mental representation", like to the object capable of satisfying it, for constructing valid

propositions. And this is unavoidable in a functionalist theory of mind, precisely for the same reason for which in formal logic and in formal semantics, "the fundamental conventions regarding the use of any language require that in any utterance we make about an object, it is the name of the object which must be employed and not the object itself" ([16], p. 55). This is the core of the mentalist *representationalism* and of the logic *nominalism* intrinsic to the functionalist approach [11.13-14]. It justifies completely Fodor's pretension that the functionalist approach is an operational counterpart of the Kantian theory of mind and of his epistemological representationalism, against the epistemological realism.

If the formal semantics constitutes the operational counterpart of the intentionality in the functionalist theory of mind, there is no room in the functionalist approach for Piaget's *inductive schematism.* A process of scheme accommodation poses itself at the level of the formal constitution of the scheme itself. But it is precisely about this procedure of formal constitution of the logical symbols that formal semantics has *in principle* nothing to say.

11.2.1.3 *Intentionality and the Metaphor of the Three "Rooms"*

It is hard to defend the functionalist pretension of using Tarski's semantics for dealing with psychological intentionality, overall in the study of perception. Indeed, what we mean by "intentionality" is not only the act of *reflexive thought* of formal manipulation of logical symbols and relations already otherwise constituted in our mind. In this sense intentionality could be in agreement with the methodological solipsism of functionalist theory, as well as with nominalism of Tarski's formal semantics. On the contrary, intentionality essentially means the act of *productive thinking* of new logical symbols and hence of new logical relations. In short, intentionality is essentially related to the act of *constitution of symbols,* and in the case of constitution of *true* symbols, intentionality is essentially related to the problem of constitution of symbols *adequate* to the singular context of their use. So, any scientific theory of intentionality must deal with the problem of intentionality at the *pre–symbolic level.*

In summary, the formalist method requires that functionalism posits intentionality only at the symbolic level of mental information processing rather than at the more fundamental pre–symbolic level of the constitution of symbols.

This criticism against the functionalist approach to intentionality has been developed in the last twenty years. In this regard, two other counterexamples of the famous "Turing room" metaphor have been proposed: the "Searle room"

and, more recently, the "Putnam room". These two metaphors exemplify indeed two main criticisms that can be posed to the symbolic treatment of the intentionality problems in the functionalist approach. These criticisms are, respectively:

1. from the standpoint of the *intensional* (with *s*) *logic* approach to the theory of intentionality;
2. from the standpoint of the *theory of coding* in the logical *foundations of computability theory.*

Let us begin with J. Searle's criticism.

11.2.1.4 *Searle's "Room" and the Intensional Approach to Intentionality*

In order to exemplify in which sense the Turing test fails in proposing a valid proof of the equivalence between a mind and a computer, J. Searle proposed the counterexample of his "Chinese room" [22-23]. Let us imagine that a person, who does not know at all Chinese, has to translate a English text into Chinese. Let us suppose to give him a dictionary as well as the complete set of rules sufficient for the exact translation of the text concerned. Even though this person produced a text resulting in an absolutely correct translation for Chinese people, nevertheless this person, *just like a machine,* would have not understood anything of what he produced. In other words, even though a Turing test satisfies the criteria of an extensional approach to the problem of meaning, nevertheless it is impossible to affirm that this approach can be considered as a satisfying operational translation of what we designate as an intentional act of knowing [22-23]. The "relation to a content" as characteristic of any intentional act implies not only the *extensional* reference to names of objects, but also the *intension* of a conscious significance by which we associate names and objects in different contexts. *Intending* a meaning and by it *referring* to an object are not the same thing, even though they are effectively always together in any conscious intentional act.

This reciprocal irreducible character of the intensional and of the extensional components of any intentional act is evident also in the logic of their linguistic expression. In the intensional logic indeed the *extensionality* axiom and the related *substitution* axiom do not hold [24]. For instance, from the extensional standpoint, the notion of "water" and the notion of "H_2O" are to be considered as synonyms, since they apply to the same collection of objects. From the intensional standpoint, however, they do not have the same meaning; just substitute the term "water" with the scientific term "H_2O" in some poetic or religious

discourse. The result is meaningless. Owing to the exclusively extensional character of the treatment of the semantic content in the functionalist approach, this approach is absolutely not sufficient for cognitive psychology.

Unfortunately, the constructive part of Searle's theory of intentionality is void of any theoretic and scientific significance. Nevertheless, what Searle's criticism rightly emphasizes is that the functionalist approach to the study of mind cannot be at all adequate owing to its exclusively extensional approach to semantic problems.

For these very same reasons W. V. O. Quine stated that the mind–body problem is essentially a *linguistic* and not ontological problem [25]. So, because of the extensional character of any scientific language, for him intentionality cannot be at all object of scientific inquiry [26]. This reductionism, typical of the logical empiricism of Quine's philosophy is typical also of P. Churchland's interpretation of the connectionist approach to cognitive neuroscience [27].

On the other hand, E. Husserl's early attempt of an intensional approach to foundations of formal logic cannot in principle lead to any constructive approach to the semantic problems of *truth* and of *reference.* Indeed, also the intensional logic solution to the problem of truth supposes a sort of axiom of completeness in formal logic. It supposes completeness at least in the fundamental sense of an equivalence principle between the *non–contradiction principle* and *excluded middle principle.* Only by this equivalence can truth be intensionally founded on the conscious *evidence* [28]. The necessity of this equivalence for any intensional theory of truth as evidence is the deep formal reason for which Husserl abandoned his early attempts of an intensional foundation of formal logic after the publication of Gödel results two years later the publication of his main work on formal logic, *Formal Logic and Transcendental Logic* [29]. Indeed, the incompleteness of any formal language implies the unavoidable presence in it of undecidable statements. That is, in any formal language there is the unavoidable presence of true statements for which it is not possible to demonstrate them or their negation, so to violate the excluded middle principle. Also for late Husserl works, just like for Tarski and Gödel, truth can be thus only a sort of regulative idea in Kantian sense: something that is "beyond" any formal language and demonstration procedure as Husserl's late idea of *universal teleology* exemplifies.

11.2.1.5 *Putnam's "Room": Intentionality and the Problem of Coding*

One of the most exciting events in the brief history of cognitive sciences is the

abandonment of the functionalist approach by its supporter who introduced it into the scientific and philosophical debate: the mathematician and philosopher Hilary Putnam. This is related to the unsolvable problems of *reference and truth* characterizing any intentional act, when approached from the standpoint of the computability theory [30-31].

As we can expect from a cultivated logician and mathematician as Putnam is, his complete theoretical conversion form the early functionalism posed the intentionality question at the right place, both from the computational and from the logic points of view. As we saw before (See pp. 245), any formal theory of reference and truth is faced with the Gödelian limits making impossible a recursive procedure of satisfaction in a semantically closed formal language (see also note 1). What we emphasized as the core of the problem is that such a recursive procedure for being complete would imply the solution of the *coding* problem through a diagonalization procedure; that is, the solution of the so–called "Gödel numbering" problem. In computational terms, the impossibility of solving the coding problem through a diagonalization procedure means that no TM can constitute by itself the "basic symbols", the primitives, of it own computations. For this reason Tarski rightly stated that, at the level of the propositional calculus, the semantic theory of truth has nothing to say about the conditions under which a given simple ("atomic" in L. Wittengstein's terms) proposition can be asserted. And for this very same reason, in a fundamental paper about *The meaning of "meaning"* [30], Putnam stated that no ultimate solution exists either in extensional or in intensional logic both of the problem of *reference* and, at the level of linguistic analysis, of the problem of *naming.*

In this sense, Putnam stated, we would have to consider ultimately names as *rigid designators* "one – to - one" of objects in S. Kripke's sense [32]. But no room exists both in intensional and in extensional logic for defining this natural language notion of *rigid designation* in terms of a logical relation, since any logical relation only holds among terms and not between terms and objects, as Tarski reminded us. Hence a formal language has always to suppose the existence of names as rigid designators and cannot give them a foundation.

To explain by an example the destructive consequences of this point for a functionalist theory of mind, Putnam suggested a sort of third version of the famous "room–metaphor", after the original "Turing test" version of this metaphor and J. Searle's "Chinese–room" version of it. Effectively, Putnam proposed by his metaphor a further test that a TM cannot solve and that, for the reasons just explained, has much deeper implications than the counterexample to the Turing test proposed by Searle. For instance, Putnam said, if we ask

"how many objects are in this room?", the answer supposes a previous decision about which are to be considered the "real" objects to be enumerated — i.e., rigidly designated by numerical units. So, one could answer that the objects in that room are only three (a desk, a chair and a lamp over the desk). However, by changing the enumeration axiom, another one could answer that the objects are many billions, because we have to consider also the molecules of which the former objects are constituted.

Out of metaphor, any computational procedure of a TM (and any computational procedure at all, if we accept Church's thesis) supposes the determination of the basic symbols on which the computations have to be carried on. Hence, from the semantic standpoint, any computational procedure supposes that such numbers are *encoding* (i.e., unambiguously naming as rigid designators) as many "real objects" of the computation domain (See [31], p. 116). In short, owing to the coding problem, the determination of the *basic symbols* (numbers) on which the computation is carried on, *cannot have any computational solution* at the actual state of development of the formal computability theory.

To sum up, for Putnam's analysis, the functionalist approach to cognitive intentionality has to do essentially with an *inductive schematism* of concepts and therefore with the act of *productive thinking* for the constitution of logic symbols (See p. 250). On the contrary, the functionalist approach can at last give some limited operational version of the *deductive schematism* and hence of the intentional act intended as an act of *reflexive thought* on symbols already constituted. In other words, neither the problem of real reference nor of inductive schematism, essential for a scientific theory of human and animal perception, have in principle any solution from the functionalist approach to cognitive science.

11.2.2 *The Connectionist Approach in Cognitive Neuroscience*

11.2.2.1 *What is the Connectionist Approach?*

Generally a Neural Network (NN) is conceived as a computational architecture simulating brain dynamics and in a *pre–symbolic* form cognitive behaviors. Where "pre–symbolic" has to be intended in the "weak" sense that this computational architecture is conceived for reducing the relevance of the programming operation, not in the "strong" sense of "constitution of logic symbols".

From the engineering standpoint, NN's are useful for their capability of *automatic extraction* of statistical relations in the input data of a *higher order*

than simple averages, so to perform operations generally very difficult for classical symbolic AI models, such as *pattern recognition* and temporal series previsions in complex systems.

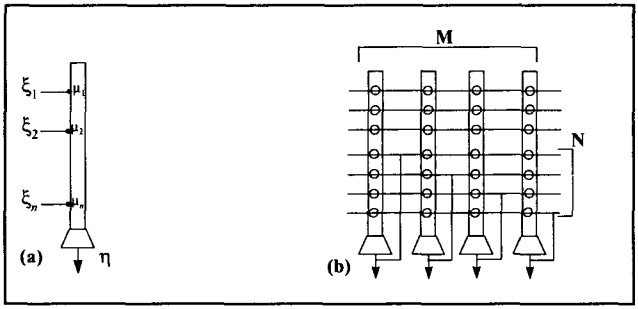


Fig. 11.1 (a): Scheme of the "Formal Neuron" of McCulloch and Pitts. (b): Scheme of a Self–organizing Neuron Module According to the General Eqs. (1) - (3).

From the architectural standpoint, artificial NN's are networks interconnected in *parallel ways* composed of simple *adaptive* elements (neurons) and by their hierarchical organization, designed for interacting with real world in a way similar to *biological* NN's. A "neuron" of an artificial NN is effectively a *threshold logic unit* of the logical circuit implemented in a classical digital computer. The different architectures of NN's depend on different modalities of determination of the *threshold* and of the *interconnections* among neurons. For instance, in the "formal neuron" (see Fig. 11.1 (a)) of the first model of artificial NN, i.e., McCulloch's and Pitt's "formal" NN, the output frequency η of each unit as a function of an input ξ_j, $j = 1, 2, ..., n$ is given by the following function:

$$\eta = \text{cost}\,\mathbf{1}\left(\sum_{j=1}^{n} \mu_j \xi_j - \theta\right)$$

where $\mathbf{1}(\bullet)$ is a classical Heaviside step function, μ_j are the statistical weights among the connections and θ is the threshold. Because of the threshold θ the output is discretized, so that, if the input is discretized too, it is possible to demonstrate that, for suitable values of μ and θ, a net composed of these neurons can compute whichever Boolean function.

The *adaptive* procedure in artificial NN's essentially consists in making the weights of the connections among the units *variable in time* as a function of the statistics of the neuron output. The fundamental rule by which this modification is performed is the so-called Hebbian rule [33]. This is a frequentistic rule according to which the weights μ_j change as a function of the product between input and output among the elements, so to reinforce inputs that produced stable outputs. In this way, the spontaneous formation of modules of reciprocally exciting neurons becomes possible, formally corresponding to the presence of statistic correlations intrinsic to different components of the input data. Mathematically, the Hebbian rule implies:

$$dw_j/dt = \alpha y x_j - \beta(y) \cdot w_j$$

where w_j are variable weights, x is the input, y is the output and $\beta(x)$ is a positive function of x.

Hence, given the matrix notation $\mathbf{x} = [x_1, x_2, ..., x_n]^T$ and $\mathbf{w} = [w_1, w_2, ..., w_n]^T$, where T is the transposed, if $\mathbf{C}_{xx}$ is the correlation matrix of $\mathbf{x}$, if x_j are stochastic variables with statistical stationary properties, then the w_j converge asymptotically to values such that $\mathbf{w}$ represents the eigenvector of the maximum eigenvalue of $\mathbf{C}_{xx}$. In this way, through the presence of lateral feedbacks among neuron arrays, it becomes possible to speak formally of *self - organization* of computation modules in NN's (see Fig. 11.1 (b)), according to the following equations [34]:

$$d\mathbf{y}/dt = f(\mathbf{x}, \mathbf{y}, \mathbf{M}, \mathbf{N}) \quad (1)$$

$$d\mathbf{M}/dt = g(\mathbf{x}, \mathbf{y}, \mathbf{M}) \quad (2)$$

$$d\mathbf{N}/dt = h(\mathbf{y}, \mathbf{N}) \quad (3)$$

where $\mathbf{x}$ is the vector of all the inputs at the neural module concerned, $\mathbf{y}$ is the vector of all the outputs and $\mathbf{M}$ e $\mathbf{N}$ are two adaptive connection matrixes. *Biologically,* Eq.(1) is a relaxation equation of the electrical activities of neuron modules for short t; Eqs. (2) and (3) are adaptive equations evolving on longer time scales and concerning *structural modifications* of the net. In particular, Eq. (3) represents the function of an *associative memory.* To understand this essential notion, it is necessary to introduce the distinction among two different dynamics concerning the effective functioning of an artificial NN:

1. *A learning phase* concerning the dynamics on the weights by which the net self - organizes its internal computational modules;
2. *A test phase* by which the net, after the learning, performs its own task (e.g., pattern recognition). In this phase, if we consider a NN as a dynamic system characterized by a given set of differential equations. The dynamics concerns the activation of different neurons and/or of different neuron modules according to equations that assume generally the following form for a single layer NN:

$$z_j = f\left(\sum_{i=1}^{n} w_{ji} x_i - h_i\right) \quad j = 1,\ldots,k$$

where z_j is the output of the j-th neuron, w_{ij} the connection weight between two neurons x_i is the input of the i-th neuron, h_i is a threshold and f is a non-linear function. It is thus evident that such a dynamic system effectively operates a non - linear mapping T_W between the input set **X** and the output set **Y** exemplifying the notion of *associative memory*:

$$T_W: \mathbf{X} \rightarrow \mathbf{Y}$$

From these very simple notions it is easy to understand the core of a connectionist architecture of calculation with respect to classical sequential architectures of AI. While in a sequential architecture there is a strong distinction between the logic unit of calculation (the CPU of a normal computer) and the unity(ies) of information storage (the hard disk(s) and RAM devices of a normal computer), in a connectionist architecture no distinction exists between these two components. The same units (neuron modules) devoted to process information are those devoted to the information storage too. The information stored is distributed along the weight connections where it is processed. For this reason, in the connectionist realm, we speak of *parallel distributed processing* of information in such architectures [35].

11.2.2.2 *Theoretical Limitations of Connectionism*

From the logic and computational standpoint, a NN after the learning is equivalent to a TM, reproducing in itself all the theoretical limitations we discussed above, with respect to *reference* and *truth*. Of course, the novelty with respect to classical symbolic methods of AI is the pre–symbolic task of the learning phase by which a NN seems to constitute by itself the logical symbols

of its predicate calculus. The theoretical problem is the following: is a connectionist NN in learning a computational architecture able to constitute formally its own basic symbols intended as *rigid designators* of changing objects of the real world? The answer is evidently negative. A NN could be effectively able to constitute its own basic symbols *iff,* during the learning phase, was able to modify, *depending on input,* not only the statistical weights of its fixed topology of connections, but the same geometrical topology of the connections. On the other hand, only in this case a NN will assume the typical *dynamic* and *computational* characteristics of biological networking. That is:

1. From the *dynamic standpoint,* it will assume the characteristics of an *unstable* and even *non–stationary* dynamics. Indeed, in the connectionist NN's, despite the non–linear character of such dynamic systems, the information (e.g., a pattern) is stored in each stable final state (fixed point attractor) of its dynamics. That is, it is stored in some absolute minimum of the "energy" landscape (i.e., of some complex function measuring the distance between the actual state and some target state) of the dynamics. On the contrary, what is typical of real brain networking is the *unstable* character of the signal transmission and processing among neurons. For instance, in real brains, the firing rate of neuron spikes is continuously changing. In this way, it becomes despairing any attempt to interpret in a frequentistic way the learning rule for weight connections as, on the contrary, the Hebbian rule pretends to do. Moreover, there is evidence in real brains of more complex oscillatory and even chaotic (i.e., unstable in itself, even though pseudo–stable or pseudo–periodic with respect to a properly chosen interval ε): see Fig. 11.2) global dynamic behaviors [37-40].

   The informational advantages of chaotic behavior in neural dynamics, become evident as soon as we consider the information richness hidden in the pseudo–cycles of a chaotic dynamics. Roughly speaking, in the energy landscape of a classical non-linear neural net, such as a Hopfield net, it is possible to memorize less than one pattern for each of the n minima [41]. In a chaotic memory it would be possible to profit *in real time*[2], on a deterministic basis, of all the cyclic combinations of these minima, with an

[2] Because a chaotic net does not memorize patterns "statically" into fixed points of the dynamics but into unstable cycles that can be recovered on a deterministic basis, it is not necessary to reset the net after a recognition for the next one, as with static nets. It is sufficient to change a parameter value for switching from a cycle into another.

exponential increment of the memory capability (theoretically it is possible to improve the memory capacity till 2^n patterns. See Fig. 11.3). In our view, in this dynamic use of the brain dynamic instability is hidden the secret of straightforward memorization capacity of the biological and specifically the human brain. Computationally, the main difficulty is that till now there were no effective computational techniques of pseudo–cycle extraction of any length, because of the complexity of chaotic behavior. This complexity indeed makes inapplicable to deterministic chaos classical statistical methods of signal analysis. In the last four years, however, one of us, developed a new effective technique of pseudo–cycle extraction of any length, with a computation time growing only linearly with the cycle length [42-45]. We have the definitive experimental evidence that this method, based on the new foundational ideas discussed in the next Section, can extract practically *all* the pseudo–cycles of a chaotic dynamics.

Finally, there is an amazing *evidence* of the *non – stationary character* of real brain networking. For instance, Positron Emission Tomography (PET) techniques of inquiry give a sort of biological evidence of what logicians intend with the notion of names as *rigid designators* of objects. Namely, in cognition tasks, such as attention focusing or moving object tracking, completely different neuron networks are excited to designate the very same object [46]. It is as if the real brain is continuously modifying the geometrical connection topology of its computation network, to match the object modifications. On the other hand, this sort of accommodation of the basic symbol space for matching varying objects is precisely what is needed from a NN for being able of performing *really parallel computations.* Let us illustrate briefly this essential point.

2. *From the computational standpoint,* a connectionist NN cannot be considered as a really parallel computational architecture because the inner units are *fully connected* with the input units x_k (see Fig. 11.4 (a)). A really parallel computation implies that the inner units compute functions $p_i(X)$ defined only on some subset of the input units [47]. For considering such functions as rigid designators of varying external objects it is thus necessary that the supports $S_{p_i}(X)$ of these functions are varying with the objects (See Fig. 11.4 (b)). The non–stationary character of brain networking displays all its intrinsic computational value, if interpreted in this sense [48-52]. In the next Section we hint briefly to such a neural net, $\Psi^D(X)$, called *dynamic perceptron* (See Fig. 11.4 (b_1-b_2)). It is characterized by an automatic pre–processing devoted to modify the net connection geometry,

depending on the correlations of each singular input — practically it is in continuous learning, not on the weights, but on the connection topology. This architecture was developed by one of us [42.45.48], as a partial implementation of some ideas of Thomas Aquinas's theory of intentionality. In any case, there are straightforward neurophysiological evidences of the so–called "dynamic receptive field" of neurons belonging to different sensory systems of mammalians that could find by the notion of "dynamic perceptron" their computational model, showing the informational relevance of such a strange behavior[3]. The dynamic receptive field has been observed in mammalian retina [53], auditory cortex [54-55]; primary visual cortex [56-57]. It was found that there exist subfields, some of which are activated only during 20-50 msec for a continual presentation of stimuli, and the combination of activated subfields varies even for a static presentation of stimuli. In primary visual cortex, it is well known that there exist neurons with orientation specificity. Another type of neurons, whose orientation specificity — i.e., a tuning — is dynamically changeable, was found in relation to the dynamic receptive field [56]. In this context, a classical receptive field can be reformulated as a spatio-temporal summation of dynamic receptive fields. The spatial summation is taken over an entire receptive field, and the temporal summation over a few hundreds miliseconds. Since the time scale 20-50 msec is almost equal to a "unit" of psychological time, the dynamic receptive field may be considered as a neural correlate of internal dynamics for the reorganization of mental space. Namely, the presence of dynamic receptive field suggests the presence of the process of dynamic re–modelling due to dynamic interactions between higher and lower levels of information processing [56-57].

[3] We thank prof. I. Tsuda of the Dept. Of Mathematics of Hokkaido University in Sapporo (Japan) for this personal communication about the relationship between the pre–processing of our "dynamic perceptron" and the neurophysiological evidence of the "dynamic receptive field" in sensory cortex. We are preparing with prof. Tsuda a specific paper on this topics.

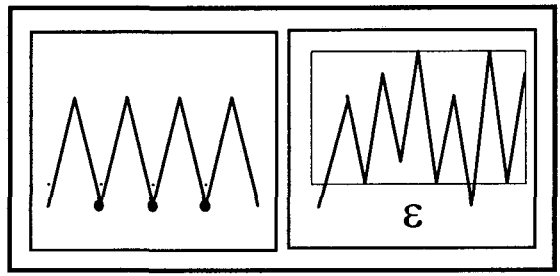


Fig. 11.2 The Difference between a Stable (periodic) [left] and an Unstable (aperiodic) [right] Time Series. An unstable time series can be intuitively defined as pseudo–periodic or *chaotic* if it can be characterized by recurrences that are periodic within a given interval ε.

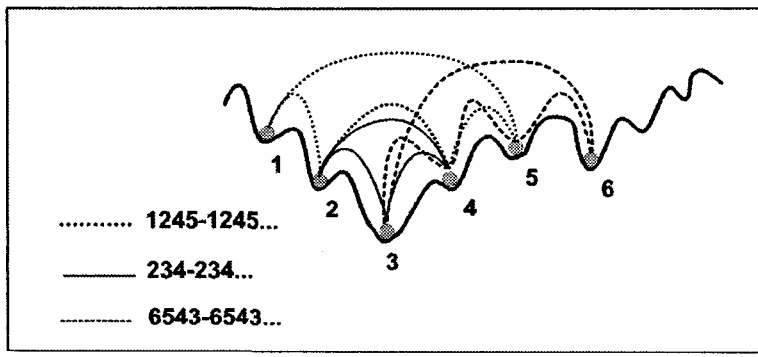


Fig. 11.3 Intuitive Representation of the Storing Capacity of a Chaotic Dynamics into the 2^n Pseudo–cycles among the n Minima of Its Energy Landscape. For instance, if we imagine that each minimum corresponds to a memorized feature of a visual object, it is easy to understand that each class of object corresponds to a cycle, i.e., a given combination of features. Moreover, by a simple phase change (e.g., a change in the ordering of minima within a give cycle) the net could easily recognize the sameness of the object also under three-dimensional rotation in the space. Finally, because we are faced here with pseudo–cycles and not with cycles, it becomes easy to explain also the physical basis of the phenomenon of similarity recognition (analogy) through such a dynamic structure of recognition.

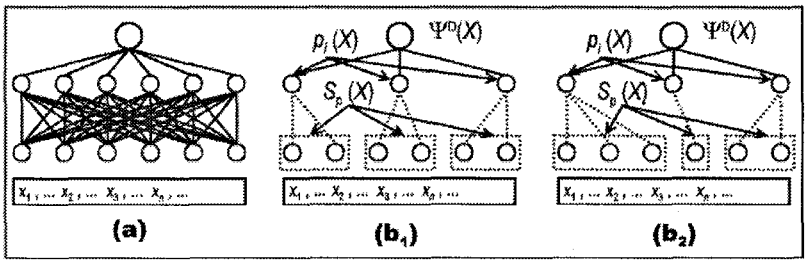


Fig. 11.4 (a) Fully Connected Topology of a Classical Connectionist NN (e.g., back-propagation) Displaying Its not Really Parallel Nature; (b1-b2) Dynamic Re-definition of Support in Dynamic Perceptron (see text).

11.2.2.3 *A First Conclusion*

In cognitive neuroscience it is generally held that a given neural circuitry is a *code* of some given perceptual *belief.* However, in the light of all the precedent discussion, we feel that any honest computational approach to the study of mind cannot limit itself to state simply that a given brain circuitry is a "code" of a "belief", i.e., of a mental representation of a given thing. At the actual state of development of the computability theory, *there is not* and *cannot be* any formal demonstration of this *threefold correspondence* among the *referential thing*, the *neural code* and the *belief.*

This sort of correspondence can be only a *matter of convention*, depending on the meta–language we choose to define this correspondence. Namely, this correspondence is only an *interpretation* in the technical sense of the model theory, just as to say that a given activated circuitry in a computer or a sequence of signals in a telegraph corresponds to the letter "A".

However, a distinction is necessary.

1. The problem of formally defining the *coreference* (i.e., to have the same reference) between a belief statement, expressed in intentional language (*I-talk,* e.g.: "I (believe to) see a red colour") according to intensional logic, and a related observation statement (*O-talk*), of some neurological (e.g.: "a modification in the variable y is measured at time t in the brain location z as a response to a given input x"), computational, psychological etc. theory, according to extensional logic, *is not a solvable problem* (See [25], pp. 132-134; [31], p. 116).

2. On the contrary, to solve the problem of the *real reference,* that is the problem of the correspondence between *a neural code* – not necessarily constituted according to a Hebbian law – and an *external thing,* it is sufficient to demonstrate that a biological brain is able to compute functions not computable for a TM, as opposed to Church's thesis. In other words, to solve the real reference problem for a scientific theory of perception it is sufficient to demonstrate that what characterizes a biological brain (and more generally any biological organism) is its *capability of redefining the basic symbols*, the codes, of its own computations, in dependence of singular different occurrences of their own objects.

To understand this point, we need a completely different approach to the real reference problem in the light of the pre–modern logic and particularly in the light of the classical Aristotelian–Thomistic theory of intentionality.

11.3 Intentionality and Foundations of Logic after Gödel

11.3.1 *Analytic versus Axiomatic Method in Logic after Gödel*

The preceding discussion about the core of the human intentionality is expressed in the language of cognitive science as the capability of human mind of re–defining the basic symbols of its computations (See pp. 258ff.). The double opposition between *inductive* versus *deductive* schematism (See p. 243) and *productive* versus *reflexive* thought (See p. 250), has a logical counterpart in the opposition between *analytic* and *axiomatic* method in logic. Namely, in the opposition between a logic defining its own role as *logic of discovery* of new hypotheses, and a logic reducing its role to the simple *logic of justification,* the logic of proving statements by deductive procedures, starting from fixed premises or *axioms.* Effectively, after Gödel — and, more recently, in the heterogeneous universe of the computer sciences — the necessity of studying logical procedures allowing change in axioms during calculations is an argument of ever growing importance. In fact, for contemporary logic, computer science and cognitive sciences there is the shared necessity of avoiding the multifarious limitation theorems that have their formal origins in Gödel's [58-60].

The interest for recovering to modern logic and modern sciences *the analytic method*[4] *of classical, pre–modern logic* depends on the fact that it is in principle impossible to allow axiom changes within formal systems. Following Cel-

[4] As we explain after (See note 7), "analytic" has here to be intended in a radically different way as to its modern sense, the sense used by Pappo, Descartes, Newton and Leibniz.

lucci's reconstruction, the historical origin of the analytic method is in Plato's logic and it consists in affirming that the premises of any deductive procedure consist in pure *hypotheses,* since it is impossible to attain the truth of any mathematical entity. Each hypothesis consists thus in a "step" toward the further, more general one in a never ended bottom-up process. The aim of logic would consist in the continuous progress toward ever more general principles, without the possibility of stopping such a process[5].

The historical origin of the *axiomatic method* is in Greek geometry and namely in the prototype of any axiomatic system: Euclid's *Elements.* It is based on the supposition that we can attain self–evident principles, without developing research toward more fundamental hypotheses. Further, Aristotle's logic transformed the axiomatic method into the proper object of logic, and proposed the axiomatic method in mathematics as a model for any other science. On the other hand, he refused the idea of a mathematical science of nature, typical of Pythagorean and Platonic traditions. Nevertheless, for Aristotle, the analysis still plays an important role as method of discovery of the so–called "middle–term" in any syllogistic procedure, that is the term connecting the "major premise" of the syllogism to its "conclusion"[6]. This use of the analytic method, related with inductive strategies, is functional to the axiomatic one. For Aristotle, analysis is a terminating procedure, or a "reduction" procedure, whose end is some new axiomatic definition — characterized by an immediate relation subject–predicate, i.e., axiomatic definitions are essence definitions — and/or some statement easily reducible to some axiomatic truth (See [61-62]. See also [60], pp. 291ff.). The construction of a deduction system following the axiomatic method in its syllogistic version within each scientific discipline constitutes the deductive "synthetic" moment, after the "analytic" devoted to the principle discovery. In this sense, the synthetic component has functions to make scientist's discoveries *rigorously expressible and profitable for all.* So, for Aristotle cannot exist one only axiomatic system for expressing all the mathematical truths or the true contents of any science. The analytic method for discovering new principles and finding new truths plays thus an essential, though subordinate, role in Aristotle's logical and epistemological theory.

---

[5] See, for instance Plato, *Parmenides,* 136c,1-7, *Letters,* VII, 342a-343c. The necessity of an infinite character of this process is however negated in *Republic*, VI, 511b6-8 where it is said that knowledge is a sort of ascension–descent through a sort of universal deduction tree. That is, knowledge is intended in *Republic,* before as a bottom-up process by the *resolution* (finitely *analytic*) method toward a final not–hypotetical principle, for re–descending thereafter to all the consequences through a top–down process by the *synthetic* (deductive) method. This program, at least for geometry, was effectively fulfilled by Euclid's *Elements.*

[6] See Aristotle, *Post. An.*, I, 22, 83b,39-84a,2.

In the modern age, the axiomatic method was established with important differences from Aristotelian teaching. The most important one was the rejection of axioms as "real definitions" or essence definitions, because of Galileian science self–limitation to *quantitative properties* of the physical things. This rejection was confirmed by Newtonian physics, vindicating *the absolute phenomenal character* of the new physics, and the *purely formal character* of the three laws of dynamics, as conditions for justifying the calculus and geometrical predictability of quantitative phenomena. The *Logique* of Port–Royal, re–proposing former reflections of B. Pascal, asserted the necessity for mathematics of using only "nominal definitions", by a separation between "definitions" and "existence assertions".

In modern mathematical logic and in Hilbert's formalism, the nominal character of definitions implied the rejection of Frege's logicism, by renouncing the necessity of supposing "truth" and "meaningfulness" of formal system axioms for maintaining only their *coherence*. "Truth", "meaningfulness", as well as "coherence" are metalogical properties of formal systems and must be metalogically checked by algorithmic procedures. A set of axioms is not *coherent* because it is *true* and *existent the objects* to which these axioms refer. On the contrary, because the set of axioms is coherent and its coherence can be proved by a finite recursive (algorithmic) procedure, they are also *true* and *existent* their objects. On this basis, Hilbert pursued the possibility of constructing one only formal system for all mathematics. He stated also the possibility of using the axiomatic method for the "logic of discovery", by supposing the possibility of an algorithm able to determine, for each statement expressible in a given formal language, whether is it demonstrable or not within this language. Church–Rosser theorem denies such an algorithm can exist in formal systems. Moreover, Gödel's incompleteness theorems for arithmetic and their extension to all formal systems in the work of Turing and Tarski, ruled out the idea that the notion of mathematical truth can be exhausted by any formal system.

In this sense it has been asserted that logic and mathematical systems must be *open* systems in which the analytic method must recover its ancient role as logical method of new axiom discovery [58-63][7]. In other words, the incom-

[7] This is true, though some distinctions have to be made with respect to the difference between: 1) Plato's and Aristotle's definition of the analytic method as bottom–up process for the definition of new hypotheses and/or as process for the definition of new axioms for making possible a demonstration; and 2) the modern definition of analysis, all depending on Pappo's definition of it, as top–down process of decomposition of a compound in its parts. Of course (2) cannot be reduced to (1). See [60], pp. 292-299 and pp. 349-351.

pleteness destiny for logical systems is unavoidable only *iff* we want to maintain *fixed* principles of demonstration, affirming that the formal systems are the *only* logical systems and the axiomatic method is the *only* method of logic. The *logic of discovery,* the logical method for new principle detection for the continuous construction of scientific (demonstrative) procedures, is the most important part of logic, *since only this type of non–determinism can avoid undecidability spectres.*

According to Cellucci, the reasons for which the discovery of limitation theorems for formal systems can be interpreted as a necessity for recovering the analytic method in its early Platonic version (see note 7) against the monism of the axiomatic method are very deep. They are essentially three:

1. *For avoiding the incompleteness*, it is not sufficient to construct a series of formal systems, each obtained by adding as new axiom the undecidable proposition of the precedent one. Indeed the main question is whether there are complete formal systems successions obtained in such a way. The answer is very limited and substantially negative. Formulas of the type $\forall x$ A(x), where A(x) is *a decidable* property, are demonstrable, even though there cannot be an algorithmic (finite) procedure for deciding the truth of the formula $\forall x$ A(x). On the contrary, formulas of the type $\forall x \exists y$ A(x, y), where A(x, y) is a decidable relation, are not demonstrable, though they are true in the system [63]. More generally, the solution to Gödel's incompleness theorems for formal systems cannot consist in a series of systems chosen through an effective procedure. Some sort of non–determinism is necessary in the construction of the systems and hence in the construction of the axioms.
2. *The only non–determinism sufficient for avoiding Gödel's incompleteness* in formal systems consists *in the introduction of new axioms* and not in the simple possibility of non–deterministic multiple choices [64]. "The non–determinism required by Gödel result is the non–determinism related to the possibility of introducing at each step new axioms in a non algorithmic way" (See [60], p. 326). This denies that the system might be considered "formal" in classical sense and that the method used might be axiomatic — or "analytic" in modern sense (Gentzen's natural deduction methods included. See note 7).
3. *Turning to "mathematical intuition"* for justifying the discovery of new axioms, as Gödel himself did, implies a double unpleasant consequence. Before all, it means that there must exist ultimately in logic and mathematics (and hence in any science, metaphysics included) an irrational,

subjective component [65-66]. "Here we are not in the realm of science, but of poetry" [67]. On the other hand, also if we intend the "intuition" in the strong sense of Gödel's "ideal intuition" of abstract concepts — that would be relative to infinite mathematical objects and that would be the result of a difficult training of the mathematician — we are faced with unavoidable limitations. The certainty thus obtained is unusable for granting that *certainty* in concrete mathematical choices we are searching for. For instance, let us suppose that in the system *S* there exists a true formula *A*, undecidable (i.e., both *A* and non-A cannot be demonstrated in this system) and a given abstract concept of set, $\Im$, known for intuition and for which the axioms of *S* are equally true. For a corollary of the first incompleteness theorem of Gödel, there must exist also another formula *B* and another intuitive notion of set, $\Im'$, for which the axioms of *S* are equally true, but such that *B* is *true* for $\Im'$ and false for $\Im$ (it is sufficient that we pose as *B* the statement not–*A*). In this case the ideal intuition cannot be used for deciding whether $\Im$ or $\Im'$ is the correct set notion. An examplification of such a case is whether we pose *S* as the Zermelo-Franklin set theory (ZF), $\Im$ as ZF notion of set and $\Im'$ as Cohen notion of set, after his demonstration of the independence of continuum hypothesis from the axiom of choice [70]. It is thus evident that ideal intuition, in spite of its implicit reference to infinitary method of demonstration[8], cannot grant that absolute mathematical certainty which formal systems are searching for (See [60], p. 254).

However, the reference to *infinitary methods* in mathematics, which can abstractly grant coherence and truth but with weak effectiveness, can offer another contribution to a better understanding of Plato's analytic method limitations. Plato's impossibility of reaching mathematical truth and his preference for hypotheses and not for self–evident axioms are both based on an *ontological assumption*. This assumption is in many senses equivalent to the core of Tarski's demonstration of impossibility of defining semantic notions such as truth, coherence, reference, etc. by a purely formal recursive procedure of satisfaction, without attaining formal languages of ever higher logical types (see above § 11.2.1.2 and note 1). The common ontological assumption here con-

[8] It is to be remembered that G. Gentzen demonstrated the coherence of number theory by extending the mathematical induction till ε_0 [68]. In this way he explained why Hilbert's finitary arithmetic cannot give a similar demonstration of number theory coherence, so to satisfy Gödel's second incompleteness theorem.

cerned is summarized in the following quotation from Plato's Dialogue, *Parmenides*:

> (If you pose the essence and the existence of a given object), you have to examine *simultaneously all* the consequences of such a hypothesis, both with respect to the object itself, and with respect to each other object individually considered, as well as with respect to whichever collection of these objects and with respect to the other objects considered all together. (...) This task is never ending (*Parmenides* 136c, 1-6).

It is evident that this is the same problem identified by Tarski for demonstrating the undecidabilty of semantic notions in formal systems because of the necessity of posing "infinitely many notions of satisfaction that must be introduced simultaneously because they cannot be defined independently" (see § 11.2.1.2 and note 1).

The solution suggested by Cellucci for avoiding this limitation would recover in a post–modern (post–Gödel) way the core of Plato's analytic method. Effectively it results very close to M. Minsky's "society of mind" [69]. It consists in supposing many systems connected together in a variable way, without, of course, because of Gödel's, Church's and Turing's limitation theorems, any possible formal rule and/or algorithmic procedure governing this variation and/or the choice among the systems. Logically, it means to attribute only *a hypothetical existence and essence* to the different logical objects expressed in the different axiom collections, available in this way. The advantage of such an approach is that, using different discovery inferences — overall "induction" and "analogy" —, this method chooses in the dichotomy characterizing the so–called "inference paradox" (either *certainty* or *knowledge amplification*) the second alternative.

In other words, Cellucci concludes, we are faced today with a constraining alternative.

1. From one side, there is logic following *the axiomatic method* that not only gives no knowledge amplification, but it is no longer able, after Gödel, to grant absolute certainty. So, to avoid contradictions, this approach weakens the strength of logical implications (See, for instance, P.J. Cohen's "generic set" theory [70]), demonstrating only generic and therefore *useless* propositions.
2. On the other side, we have *the analytic method* that, without granting certainty, is able at least to amplify knowledge by providing, new hypotheses for demonstrating useful propositions. It may be obvious that

post–modern logic has to choose the second alternative: if it is possible to have logical systems only with a local and limited demonstrative power, at least they showed be able to demonstrate useful propositions! This is the reason for Cellucci's preference for the analytic method.

Nevertheless, it is hard to understand how different this solution is from the one criticized by Cellucci himself and appealing to subjectivity of mathematical intuition. In the next Section we want to suggest another strategy, which recovers the Aristotelian–Thomistic integration between the analytic and the axiomatic method.

It is useful to conclude this subsection by recalling the result of this overview about the opposition of the two logical methods (the analytic one and the axiomatic one), in the light of Gödel's theorems. This result is that the true problem is only one: how to grant an "open character" to logical systems, that is, *a procedure making a logical system able to change its axioms* to avoid undecidable situations in formal systems. In this light it is easy to understand that this is the same problem we faced in cognitive science, discussing H. Putnam's approach to intentionality problem (See § 11.2.1.5). Not casually he is the more trained in foundational questions among cognitive scientists. Because the notion of intentionality was introduced and discussed for the first time in western thought by Scholastic philosophy in the Middle Ages, it gives us useful suggestions for a solution to the related problems of *intentionality* in cognitive systems and *openness* in logical systems.

11.3.2 *An after Gödel Reconsideration of Thomas Aquinas' Theory of Logic*

To comply with this double unique problem, we deepen Thomas Aquinas's[9] logic, in its more original suggestions with respect to Plato's and Aristotle's logical theories.

In addition, according to Cellucci's reconstruction, it is difficult to find in modern logic what we need. Gödel's theorems constrain modern logicians "to see beyond modern age", both forward and backward. If we want to build a "post–modern" age that is not the irrational realm of "weak thought", we have to solve the problem of *the logic of discovery,* without falling into a purely subjective approach to the problem of axiom change in deductive systems. Given the prevalence of the axiomatic method in modern science since Des-

[9] Thomas Aquinas (1225-1274) was an Italian philosopher and Theologian, who lived and worked between Paris' (France) and Naples' (Italy) universities, during the first half of XIII Century.

cartes, Galilei and Newton, for a post–modern approach to the logic of discovery, we are obliged to search for it within the "pre–modern" age, without any Enlightenment preclusion for the Middle Age.

In this research of a suitable logic of discovery, we stopped with Cellucci by the classic analytic method. We agree with him in emphasizing the strong distinction between the use of analytic method in Plato and in Aristotle. We agree also in affirming that all the differences ultimately consist in their opposite approach to the notion of "essence" knowledge. For Plato, this knowledge is ultimately unreachable; for Aristotle it is something available by a process of abstraction–intuition. His axiomatic method in formal logic depends precisely on this, so that each principle of a categorical demonstration by his syllogistic method consists in an "essence definition". Formally this definition consists in *a non–tautological identity*.

To modern people affected by the Galileian–Newtonian refusal of the "essences" it is sufficient to recall that by "essence" both Plato and Aristotle intended *the infinite totality of relations* making each thing[10] identical with itself and different with respect to other things, or collections of "things" (see note 10), in the universe. It is evident that, when we are faced with the problem of dealing simultaneously with infinitely many satisfaction relations, as in Tarski–Gödel formal theory of semantics (See note 1) — apart from different words and cultural contexts —, we are effectively faced with the very same problem of truth as "*essence* knowledge" of our Ancestors. In both cases, however, what is effectively being discussed is a valid justification of universality and necessity in logic. This is the problem of *certainty* in scientific knowledge.

Generally, in the history of philosophy of the Middle Ages, Thomas Aquinas's philosophy is considered as an original synthesis of Platonic and Aristotelian traditions. The core of Aquinas's originality is the doctrine of *real distinction between essence and existence* in the notion of "being" (See note 10). According to Thomas Aquinas, the error of both Aristotle and Plato in dealing with the essence problem — and hence with the justification of universality in logic — consists in not distinguishing adequately between *essence* and *existence* of a given thing.

For Plato essence and existence are not really distinguished: the essences

[10] From now on, for sake of simplicity and clarity, we name as "thing" each existent being (whether it is a "substance" or an "event", or a "relation" or a "quality" or a "quantity" or a "collection" or whichever else). This denotation is aimed to avoid confusions with the term "being", from now on intended exclusively as nominal form of the verb "to be".

exist as immaterial individuals and the "being" of each material "thing" is only a limited participation to this ultimate way of existing. In fact, he — with the greatest majority of Western logical and mathematical thinkers — established the existence of a given individual by the satisfaction of the formal relation of self–identity. It is intended as negation of any negation of identity, i.e., as negation of any qualitative difference[11] with everything else, and hence by *an actual infinity of relations* (See [71] 185a; [72] 139b-e; 146a-147b). In this way, only the immaterial essences fully exist as individuals. Through this opposition between what is relative (the quality) and what is absolute (the essence), the ultimate being and truth of each thing become for Plato unknowable. The knowledge of essence would require *the simultaneous exhaustion of all the infinite qualitative differences* from which only the absoluteness of self–identity and individuality of a thing can emerge.

Aristotle tried to solve Plato's problem in two steps:

1. by distinguishing between *substance,* intended as individual existent thing, and its *essence* that could be common to many individuals, so to deny that essences are existent individuals on their turn, belonging to some immaterial world,
2. by putting in the *material* constituent of any essence the root of the differentiation process, both of the different essences and of different individuals, sharing the same essence.

For Aristotle, the knowledge of essence becomes possible for humans in this way. Knowing an essence does not mean, as for Plato, dealing *actually* with an infinity of relations, but only each time with a finite totality of relations, since all the other ones exist only *in potency*, hidden in the common indefinite material substratum of all the things. It is evident that Aristotelian ontology is perfectly coherent with his treatment of the analytic–synthetic method in logic as heuristic component of an overall axiomatic method[12]. If we applied Aristotelian ontology to our post–Gödelian problems in foundations, we would obtain

---

[11] The qualitative difference, is that allows to say that a given individual is something and is *not* anything else. To say it in extensional logic (class theory) terms, any class must to be close to other classes, that is it must contain as null–class the class of all the elements not belonging to it. This necessity of negative definitions for consistent logical constructions is the ultimate formal root of all the logical antinomies.

[12] Namely, this ontology explains why Aristotle's application of the analytic method for the discovery of the lacking "middle–term" for constructing a syllogistic demonstration (see § 0), is only a reduction procedure. That is a procedure always terminating in some universal statement directly derivable from some main axiom, as far as this statement was "implicit", "hidden" or "potentially existing" in it.

at most very weak results. It would be possible to preserve a formal system despite its inner incoherent statements — whose presence is granted by Gödel theorems — as long as it is possible to maintain these statements "implicit" or "in potency". In other words, Aristotelian ontology is supposed in any modern attempt to solve logical antinomies by weakening the strength of the logical implication (See § 11.3.1).

So, what we need is that, from one side, instead of being hidden or existent in potency in some universal collection, the relations not concerned in some effective calculation and/or demonstration *do not exist* at all. Nevertheless, *universality* could be granted if the essence of a given object, instead of being conceived as the simultaneous existence of an infinity of relations, was conceived as another primitive besides relations irreducible to them. Universality could be thus granted if we were able to attribute to essences the *capability* (of course passive, i.e., relative to an active power as it is in any causal relationship) of generating relations, each time it was *necessary* for converging in calculations and/or for making a demonstration effective. The astonishing plasticity of human brain and of human cortex in redefining continually its finite connection topology, with rapid responses to impinging inputs, without loosing time in combinatorial searches, would be a limited but efficacious neurophysiological "icon" of such a metalogical and metaphysical idea.

On the other hand, how could we wish, without falling into subjectivism, to make "open" the logic systems, and simultaneously pretend not to insert as primitives of these "open" systems logical objects with the power of *generating* other logical objects? What we need is an ontology able to make the existence of infinite logical relations *virtual*. The relations and the terms they connect have to be conceived neither as existent "in potency" nor as existent "in act". They have to be conceived as *virtually existent*, i.e., *relatively to some "principle" with the power of making existent* a different subset of the infinite totality concerned, for each different concrete context, for satisfying universal logical laws. Only at this price it is possible to give back to a post–modern logic of discovery all the rigor of *logical* method. That is, to make this method a set of logical rules deriving in a strong deductive sense by universal logical laws, without any concession to irrationalism.

Thomas Aquinas ontology is useful at this point. For different historical motivations from ours (more theological and metaphysical than metalogical), both his metaphysics and his logic are based on the definition of causal principles for the existence of each thing (both physical and logical or linguistic) belonging to the universe. From one side Thomas' ontology accepts Plato's

instance that universality depends on essence and on its capability of embracing an infinity of relations. On the other hand, it accepts Aristotle's instance that, for each concrete, individual application, only a finite subset of the infinity is effective, even though this subset is always changing. Nevertheless, if on one side Thomas criticized the unattainable character of truth in Plato's philosophy, from the other side he negated Aristotle's solution of distinguishing different senses of existence — "in potency" and "in act", with a continuum of intermediate states — was logically and ontologically consistent. Particularly, he criticized Aristotle's justification of existence contingency (i.e., the "being–in–potency" of a thing) as a supposed "indifference to being and not–being", because violating non–contradiction principle (See [73] n. 184; [61] pp. 50-69). In fact, Aristotelian ontology can grant at most a non–determinism in choosing among a set of alternatives already fixed — i.e., existent in potency in some universal substratum. But we have demonstrated that this is insufficient in principle for avoiding the limitations related with Gödel theorems.

Aquinas's solution is more radical than Aristotle's. From one side, he distinguishes another sense in the notion of "being" absolutely different from those identified by Aristotle. The different senses of being identified by Aristotle, making the notion of being an "analogical" ("multivocal") and not "univocal" notion, are only *different modalities of existing* (necessarily, contingently, potentially, actually, etc.). When we speak about essences we need another sense of "being" distinct from all senses of "being" as "existing" in the different modalities detected by Aristotle. This sense is related to the use of the copula "is" in the construction of elementary definition statements (See [61] I, v, 71ff.; [74] II, 23). For instance, when we say that "the phoenix is the bird reborn from its ashes" we are saying nothing about its existence. Similarly, to define "the runner" as "who (or what) runs" says nothing about the existence of somebody (or of something) effectively running. In the construction of definition statements we are dealing with the "beingness" (*entitas*) of the object concerned, with "*what* it is" not with its "existence" (*existentia*), with the "*it is*" of some "what". When in a realistic epistemology we speak about "real reference" of a given elementary (subject–predicate) statement, the "being" the statement is referring to, is properly the "beingness" of the object concerned, not its "existence". It is thus evident that definition statements do not catch all the essence of the object but only its "whatness" (*quidditas*), a finite subset of relations to distinguish the object within the finite semantic context of a given linguistic occurrence. E.g., Aquinas said (See [75] II,vii,472-475), defining humans as "rational animals" is sufficient to distinguish them both from immaterial things and from non–living bodies, as well as, among organisms, from plants and irrational animals.

But this definition is not able to fulfil all the human essence. In some case, we could be constrained to adequate this definition, by coming back to human "beingness" to "pick up" some other quality from the human essence. For instance — to give a modern example of this ancient idea — if some extraterrestrial individual arrived on the earth and fulfilled the definition of human "whatness" as "rational animal", it would be necessary to change such a definition to avoid an undecidable situation. That is, we would be constrained to come back to human "beingness" to extract from human essence other properties to improve the discriminating power of our human "whatness" definition. So we could define humans as "*terrestrial* rational animals", a definition absolutely redundant in the actual context where no E. T.s are officially discovered! It is evident that we are faced with the realistic epistemological counterpart of what in the precedent subsection we defined the discovery of new axioms for solving undecidable situations.

Aquinas explains (See [76], VII, 2, ad 1) that we can properly use in logic the meta–predicate "to exist" — as coextensive to the meta–predicate "to be true" (See [77], I, 1c) — only after having properly constructed the "whatness" statement that is argument of these meta–predicates (See [78], VIII, II, 1d, ad 1). By this rule, we can state, for instance, that "it is false that 'the phoenix is the bird reborn from its ashes' and therefore (this sort of) phoenix does not exist". On the contrary, "it is true that 'the phoenix is the *mythological* bird reborn from its ashes' and therefore (this sort of) phoenix exists". It is evident that, besides the extensional sense of being as *existence* with all its modalities — whose linguistic counterparts are object of several modal logic theories —, there is another purely intensional sense of being. Aquinas defines it as "beingness" (*entitas*), or "essence being", being–of–the–essence, for distinguishing it from "existence" (*existentia*), or "existence being", being–of–the–existence. The former is involved in all the answers to the question "what is it?" (*quid est*), the latter in all the answers to the question "is there?" (*an est*). But for Aquinas it is possible to answer the second question *only* after having answered the first one. In this way, it is easy to solve all linguistic paradoxes related to the use of negative terms, such as "liar paradox", in as many confused uses of "being" notion as "beingness" or as "existence". A typical case concerns the theological paradox of the "existence of evil", as far as the beingness of evil consists in a "privation of being", i.e. in a privation of some qualities characterizing the beingness of a given thing. Though "evil is a not–being", nevertheless "it exists" in given contexts (e.g., physiologically as sickness, morally as sin, physi-

cally as natural disaster, etc. See Thomas Aquinas, [79], III, 7).

This solution of the formal and semantic logical antinomies does not involve any "type theory", either in Russell's "ramified version" — attaining languages of higher logical order — or in its "simple version" — avoiding higher order logic, by supposing in the meta–language an appropriate choice of primitive terms [80]. By contrast, Aquinas's metalogical distinction between existential and essential (definitory) statements is based on a constructive logical theory of the essential statements directly depending on his metaphysics.

The main consequence of this logic is however the following. Because of his strong distinction between beingness and existence, Aquinas can change the logical notion of *identity* in a fundamental point. Two existent things are identical not because they are "the same thing" (two individuals cannot be at all the same one) but because they have *one only essence.* In Aquinas's logic, the symbol "=" interposed between two equiform tokens (in formulae of the type "$a = a$") or not equiform tokens (in formulae of the type "$a = b$"), cannot be metalinguistically interpreted as a sign that the two symbols it connects "are denoting the same thing" (e.g., two equivalent classes, if they are class symbols). The equality symbol has to be interpreted as a sign that the two symbols it connects are "referring to *one only essence,* even though denoting two distinct things", also when the two symbols are equiform (See [81], V, xvii, 1021). Where "reference" in Aquinas's semantic theory is a *constructive operation* and not a binary relation. That is "f refers to e" means that "e constitutes f as true", where e is (an essence determining the beingness of) the denoted object and f is a formula of a given language.

What in the classical axiomatic approach is a contradictory formula could become here only an equivocal formula. It could be possible indeed to find appropriate conditions under which to attain the essence for generating new symbols to remove the ambiguity. This is the main property of an open constructive theory such as Aquinas's[13]. E.g., if I say "Andrea is a man" and "An-

[13] "It is not sufficient name identity with the difference of the thing that it denotes: this brings *to equivocating* (not to contradicting)" (See [73], I, ix, 116). This has for Thomas an immediate consequence for mathematics, as to numbers applied to concrete measurements and/or calculations: "A number, as far as existing in numbered things, is not the same for all, but different by different things" (See [82], I,10,1c). E.g., the "two" used for numbering "two horses" is not the *same* "two" used for numbering "two mosquitoes". The two "two's" share only the essence of "two–ness", as their common generating principle. This means that in mathematics, by denoting the "two's" with the very same digit "2", I am referring to the common essence of "two–ness", but for denoting in my applied calculations two different existential instantiations of the same essence. If I pretend to use in my applied calculations the same instantiation, they cannot converge to a solution.

drea is not a man", the contradiction is a simple ambiguity if by "man" I am always referring to Andrea's same humanity (his beingness). But in the first formula I am denoting Andrea as a living body and in the second Andrea as a corpse. We have ultimately and inevitably contradictions *iff* we fix the axioms (and/or the definitions), that is if we pretend that identity expresses sameness with respect to existent things.

Allow us to put this in terms of modern computability theory. Against axiomatic method, to perform effective, concrete calculations, it is inconceivable to suppose *one only axiomatic system of natural numbers.* In this connection, Gödel's theorems result a confirmation of this ancient foundational idea. The *equality* token in arithmetic and, more generally, in logic and mathematics, cannot mean "sameness" with respect to existing things!

From the axiomatic system theory standpoint, following A. L. Perrone [42-43], a theory of logical foundations incorporating Aquinas's distinction between "being as essence" and "being as existence" is characterized by two kinds of primitives, *essences* and *relations*, and not only one, the relations, as in modern mathematics [83]. The foundational theories proposed during last years by E. De Giorgi and his colleagues are a useful approximation to this idea. Also they recognize another kind of primitives besides relations: the "qualities" [84]–[86]. Qualities characterize each object belonging to the theory: e.g., there exists "the quality of being a set", "the quality of being a relation", "the quality of being a natural number" and so on. Also "'the quality of being a quality' is a quality", where this self–referential properties of qualities is granted in that axiomatic language by allowing the graph of relations among all the objects belonging to the universe of this theory to remain partially undetermined.

In other words, also for De Giorgi the way to avoid inconsistencies is to allow the demonstration of only *generic* propositions. This foundational approach is biased by the Platonic prejudice of considering the "qualities", like Plato's "essences", as existing objects in a purely *extensional* sense. That is, they are reciprocally distinguished by their property to determine different collections within the Universal Collection *V.* For these collections, *the extensionality axiom* holds, that is, two equivalent collections are the *same* collection[14]. This purely extensional definition of identity emphasizes that De Giorgi's "qualities" do not differentiate another realm of being besides the (relational) existence like Aquinas's.

---

[14] In Aquinas's as in ours approaches, two equivalent collections are not he same, they refer to (i.e., they are generated by) the same essence.

Aquinas's essences are, on the contrary, absolutely *monadic*. The only relation they have, in Perrone's formalization, is each one with itself. This relation however is not *self-referential*, so to determine the *existence* of the relative collection in V. Their auto-relational property is only to emphasize that they constitute the ultimate anti-predicative level in any chain of predicative definitions. In other words, they emphasize the only proper level at which identities occur. In this way, in Aquinas's approach, self-referential expressions are strongly prohibited for essences (See [74], II, 23; [81], IV, viii, 649). As it is inadmissible to say that "the race (intended as 'running') runs", one cannot say that "the essence of being an essence is an essence". Hence it cannot exist as an individual or as a thing. "It is" as a generating co-principle entering into the ontological constitution of each existent thing[15]. In this logic collections (classes, sets, families, etc.), both finite and infinite, are thus to be conceived as *evolving* objects. They do not contain as existent all their elements, but they contain virtually all the things that can be made progressively existent (generated) according to given modalities (See [88], 113; [82], I, 18, 4 ad 3; [89], III, 11, 385).

11.3.3 *An Application to Foundations of Arithmetic*

Here, we are faced with the "dynamic" character of the collections (sets, classes, etc.), because they are made able to enrich themselves of new objects, as far as the conditions making necessary the existence of new element(s) in them occur [42]-[43]. This implies the definition of a "dynamic" counterpart "$\overset{\wp}{=}$" of the "equality" relation "=" because two distinct things now can be posed as "equal" with respect to a given operation r on which the equality "$\overset{\wp}{=}$" is defined (see Thomas' quotation in note 15).

The main axiom of this foundational theory concerns the existence binary operator $\dot{\exists}$ whose action consists in making existent a given object x within the universal collection V, $x \in V$, by applying itself to the essence of x, Ex, every time the conditions c making necessary the existence of x occur, i.e., $c = 1$ (See [43], 270):

> **Axiom 1**: $\dot{\exists}x \overset{\wp}{=} \dot{\exists}(E,x,c)$ is an existence binary operator. It applies to the essence Ex and gives the object $x \in V$ as existent ($\exists x$) or non-existent ($\sim\exists x$), depending on conditions necessitating the existence of x through

[15] "The essence has not directly the existence: it passes to existence through some individual thing to which only existence pertains because the producing action terminates onto it" [87].

the operation r on which the equality $\overset{\wp}{=}$ is defined. These conditions are summarized in the value of the constant c, i.e., *iff* $c = 1$ the passage from Ex to $\exists x$ occurs, otherwise $c = 0$.

In this foundational theory, the *non–contradictory* character of a given formula is insufficient for granting its *truth* and *existence* of the relative object. I.e., it is not true that: $\forall x \neg(\neg P(x)) \Rightarrow \exists x P(x)$. This differs from intuitionistic mathematics, however, because in the intuitionistic approach the existential operator acts only if there exists already an effective calculation for the single x value. In our approach, there exists in principle the possibility to construct an effective procedure for calculating what we need in each given condition.

This depends essentially on the possibility of defining a relation of dynamic equality between natural numbers defined as successors on different axiomatic arithmetic's[16] $g_i, g_j \in G$, where G is the collection containing virtually all the axiomatic arithmetic's. For obtaining this result it is sufficient to define the successor relation S as follows ([43], p. 272):

Axiom 2: S_g is a binary relation. It is defined as follows: $\overset{\wp}{\exists} i, j : \forall x_{g_i}, y_{g_i} \in \mathbb{N}_{g_i}; \forall x_{g_j}, y_{g_j} \in \mathbb{N}_{g_j}$ such that the following holds:

$$S_{g_i}(x_{g_i} + y_{g_i}) \overset{\wp}{=} S_{g_j}(x_{g_j} + y_{g_j})$$

All this means that the integers i, j, or better the correspondent axiomatic theories of natural numbers g_i, g_j belonging to the collection G there exist ($c = 1$), *iff* there is a relation $S_{g_i}(x_{g_i} + y_{g_i}) \overset{\wp}{=} S_{g_j}(x_{g_j} + y_{g_j})$ to be fulfilled. In other words, the collection G *evolves* by specifying its own elements g_k depending on the *necessity* ($c = 1$) imposed by the relations to be fulfilled. In the case that the two axiomatic theories are the same g_i, the following holds:

Lemma 1: $S_{g_i}(x_{g_i} + y_{g_i}) = x_{g_j} + S_{g_j}(y_{g_j})$

Demonstration: it is sufficient to consider the Axiom 2 by positing $i = j$ allowing Peano's classical successor.

---

[16] We remember that a corollary of Gödel's first incompleteness theorem is that does not exist and cannot exist one only axiomatic arithmetic in which it is possible to demonstrate all the true arithmetic propositions.

In other terms, Peano's axiomatic arithmetic is a subset of this "open" arithmetic, given the successor operator defined on one only "closed" axiomatic system. Following Perrone's demonstration it is possible to see how "open" arithmetic, by such an operational version of Aquinas's ontology here briefly discussed, can be interpreted as a collection of axiomatic systems. Effectively, they are a collection of arithmetical systems "in progressive construction". The construction of each is governed by rules, satisfying a semantic interpretation of universal laws of logic, even though these rules are not "algorithmic" in classical Church–Turing sense [42]-[43].

Moreover, it can be demonstrated that the recursive functions constructed by such a "dynamic" approach are defined within different axiomatic theories of natural numbers. Such functions are endowed with a higher computational power than the partial recursive ones $\phi(x)$ that are defined not for all the values of x [20]. Partial recursive functions allow only recursive calculation schemes characterized by some *aleatory definition* of the codomain, as in the *non–deterministic Turing Machine*. On the contrary, in Perrone's approach, the relation defined in Axiom 2 grants that the choice of the number succession on which the function develops its computation at the next calculation step is not aleatory. It depends on what we have to calculate (the input), and within which conditions. On this basis, with further axiomatic constraints that we cannot discuss here (See [43], p. 276f.), it is possible to demonstrate that such recursive functions $\Psi(x)$ are *virtually* general recursive. I.e., they are defined on *all* x values (they are not partial), even though such a definition is not given *simultaneously*. That is, they cannot be general recursive in the classical axiomatic sense. Gödel theorems prevent a diagonalization procedure for a general recursive function defined by only one "closed" axiomatic system of natural numbers. However because they are ranging on "open" systems, they are general in the sense they have the power of being defined on all the domain, even though, each time, *only the part of this domain necessary to conclude an effective computation* is given.

Perrone has developed several applications of these foundational ideas in different fields of computer science. They concern:

- *Automatic pattern recognition* in high energy physics experiments by the application of "dynamic perceptron" scheme (see § 11.2.2.2) [48].
- *Chaotic systems* characterization based on an effectively computable technique of the pseudo–cycles of any length [42], [43], [45].
- *Data compression techniques* based on the possibility of a "dynamic quantization" of the coefficients of the mathematical transform (wavelet, DCT, etc.) used [44], [90].

11.3.4 *Thomas Aquinas' Theory of Intentionality*

Quoting Putnam's work about *The meaning of "meaning"* [30], R. McIntyre [91] rightly emphasizes the core of any realistic theory of intentionality as to the problem of real reference. In any extensional and/or intensional approach to the problem of reference, *logical domain* of a given symbol determines the object. On the contrary, in a realistic approach *the real object must determine* the logical domain of a symbol. We gave different illustrations of such an idea in this paper, producing evidence both from the theoretical and from the experimental standpoints to sustain it. The preceding statement, however, recovers the core of the notion of intentional reference of Aquinas, which differs from the modern treatment of this notion.

In light of the previous discussion about Aquinas's approach to logic foundations, we have abandoned the idea that the logical notion of *reference* can be interpreted as a *logical twofold relation* between names and real objects. Generally, a logical relation has always its *reciprocal*. E.g., if $A = 2B$, then $B = ½ A$; if $A \geq B$, then $B < A$, if A causes B, then B is an effect of A, etc. On the contrary, it is well known that reference relation is without reciprocal: if A refers to B, B is not referring to A. It is related to B by some other relation. In Aquinas's foundational theory this picture is further complicated by the fact that the relation linking the object to the symbol referring to it is a *symbol constitution operation*, of which logical and ontological "machinery" we discussed in the previous subsection. Let us see the same idea from the epistemological standpoint.

As Aquinas emphasizes (See [77], I, 1; [78], b. I, XIX, 5, 2 ad 2um; [81], V, xvii, 1027), for granting real reference, we must consider the referential object *constituting* the symbol that *refers to* object "beingness" and hence that *names* the object. Particularly, (See [79], II, chs. 12-15) we must consider the referential object to be what *makes existing* in a logical sense the *true* proposition naming it. For instance, following Aquinas, the proposition "the sky is blue" is a true logical symbol of the object I am observing *iff* the blue sky I am *actually* observing is able to modify both the extension and the intension of the predicate "being blue", so to include in its domain the singularity of this object with its absolute novelty. But the reciprocal of such an act of constitution does not hold. In fact, if for any reason we pretended to designate the same object by the false proposition "the sky is yellow", the blue sky is not made yellow. Logically, we are thus constrained to say that the reference is neither *a logical nor a causal twofold relation,* but a *metalogical operation* of symbol constitution by

the referential object.

We know from the discussion above that the referential object here concerned is not the "existent thing" but its "essence" in its being a constitutive metaphysical principle of the *beingness* such an existent thing. By this idea of reference as an operation of logical constitution, we can understand how for Aquinas the same object in different contexts – effectively, the same essence in different existential istantiation — will modify the universal symbol that designates it as a "one – to – one universal"[17], i.e., as a "rigid designator". Psychologically, such a logical operation corresponds to the famous theory of truth as *self – conforming* (*adaequatio* in Latin) of the intellect to the thing. The knowing act is the operation by which the senses and the intellect inner state *assimilate themselves* continuously to the changing referential thing. It is not the thing that must accommodate itself to the *a priori* of human minds, as in modern approach after Kant, but it is the *a priori* of the human mind modified continuously to make itself adequate to the referential thing. The domains of the predicates are not constituted *a priori*, but they are constructed step by step for including symbols designating new objects and/or states of affairs. The "logical machinery" of such an epistemological and psychological theory of intentionality could now be more intelligible in the light of the previous subsections.

Because of the formal justification of real reference as *a constitution operation* (generation of symbols) and not as a simple asymmetric relation, it is possible to define formal languages as "top – down" (to the referential object) and not "bottom – up" (to higher order meta–language) *semantically open*. In this way, one can imagine also new promising approaches both to the problem of inductive schematism in cognitive psychology of perception (see § 11.2.1.2) and to the two difficult formal problems with which both computer science and cognitive neuroscience are today faced (see § 11.2.2.2). I.e., the problem of an effective mathematical characterization of unstable and non - stationary dynamic systems and the problem of really parallel computation in natural and artificial NN's.

11.4 Conclusion

In this paper we deepened the relationship existing between the intentionality problem in cognitive science, and the problem of foundations in logic and

[17] A "one–to–one universal" is a name that designates universally a singular object. Classical universals of such a type are the proper names.

computer science. The main result of this research is the necessity of overcoming the "axiomatic ideology" both in logic and computer science for allowing the construction of "open" logical and mathematical systems. In this way, also the problem of simulating intentional behavior can hope to find a solution in cognitive science. What is indeed characteristic of the intentional mind is its capability of "changing the basic symbols" of its logical computations for locking itself onto the changing reality. This idea recovers the essential of Aquinas's approach to foundations of logic as well as to intentionality problem.

In other words, Immanuel Kant's philosophical "Copernican Revolution" placed human intellect not the object at the center of modern science construction, just as Copernicus placed sun in the center of the solar system instead of earth. This philosophical and cultural revolution was justified by the wondrous victories of Newtonian calculus and of the axiomatic method after Descartes. Euclidean geometry became the paradigm of the new Galileian science. The evolution of numerical calculus; the necessity of overcoming the "fixity" of classical axiomatic method and hence the "stupidity" of actual computers, as well as the necessity of not abandoning logic and mathematics foundations to the "weakness" of subjectivism, all this imposes today a counterrevolution. This revolution however is not and cannot be the counterpart of a return to Ptolemy in cosmology. Einstein's cosmology discovered that universe has no center, because it is not static. What post–modern science needs for growing up, with an higher awareness of its limitations, but just for this with a more effective control on its ever increasing power, is a logic of "open" formal systems. From that, an epistemology of truth as unending process of self–conforming of intentional mind to an always-changing reality can suggest new more effective solutions to artificial simulations of cognitive behavior.

References

[1] Gardner, H., The new cognitive science. A History of Cognitive Revolution, Cambridge, MA, MIT Press, (1985).

[2] Putnam, H., Minds and machines, in: Dimensions of mind, S. Hook, ed., New York, Cambridge University Press, (1960).

[3] Kant, I., Logic, translated by Hartman R. S. and W. Schwarz, New York, Bobbs – Merrill, (1800, 1974).

[4] Kant, I., Kritik der reinen Vernunft, in: Kant's gesammelte Schriften, hrsg. von Deutschen Akademie der Wissenschaften, Berlin, pp. A142/B181, (1781/7, 1900).

[5] Turing, A.M., On computable numbers with an application to the Entscheidung problem, Proceedings of the London Mathematical Society 42, pp.230-265, (1937).

[6] Gödel, K., Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, Montash. für Math. u. Phys. 38, pp.173-198, (1931).

[7] Hofstadter, D.R., Gödel, Hescher, Bach. An Eternal Golden Braid, New York, Basic Books, (1979).

[8] Piaget, J., La Psychologie de l'Intelligence, Paris, Colin, (1952^5).

[9] Neisser, U., Psicologia cognitivista, tr.it., Firenze, Giunti-Martello, (1976).

[10]Fabro, C., Percezione e pensiero, Milano, Vita e Pensiero, (1941).

[11]Fodor, J.A., Fodor's guide to mental representation: the intelligent auntie's vademecum, Mind 94, pp.76-100, (1985).

[12]Turing, A.M., Computing machinery and intelligence, Mind 59, pp.433-460, (1950).

[13]Fodor, J.A., Methodological solipsism considered as a research strategy in cognitive psychology, The Behavioral and Brain Sciences 3, pp.63-109, (1980).

[14]Fodor, J.A., Psychosemantics. The problem of meaning in the philosophy of mind, Cambridge, MA, MIT Press, (1987).

[15]Tarski, A., Der Wahreitsbegriff in formalisierten Sprachen, (German translation of a book in Polish), Studia philosophica 1, pp.261-405, (1935).

[16]Tarski, A., The semantic conception of truth and the foundations of semantics, in: Readings in philosophical analysis, H. Feigl, ed., New York, Cambridge University Press, pp.52-84, (1944, 1949).

[17]Carnap, R., Studies in semantics: introduction to semantics, 3 vv., Cambridge, MA, MIT Press, (1942).

[18]Brentano, F., Psychologie vom empirischen Standpunkt, Leipzig, Hahn, (1874, 1924-28).

[19]Webb, J.C., Mechanism, mentalism and metamathematics, Dordrecht, Reidel, (1980).

[20]Kleene, S.C., Introduction to Metamathematics, Amsterdam, North-Holland, (1952).

[21]Popper, K.R., Conjectures and Refutations, London, Routledge and Keegan Paul, (1969).

[22]Searle, J.R., Mind, brains and programs. A debate on artificial intelligence, The Behavioral and Brain Science 3, pp.128-135, (1980).

[23]Searle, J.R., Intentionality. An essay in the philosophy of mind, New York, Cambridge University Press, (1983).

[24]Zalta, E., Intensional logic and the metaphysics of intentionality, Cambridge, MA, MIT Press, (1988).

[25]Quine, W.V.O., Quiddities. An intermittently philosophical dictionary, Cambridge, MA, Harvard University Press, (1987).

[26]Quine, W.V.O., Word and object, Cambridge, MA, MIT Press, (1960).

[27]Churchland, P.S., Neurophilosophy. Toward a unified science of the mind-brain, Cambridge MA, MIT Press, (1986).

[28]Husserl, E., Recherches Logiques, translated by H. Elie, Vol. I: Prolégomènes à la logique pure, Paris, P.U.F., pp.119-203, (1913, 1962).

[29]Husserl, E., Formal and Transcendental Logic, translated by D. Cairns, Le Hague, Nijhoff, (1929, 1969).

[30]Putnam, H., The mening of 'meaning'. In: Philosophical papers: mind, language and reality, New York, Cambridge University Press, pp.215-271, (1972).

[31]Putnam, H., Representations and reality, Cambridge, MA, MIT Press, (1988).

[32]Kripke, S., Naming and necessity, Cambridge MA, Harvard University Press, (1972, 1996).

[33]Hebb, D.O., Organization of behavior. A neuropsychological theory, New York, Wiley, (1949).

[34]Kohonen, T., Self-organization and associative memory. Second Edition, Berlin, Springer, (1988).

[35]McClelland, J.L., Rumelhart, D.E. and the PDP Group Parallel distributed processing, Cambridge, MA, MIT Press, (1986).

[36]Marlsburg, C. Von der, and Bienenstock, E., Statistical coding and short-term synaptic plasticity: a scheme for knowledge representation in the brain, in: Disordered systems and biological organization, NATO ASI Series F20, Berlin, Springer, pp.312-317, (1986).

[37]Gray, C.M., Koenig, P., Engel, A.K. and Singer, W. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties, Nature 338, pp.334-337, (1989).

[38]Eckhorn, R., Reitboeck, H.J., Arndt, M. and Dicke, P., Feature linking via synchronization among distributed assemblies: simulation of results from cat visual cortex, Neural Computation 2, pp.293-307, (1990).

[39]Engel, A.K., Koenig, P., Gray, C.M. and Singer, W., Synchronization of oscillatory responses: a mechanism for stimulus-dependent assembly formation in act visual cortex, in: Parallel processing in neural systems and computers, R. Eckmiller, ed., Amsterdam, Elsevier, pp.212-217, (1990).

[40]Skarda, C.A., and Freeman, W.J., How brains make chaos in order to make sense of the world, Behavioral and Brain Sciences 10, pp.161-195 (1987).

[41]Parisi, G., Asymmetric Neural Nets and the Process of Learning, Journ. of Phys. A: Math. Gen, 19, L675-680, (1986).

[42]Perrone, A.L., A formal scheme to avoid undecidabilities: an application to chaotic dynamics characterization and parallel computation in: Cognitive and dynamical systems. Lecture Notes in Computer Science, S.I. Andersson, ed., 888, pp.9-52, (1995).

[43]Perrone, A.L., Verso una teoria dinamica della computazione, in: G. Basti and A.L. Perrone, Le radici forti del pensiero debole: dalla metafisica, alla matematica, al calcolo, Padova-Roma, Il Poligrafo e Pontificia Università Lateranense, pp.255-332 (Percorsi della scienza. Storia testi e problemi, 7), (1996).

[44]Perrone, A.L., Applications of chaos theory to lossy image compression, Nuclear Instruments and Methods in Physics Research. Section A., 389, pp.221-225, (1997).

[45]Perrone, A.L., The cognitive role of chaos in neural information processing. In: Proceedings of the International School of Biocybernetics: Processes in the perception representation mechanisms, C. Taddei-Ferretti and C. Muzio eds., Singapore-London, World Scientific, (1999) In press.

[46]Posner, M.I., and Raichle, M.E., Images of mind, New York, Scientific American Library, (1994).

[47]Minsky, M. and Papert, S., Perceptrons. Second Edition, Cambridge Mass., MIT Press, (1988).

[48]Perrone, A.L., Basti, G., Messi, R., Pasqualucci E., Paoluzi L., Offline Analysis of HEP events by the 'dynamic perceptron' neural network, Nuclear Instruments and Methods in Physics Research. Section A., 389, pp.210-213, (1997).

[49]Basti, G., and Perrone, A.L., Per un ruolo costitutivo del caos deterministico nei sistemi neurali biologici ed artificiali, in: Biosistemi e complessità, E. Belardinelli and S. Ceutti, eds, Bologna, Pàtron, pp.239-289, (1993).

[50]Basti, G., Perrone, A.L. and Cocciolo, P., Using chaotic neural nets to compress, store and trasmit information, in: Applications of Artificial Neural Networks, V, SPIE–Proceedings Series, 2243, S.K. Rogers and D.W. Ruck, eds, Washington, D.C., SPIE Press, pp.540-551, (1994).

[51]Basti, G., Perrone, A.L., Pasqualucci, E., Messi, R., Picozza, P., Pecorella, W., Paoluzi, L., Principles of computational dynamics: applications to parallel and neural computations. In: Applications and Science of Artificial Neural Networks, VII, SPIE–Proceedings Series 2760, S.K. Rogers and D.W. Ruck (Eds.), Bellingham WA, SPIE Press, pp.738-752, (1996).

[52]Arecchi, F.T., Basti, G., Boccaletti, S. and Perrone, A.L., Adaptive recognition of chaos, Europhysics Letters 26, pp.327-332, (1994).

[53]Tsukada, M., Private communication for his experimental findings in late 1970s on

dynamic receptive field in cat retinal ganglion cells, (1998).

[54]Eggermont, J.J., Aertsen, A.M., Hermes, H.J. and Johannesma, P.I.M., Spectro-Temporal Characterization of Auditory Neurons: Redundant or Necessary?, Hearing Research 5, pp.109-121, (1981).

[55]Kilgard, M.P. and Merzenich, M.M., Cortical map reorganization enabled by nucleus basalis activity, Science 279, pp.1714-1718, (1998).

[56]Dinse, H., A Temporal Structure of Cortical Information Processing, Concepts in Neuroscience 1, pp.199-238, (1990).

[57]Dinse, H., A time-based approach towards cortical functions: neural mechanisms underlying dynamic aspects of information processing before and after postontogenetic plastic processes, Physica D 75, pp.129-150, (1994).

[58]Lakatos, I., Mathematics, science and epistemology, Cambridge, Cambridge University Press, (1978).

[59]Hofstadter, D.R., Fluid concepts and creative analogies, New York, Harper Collins Publ, (1995).

[60]Cellucci, C., Le ragioni della logica, Roma–Bari, Laterza, (1998).

[61]Basti, G., Per una lettura tomista dei fondamenti della logica e della matematica, in: Basti G. and Perrone A.L., Le radici forti del pensiero debole: dalla metafisica, alla matematica, al calcolo, Padova-Roma, Il Poligrafo e Pontificia Università Lateranense, pp.19-254, (1996).

[62]Basti, G., L'approccio aristotelico-tomista alle aporie dell'induzione. In: Il fare della scienza. I fondamenti e le palafitte, F. Barone, G. Basti, C. Testi eds., Padova, Il Poligrafo, pp.41-95, (1997).

[63]Feferman, S., Transfinite recursive progressions of axiomatic theories, The Journal of symbolic logic 27, pp.259-316, (1962).

[64]McCarthy, T.G., Self–reference and incompleteness in a non–monotonic setting, The Journal of philosophical logic 23, pp.423-449, (1994).

[65]Turing, A.M., Systems of logic based on ordinals, in: The Undecidable, P. Davies ed., New York, Raven Press, pp.208f., (1967).

[66]Popper, K.R., Logic of scientific discovery, London, Routledge and Keegan P., (1959).

[67]Girard, J.-Y., Le champ du signe ou la faillite du réductionisme, in: E. Nagel and J. R. Newmann, Le théorème de Gödel, Paris, Editions du Seuil, p. 161, (1989).

[68]Gentzen, G., Die Wiederspruchsfreiheit der reinen Zahlenthorie, in Mathematische Annalen, 112, pp.439-565, (1936).

[69]Minsky, M., The society of mind, New York, Simon and Schuster, (1987).

[70]Cohen, P.J., Set theory and the continuum hypothesis, New York, (1966).

[71]Platone, Tehetetus, in: Opere, G. Giannantoni, ed., Roma–Bari, Laterza, (1971ff).

[72]Platone, Parmenides, in: Opere, G. Giannantoni, ed., Roma–Bari, Laterza, (1971ff).

[73]Aquinas, T., In Aristotelis Libros Perì Hermeneias et Posteriorum Analyticorum Expositio, Spiazzi R. (Ed.). Torino, Marietti, (1964).

[74]Aquinas, T., In Librum Boetii De Hebdomadibus Expositio, M. Calcaterra ed., in: Opuscula Theologica, II v.: De Re Spirituali, R. Spiazzi ed., Torino, Marietti, (1954).

[75]Aquinas, T., In Aristotelis Libros Posteriorum Analyticorum Expositio, in: In Aristotelis Libros Perì Hermeneias et Posteriorum Analyticorum Expositio, R.M. Spiazzi, ed., Torino, Marietti, (1964).

[76]Aquinas, T., Quaestiones de Potentia, in: Quaestiones Disputatae, II v.: P. Bazzi et Al. eds., Torino, Marietti, (1965).

[77]Aquinas, T., De Veritate, in: Quaestiones Disputatae, vol. I, R.M. Spiazzi, ed., Torino, Marietti, (1953).

[78]Aquinas, T., In Quatruor Libros Sententiarum Magistri Petri Lombardi Commentarium. Liber Quartus, in: Opera Omnia, tt.VI-VII, Parma, Editio Leonina, (1865-68).

[79]Aquinas, T., Summa contra Gentiles, D.P. Marc, ed., Torino, Marietti, (1967).

[80]Ramsey, F.P., The foundations of mathematics, in: The foundations of mathematics and other logical essays, New York, (1931, 1950).

[81]Aquinas, T., In Duodecim Libros Metaphysicorum Aristotelis Expositio, R.M. Spiazzi, ed., Torino, Marietti, (1964).

[82]Aquinas, T. (S. Th.) Summa Theologiae, P. Caramello Ed., Torino, Marietti, (1952-1956).

[83]Weyl, E., Il Continuo. Indagine critiche sui fondamenti dell'analisi, A. B. Veit Riccioli ed., Napoli, Bibliopolis, (1932, 1977).

[84]De Giorgi, E., Forti, M., Lenzi, G., Verità e giudizi in una nuova prospettiva assiomatica, in: Il fare della scienza. I fondamenti e le palafitte, F. Barone, G. Basti, C. Testi eds., Padova, Il Poligrafo, pp.233-252, (1996).

[85]Forti, M., Galleni, L., An axiomatization of biological concepts within the foundational theory of Ennio De Giorgi, Biology Forum 92, pp.77-104, (1999).

[86]Forti, M., The foundational theories of Ennio De Giorgi, in: Foundations in mathematics and biology: problems, prospects, interactions, G. Basti and A.L. Perrone eds., Rome-Milan, Mursia and Pontifical Lateran University Press, (1999). In press.

[87]Aquinas, T. De Natura Generis, in: Opuscula Philosophica, R.M. Spiazzi, ed., Torino, Marietti, (1954).

[88]Aquinas, T., In librum Beati Dionysii de divinis nominibus expositio, C. Pera, ed., Torino, Marietti, (1950).

[89]Aquinas, T. (In Phys.). In Octo Libros Physicorum Aristotelis Expositio, P. Maggi-

olo Ed., Torino, Marietti, (1965).

[90]Perrone, A.L., Basti, G., Ricciardi, M., Lossy plus lossless residual encoding with dynamic pre–processing for Hubble space telescope fits images, In: Applications and Science of Computational Intelligence, II. K.L. Priddy, P.E. Keller, D.B. Fogel, J. C. Bezdek Eds., SPIE - Proceedings Series vol. 3077, SPIE- The Int. Soc. for Optical Engineer, Bellingham, WA, pp.532-541, (1999).

[91]McIntyre, R., Intending and referring, in: Husserl, intentionality and cognitive science, H.L. Dreyfus and H. Hall, eds, Cambridge MA., MIT Press, pp.219-235, (1982).

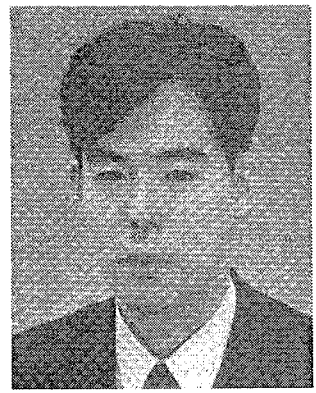
About the Authors

This page is intentionally left blank

Tetsuya Ogata

Department of Mechanical Engineering
School of Science and Engineering
Waseda University
3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan

E-mail: ogata@paradise.mech.waseda.ac.jp
Phone : +81-3-5286-3264
Fax : +81-3-5272-0948
URL : http://www.sugano.mech.waseda.ac.jp/~ogata/

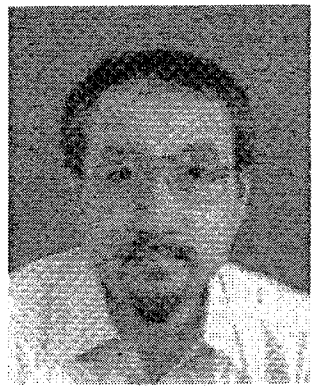


Tetsuya Ogata received the B.E. and the M.S. degree both from Waseda University, Japan in 1993 and 1995, respectively. From 1997 to 1999, he was a Research Fellow of the Japan Society for the Promotion of Science. Since 1999, he has been a Research Associate in the Department of Mechanical Engineering, School of Science and Engineering, Waseda University. His research interests include artificial minds, human-robot communication.

Tadashi Kitamura

Department of Mechanical System Engineering
Faculty of Computer Science & Systems Engineering
Kyushu Institute of Technology
680-4 Kawazu, Iizuka, Fukuoka 820-8502, Japan

E-mail: kita@mse.kyutech.ac.jp
Phone : +81-948-29-7765
Fax : +81-948-29-7751
URL : http://www.imcs.mse.kyutech.ac.jp/



Tadashi Kitamura is Professor of Mechanical System Engineering of Faculty of Computer Science and System Engineering at Kyushu Institute of Technology. He received the B.S. degree from Department of Mechanical Engineering at Waseda University, 1973, and the M.S. and Dr.Eng. degrees from Graduate School of Engineering at Kyoto University, 1975 and 1981 respectively. He was Assistant Professor of EE Department from 1984 to 87 at University of Houston, University Park, joined the faculty at Kyushu Institute of Technology in 87, and is Professor of Department of ME System Engineering there since 1988. His current research interest is design of intelligent mechatronic systems including artificial hearts, biorobots, and superconducting actuators. He is the Editor-in-Chief of Journal of Biomedical Fuzzy Systems Association since 1997. He published over 100 academic papers, edited three books, and co-authored nine books. He is a member of IEEE, ISAO, JSAO, BMFSA, JSME and JSICE.

Akifumi Tokosumi

Department of Value and Decision Science
Graduate School of Decision Science and Technology
Tokyo Institute of Technology
2-12-1 Ookayama, Meguro-ku, Tokyo 152-8552, Japan

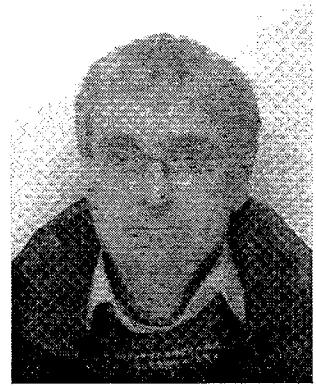
E-mail: akt@valdes.titech.ac.jp
Phone : +81-03-5734-2680
Fax : +81-03-5734-3618
URL : http://www.valdes.titech.ac.jp/~akt/

Akifumi Tokosumi was born in Japan and received M.A. in experimental psychology from Hokkaido University, in 1977. His previous affiliations include: Instructor, Hokkaido University, 1979-1986; Visiting Worker, University of Edinburgh, 1982-1983; Associate Professor, University of the Sacred Heart in Tokyo, 1986-1992. He is presently Associate Professor of Cognitive Science and Psychology at the Department of Value and Decision Science, Tokyo Institute of Technology. His research interests are in the areas of psycholinguistics, emotion, and computational psychology. He is a supporting member of the International Workshop on Literature in Cognition and Computer (iwLCC), whose web site is on "http://www.valdes.titech.ac.jp/~iwlcc/".

Michel Dufossé

Laboratoire Creare
Inserm U-483
University Pierre and Marie Curie
Paris 5, France

E-mail: michel.dufosse@pacwan.fr
Phone : +33-144-27-2624
Fax : +33-144-27-3438

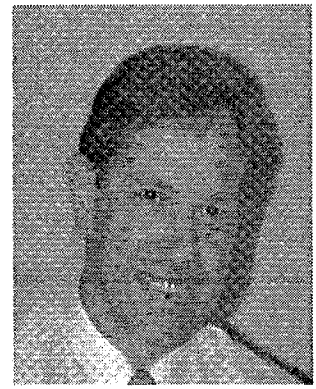


Starting from a mathematical background, Dr. Dufossé started experimental research on brain motor control. During the 1970's decade, he was involved in the hypothesis of cerebellar study. During the 1980's decade, the neural network expansion supported my goal of functional brain modelling (the Marr-Albus-Ito perceptron theory, the Basal Ganglia hypothesis). Presently, Dr. Dufossé is involved in modelling the interaction between brain structures: cerebrum, cerebellum, basal ganglia, spinal "equilibrium theory".

Nikola Kasabov

Department of Information Science
University of Otago
P.O.Box 56, Dunedin, New Zealand

E-mail: nkasabov@otago.ac.nz
Phone : +64-3-479-8319
Fax : +64-3-479-8311
URL : http://divcom.otago.ac.nz/infosci/Staff/NikolaK.htm

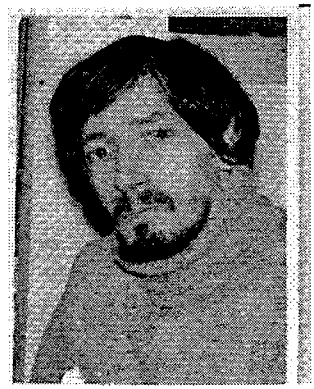


Nikola K. Kasabov is Professor of Information Science in the Department of Information Science, University of Otago, Dunedin, New Zealand. He received his MSc degree in Computer Science from the Technical University in Sofia, Bulgaria, in 1971. He obtained his PhD degree in Mathematical Sciences in 1975 from the same university. Kasabov has published over 200 works, among them over 50 journal papers, 90 conference papers, 15 book chapters, 5 text books, 2 edited research books, 3 edited conference proceedings, 19 patents and authorship certificates in the area of intelligent systems, connectionist and hybrid connectionist systems, fuzzy systems, expert systems, speech recognition, and data analysis. He is Director of the research laboratory for Knowledge Engineering and Computational Intelligence in the Department of Information Science, University of Otago. Kasabov is the immediate past President of APNNA - Asia Pacific Neural Network Assembly. He is member of the TC12 group on Artificial Intelligence of IFIP and also member of the IEEE, INNS, NZCS, NZRS, ENNS, IEEE Computer Society. He was the general chairman of the First, the Second and the Third New Zealand International Conferences on Artificial Neural Networks and Expert Systems - ANNES'93, ANNES'95 and ANNES'97 (the latter jointly held with ICONIP'97 and ANZIIS'97).

Horia-Nicolai L. Teodorescu

Department of Computer Science
Faculty of/ College of Engineering
University of South Florida
4202 E. Fowler Ave., Tampa, Fl 33620 5399, USA

E-mail: teodores@csee.usf.edu
Phone : +1-813-974-9036
Fax : +1-813-974-5456
URL : http://www.csee.usf.edu/faculty_pages/teodorescu.html

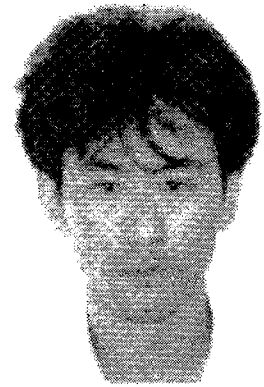


Dr. Horia-Nicolai L. Teodorescu has written about 250 papers, authored, co-authored, edited or co-edited more than 20 volumes, and holds 21 patents. He won several medals for his inventions. He is a Senior Member IEEE, and holds several honorific titles, including "Eminent scientist" of FLSI. He is a *correspondent member* of the Romanian Academy. He is a Chief Editor of *Fuzzy Systems & A.I.– Reports and Letters, International Journal for Chaos Theory and Applications* and of two other journals. He is a member of the editorial boards of *Fuzzy Sets and Systems* and of five other journals.

Yukio-Pegio Gunji

Department of Earth & Planetary Sciences
Faculty of Science
Kobe University
1-1 Rokkoudai, Nada-ku, Kobe, 657-8501, Japan

E-mail: yg@scipx.planet.sci.kobe-u.ac.jp
Phone : +81-78-803-5745
Fax : +81-78-803-5757
URL : http://www.planet.sci.kobe-u.ac.jp/

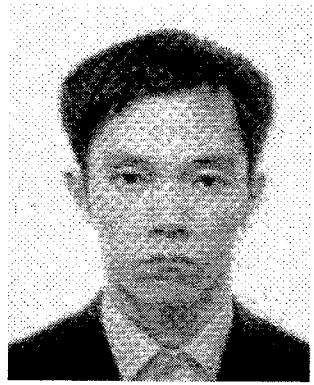


Yukio-Pegio Gunji (Professor). Born in 1959. Graduated from Tohoku University (Faculty of Science) in 1981. Doctor of Science from Tohoku University in 1986. Assistant Professor of Kobe University (1986-1993). Associated Professor of Kobe Univesity (1993-1999). Professor (1999-). Mainly interested in theory of self-organization, development, ecolution and emergence. Also in experimental studies refgarding ethology and cognitive science.

Nobuhide Kitabayashi

Department of Earth & Planetary Sciences
Faculty of Science
Kobe University
1-1 Rokkoudai, Nada-ku, Kobe, 657-8501, Japan

E-mail: wassi@shida.planet.kobe-u.ac.jp
Phone : +81-78-803-5759
Fax : +81-78-803-5757
URL : http://www.planet.sci.kobe-u.ac.jp/

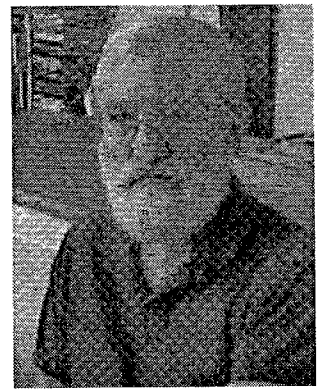


Nobuhide Kitabayashi. 1998 - Doctorate in Science in Graduate School of Science and Technology at Kobe University, Kobe, Japan. Present position: postdoctoral fellow in Graduate School of Science and Technology at Kobe University.

Walter J Freeman

Division of Neurobiology
Department of Molecular & Cell Biology, LSA-129
University of California
Berkeley California 94720-3200, USA

E-mail: wfreeman@garnet.berkeley.edu
Phone : +1-510-642-4220
Fax : +1-510-643-6791
URL : http://sulcus.berkeley.edu/

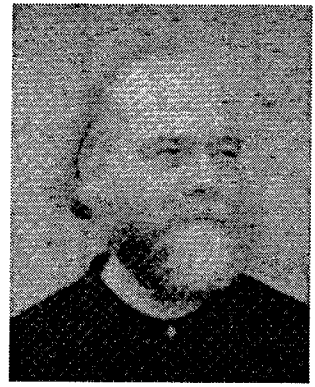


Walter J Freeman studied physics and mathematics at M.I.T., and medicine at Yale University (M.D. *cum laude* 1954). He studied medicine at Johns Hopkins and neurophysiology at UCLA. He has taught brain science in the University of California at Berkeley since 1959, where he is Professor of the Graduate School. He received the Pioneer Award from the Neural Networks Council of the IEEE, and was President of the International Neural Network Society in 1994. He is the author of >350 articles and two books: "Mass Action in the Nervous System" (1975) and "Societies of Brains" (1995).

Mark H. Bickhard

Department of Cognitive Science
Lehigh University
Bethlehem, PA 18015, USA

E-mail: mark.bickhard@lehigh.edu
Phone : +1-610-758-3633
Fax : +1-610-758-6277
URL : http://www.lehigh.edu/~mhb0/mhb0.html

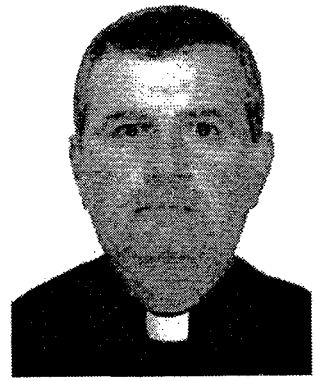


Mark Bickhard's work focuses on models of the mind and of persons more broadly. A central axis of this work is a model of the nature of representation — a pragmatic, or interactivist, model. The implications of the interactive model of representation ramify throughout all domains of the study of the mind — the classical cognitive backbone of perception, memory, cognition, reasoning, and language, but also, for example, emotions, consciousness, motivation, and so on. A graduate of the University of Chicago, 1973, he was at the University of Texas at Austin until 1990, when he moved to Lehigh University to accept the Henry R. Luce Professorship of Cognitive Robotics and the Philosophy of Knowledge.

Gianfranco Basti

Faculty of Philosophy
Pontifical Lateran University
Piazza S. Giovanni Laterano, 4 I-00184 Rome, Italy

E-mail: basti@pul.it
Phone : +39-06-4742529
Fax : +39-06-4742529
URL : http://www2.chiesacattolica.it/pul/filosofi.htm



Gianfranco Basti received the M.D. Theol. in 1980 from the Theological Faculty of the Pontifical Gregorian University in Rome, Ph. D. Philos. in 1984 from the Philosophy Dept. of the State University of Rome "La Sapienza", M.D. Philosophy in 1988 from the Philosophy Faculty of the Pontifical Gregorian University.

Since 1987 he is Professor in Charge by the Philosophy Faculty and the Institute of Religious Sciences by the Pontifical Gregorian University. From 1992 he joined the Faculty of Philosophy of the Pontifical Lateran University in Rome, where from 1992 to 1998 he was Professor in Charge of Philosophical Anthropology and received a full professorship in Philosophy of Nature and of Science in October 1996.

His main research interests lie, from the scientific standpoint, in neural networks and chaotic systems; from the philosophical standpoint, in mind–body problem and philosophy of logic. He was researcher in neural network field by the Italian Institute of Electronic Circuits (ICE) of the National Research Council (CNR) in Genoa and by the National Institute for Nuclear Physics (INFN), Section of "Rome 2, Tor Vergata". In 1995 he received by INNS a "Neural Network Leadership Award".

In 1997 he founded, by the Pontifical Lateran University, with Prof. Edward Nelson from the Mathematics Dept. of Princeton University (USA), the International Research Area on Foundations of the Sciences (IRAFS), devoted to promote the study of logic and mathematics foundations in scientific disciplines. He is member of INNS-International Neural Network Society; IEEE-Institute for Electric and Electronic Engineering Computer Society and Neural Network Council; SPIE-The International Society for Optical Engineering.

This page is intentionally left blank

Keyword Index

A

absorption 30,31,32
abstraction 78,101,245,270
active learning 85,98
adaptability 14,122,123,221
adaptive intelligent system 79
affective computing 44,45
affordance 3
ambush 25,28,32,38,39,41
analytic method 249,263,264,265, 266,267,268,269,270,271
annex language 116
anthropomorphism 184,185,187, 203
ants 184,185,188,189,190,191,192, 193,197,199,200,201,202,203
approach 25,28,32,34,39,41
artificial intelligence (AI) 2,23,24, 79,82,107,116,117,118,119,120, 142,143,208,240,241,255,257
- paradigm 107
artificial sensitivity 117
asymmetric dependence 226
autonomous robot 2,5,6,20
avalanche size 158,159,160,164
avoid 27,28,
axiomatic method 249,263,264, 265,266,268,269,270,271,276, 282

B

Bak-Sneppen model 159
basal ganglia 53,54,59,67,68,70,71
behavior 1,2,3,5,11,12,13,14,16,23, 24,25,26,27,28,29,30,31,32,33, 34,35,36,37,38,39,40,41,59,117, 120,122,123,129,132,139,144, 145,146,164,183,184,185,188, 189,190,193,194,195,197,200, 201,202,203,209,211,212,222, 227,242,244,245,258,259,260,282
- selection 25,27,28,29,33,34,35, 36,37,38,39,40
- selection criteria 25,33,34,35,36, 37,38,40
- based architecture 24
being 240,270,271,272,273,274, 276,277,279,281
bias/variance dilemma 81
biological plausibility 107
brain dynamics 208,254

C

capture 25,28,34,41
cascade-eco training 86
catastrophic forgetting 107
category theory 168,170
cell's preferred direction 57,58

central nervous system 217,218, 224,228,229,230,231,232
cerebellar cortex 54,60,61,66,72
cerebellum 53,54,56,60,61,62,63, 65,66,67,68
cerebral cortex 54,55,56,61,62,63, 64,65,66,68,71,72
chaotic dynamics 208,210,258,259, 261
coding 60,150,151,178,247,251, 253,254
cognitive model 44
cognitive science 43,44,151,137, 138,224,241,243,244,245,252, 254,263,269,281,282
co-limit 171,173,174,177,178,179
communication 1,2,3,4,5,11,14, 16,20,116,118,123,125,126,128, 129,130,131,132,133,134,135, 143,144,154,203,207,208,260
- of meaning 208
communicative computer 128
composition 165,166,167,168,170, 171,174
computer semiotic 135
connectionism 16,240
connotation 140
consciousness 24,25,26,27,28,29, 30,31,32,34,35,37,39,40,41,117, 128,131,149,150,152,217,231, 232
- based architecture (CBA) 25,26, 27,33,37,38,39,40,41
- field 26,27,28,31,32, 39
constructivism 81

D

data compression 279
desire 5,17,18,19,20,32,34,50
detour 25,28,32,36,37,38,39,41
diagonalization 247,253,279
displacement 64
dynamics 59,60,61,62,63,65,79, 103,105,106,130,158,193,208, 210,212,240,254,257,258,259, 260,261,265

E

EFuNN learning algorithm 97,103, 106
embodiment 2,24,217,232
emergent property 204
emotion 2,5,14,20,26,28,30,32,36, 40,43,44,47,48,53,120,184,232
environment 3,11,14,20,25,39,40, 54,59,78,79,80,81,83,120,126, 127,128,212,214,218,220,221, 223,224,227
epistemology 227,244,249,273,282
escape 28,32,36
essence 240,264,265,268,270,271, 272,273,274,275,276,277,281
evolution 25,39,40,62,78,82,122, 128,130,143,145,150,157,160, 210,222,230,232,282
evolving brains 78
evolving connectionist systems 79, 82,83,106,107

evolving fuzzy neural networks 86, 106
existence 47,81,144,145,150,156, 158,167,168,171,186,187,219, 240,253,265,268,270,271,272, 273,274,275,276,277,278

F

Fodor 226,227
food-retrieving behavior 184,185, 190
formal arithmetic 246
formal semantics 245,246,249,250
foundations of logic 244,282
frame of relation 130
free category 152,157,165,166,167, 168,170,172,174,177,179
functionalism 250,253
future 151,152,165,173,174,178, 179
fuzzy 3,54,82,86,87,88,89,90,92, 93,94,95,96,97,98,99,102,103, 106,107,141,142
- input 86,87,89,90,93,94,99, 106
- output 87,89,92,93,94,97,99, 106

G

generalisation accuracy 106
genetic algorithms 82,103,107
gesticulation 116,120,131
global generalisation 104,105,106
goal directedness 222
graph morphism 166,167,170
group relationship 122,128,144

H

Hebbian rule 256,258
hermit crabs 183,184,188,189,202
hesitation 32
human semiotic 135
hunger 30,31,32,35,36
Husserlian 25,40

I

identity 165,166,167,168,169,170, 171,172,175,179,243,270,271, 275,276
incompleteness 240,243,246,248, 252,265,266,267,278
incremental learning 80,81
information semantics 224
inhibited behavior 26,39
integrating level 137
intelligence 2,3,23,24,39,43,44,50, 78,107,115,116,117,118,120,129, 138,143,153,156,162,184,214, 240,243
intelligent system 23,79,144
intensional logic 251,252,253,262
intentional action 211
intentionality 230,239,240,245, 249,250,251,252,253,254,260, 263,269,280,281,282
interactivity 119

internal measurement 151,174,178, 179
internal secretion system 2,5,8,20
invalidation 162,168,169,170,171, 172,178,179

J

judgment 17,18,20

K

kinematics 56,59,64
knowledge 14,24,43,45,46,48,50, 53,70,77,79,80,82,84,86,107,119, 120,122,126,129,130,131,134, 136,137,138,141,142,143,240, 244,248,264,268,270,271
- based neural networks 82
Kripke 150,154,155,184,253

L

learning phase 62,66,85,257,258
Left-Kan extension 177,178
lifelong learning 80,81,103
limit 116,119,121,139,142,145, 146,171,173,174,177,179,253
literary computing 44,45,50
local generalisation 105,106
long-term memory 33,78,79,80,85

M

mathematical intuition 266,269
meaning 17,24,25,28,31,37,38,39, 40,117,122,123,131,132,135,137, 139,143,145,146,154,155,156, 178,186,187,208,209,210,212, 213,214,245,251,253,280
membership function 3,87,94
memory-based learning 78
mental experience 217,218
metalogic 240
methodological solipsism 249,250
Millikan 219
morality 122,144,146

N

natural language 20,117,120,134, 141,145,253
neural network 15,54,59,60,68,78, 80,82,83,86,103,106,141,254
node aggregation 100
node creation 101
normative function 218

O

off-line learning 28,104,105
1/f noise 184,203
one-pass learning 91
on-line learning 80,81,104,105, 106,107

on-line parameter optimisation 103
ontogeny 26
open problem space 105,106
open structure 80
orientation 57,260
origin of error signals 67

P

parallel distributed processing 257
parallel language 117,134,143,144
parallelism 240
partial recursive function 248,279
passive learning 85,98
past 151,152,165,173,174,178,179
perception 25,27,28,30,31,35,41, 128,210,218,232,241,243,244, 250,263,281
perceptive computer 128
personality 50,117,118,122,133, 138,144,145,218
phenomenological 25,40,239,241, 245
phenomenology 224,233,245
phylogeny 26
Piaget 118,224,243,244,250
positive capabilities 145
power-law 158,159,160,164,179
pruning rule 95,107
pseudo-cycle 258,259,261,279
psychology 4,20,79,117,118,119, 130,131,138,139,239,241,243, 244,245,249,252,281
punctuated equilibrium 152,158, 163,164,179
pursuit 32,36

R

recognition 3,14,17,18,20,36,79, 82,122,124,125,137,144,209,255, 257,258,261,279
recurrent temporal connections 88
recursive clock 152,157,162,164, 165,170
recursive function 248,279
redundancy 78
reference 245,246,249,251,252, 253,254,257,262,263,267,273, 280,281
reflexiveness 122
relation 122,126,128,130,132,134, 135,139,141,143,146,150,151, 240,242,246,247,251,253,260, 264,266,271,275,276,277,278, 279,280,281
relationship 122,124,125,126,128, 129,130,131,132,135,136,137, 140,144
- characteristics 136
representation 24,28,39,43,45,63, 78,79,83,85,87,116,130,132,135, 136,137,138,140,141,143,149, 150,151,153,173,178,179,208, 209,210,211,213,218,221,222, 224,225,226,227,228,231,261
- basis 141
- systems 128,129
representational content 151,221, 224,225,226,227

rule extraction 82,83,86,98,99
rule insertion 82,98

S

search 28,32,36
Searle 152,153,154,250,251,252, 253
selectivism 81,82
self-analysis 83,86
self-organisation 84,107
self-organizing 14,15,16,158
self-preservation 3,4,9,14,18,19,20
self-reporting 129,130,131
self-representation 133,140
self-similar return map 161
self-similarity 161,162,187
semantic 39,40,123,135,141,142, 208,240,245,246,248,249,252, 253,254,267,268,273,275,279
semiotic 116,135
sensation basis 141
sensitive machine 123
sensitivity 44,60,116,117,123
- threshold 89,92,94,102, 107
short-term memory 79,86,95,97
side language 116,133
silent neurons 230
sleep 25,36
socialized machine 123
soft AI 23,24
speech 2,5,17,20,77,78,79,92,93, 97,98,123,124,134,144
strategies for allocating rule nodes 96
subliminal language 116,128
subsumption architecture (SSA) 2, 24,39
supervised learning 87
symbol 2,3,15,16,17,28,31,125, 150,153,154,155,156,157,159, 179,186,208,209,229,241,250, 253,254,257,258,259,263,275, 280,281,282
syntax 139,153,154,156,164

T

text understanding 47
timing 228,229,230
tool 2,4,5,14,27,28,31,56,79,116, 134,136,137,140,141,143,164, 184,185,187,188,189,190,193, 195,196,197,198,199,200,201, 202,203,204,230,232
truth 220,221,222,240,245,246, 247,248,249,252,253,257,264, 265,266,267,270,271,273,278, 281,282
Turing machine 228,229,242,279
Turing test 152,153,244,251
two-passes learning 91

U

undecidability 266,268
universal learning machine 104
universality 104,158,168,175,242, 270,272,273

usage of tool 188

V

volume transmitters 230

W

wamoeba 2,5,6,7,8,9,10,11,12,13,
14,15,16,17,18,19,20
wholarchic groups 144
wish 48,49,50
Wittgenstein 150,153

Z

Zipf's law 184,185,187,188,189,
192,193,197,199,203,204

FLSI Soft Computing Series — Volume 3

What Should be Computed to Understand and Model Brain Function?

From Robotics, Soft Computing, Biology and Neuroscience to Cognitive Philosophy

Editor: **Tadashi Kitamura** *(Kyushu Institute of Technology, Japan)*

This volume is a guide to two types of transcendence of academic borders which seem necessary for understanding and modelling brain function. The first type is technical transcendence needed to make intelligent machines such as a humanoid robot, an animal-like behavior architecture, an interpreter of fiction, and an evolving learning machine. This technical erosion is conducted into areas such as biology, ethology, neuroscience and psychology, as well as robotics and soft computing. The second type of transcendence of cross-disciplinary boundaries cuts across scientific areas such as biology and cognitive science/philosophy, into comprehensive, less technical and more abstract aspects of brain function. These aspects enable us to know in what direction and how far an intelligent machine will go.

ISBN 981-02-4518-1

9 789810 245184

www.worldscientific.com

4607 hc