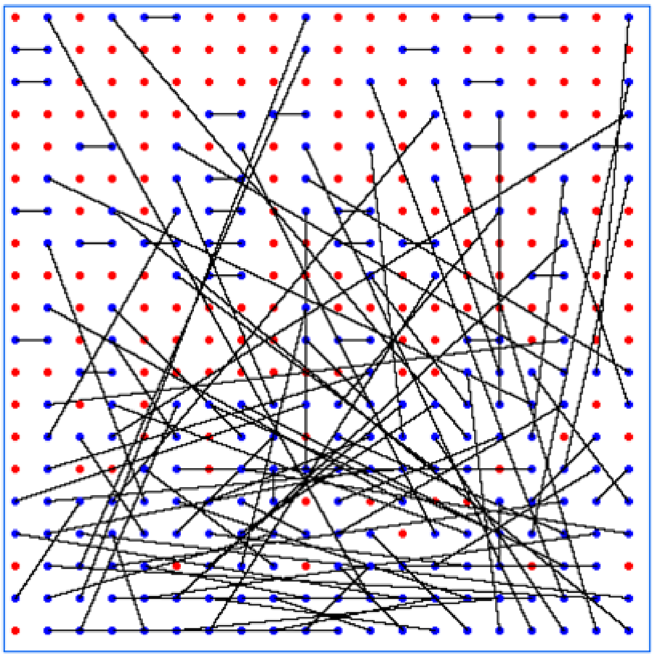
# Natural Intelligence

## The INNS Magazine

ISSN 2164-8522

Volume 2, Issue 1, July 2013



**Machine Consciousness: A Modern Approach**

**Thermodynamic Learning Rule**

**Probabilistic Adaptive Learning Mapper (PALM)**

INTERNATIONAL NEURAL NETWORK SOCIETY

Retail: US$10.00

# Natural Intelligence: the INNS Magazine
# Editorial Board



The INNS welcomes YOU who as many our members, by studying neural networks, strive to bridge the gap between natural and artificial intelligence… I welcome you to join the INNS to help us improve our collective understanding of the human brain/mind and create more powerful intelligent machines for addressing complex problems of the world.

**Danil Prokhorov, President of the International Neural Networks Society**

The International Neural Networks Society (INNS) is embarking on a new journey. Not satisfied with its own past successes, INNS is constantly looking for new ways to better itself. The goal is for INNS to be the most prestigious professional organization in fields around neural networks and natural intelligence (broadly defined), as it has been for years. To keep up with the fast changing world of relevant science and technology, a new magazine that is designed to appeal to a broader readership ---the new INNS magazine entitled "Natural Intelligence"---thus is born.

**Ron Sun, Former President of the International Neural Networks Society**

The new INNS magazine aims at bridging different communities, spreading from neuroscientists to information engineers, and also from university students to world leading researchers. We define "Natural Intelligence" to include both "intelligence existing in nature" and "intelligence based on the state of things in nature". Therefore, the new INNS magazine "Natural Intelligence" plans to cover (a) experiments, (b) computational models, and (c) applications of the intelligent functions in our brains. Also, there is an important need for well-written introductory papers targeting both young and established researchers from other academic backgrounds. The interdisciplinary nature of the many new emerging topics makes these introductory papers essential for research on Natural Intelligence. Therefore, the new INNS magazine will mainly publish (a) review papers, (b) white papers, and (c) tutorials. In addition, columns, news, and reports on the communities will also be included.

**Soo-Young Lee, Co-Editor, Natural Intelligence: the IN INNS Magazine**

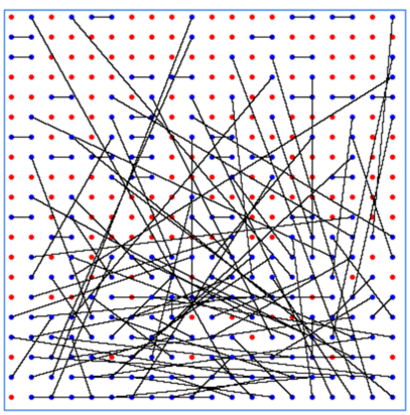
In 2013 Drs. Robert Kozma, Francesco Carlo Morabito, and Harold Szu had joined to the Co-Editors of Natural Intelligence: the INNS Magazine. The first issue of the second volume is finally online on July 31st, 2013. We would like to have the second issue of the second volume in the late 2013.

ISSN 2164-8522

# Natural Intelligence

## The INNS Magazine

Volume 2, Issue 1, July 2013



### Tutorials



### President's Message & Editor's Columns



### Meeting Report



### Call for Papers



### Conference Announcement

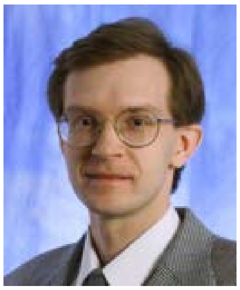


**Natural Intelligence: the INNS Magazine** is published quarterly by the International Neural Network Society (INNS) at *www.inns.org* and *www.ni-inns.info*. **Headquarters**: 2424 American Lane, Madison, WI 53704, U.S.A. Telephone: +1-608-443-2461. Fax: +1-608-443-2474 or 1-608-443-2478. E-mail addresses: inns@reesgroupinc.com.
All submission should be made to *inns.ni@gmail.com* or *ni@neuron.kaist.ac.kr*.

Message from the INNS President

# Bridge between Natural and Artificial Intelligence

**Danil Prokhorov**
The INNS President



The INNS welcomes YOU who as many our members, by studying neural networks, strive to bridge the gap between natural and artificial intelligence. Among many INNS research areas are neuroscience, connectionism, cognitive science, brain-inspired computing, neuroinformatics, brain informatics, and all kinds of neural network applications. The INNS also welcomes those who study emerging research topics, for example, bio-inspired robotics, autonomous learning and mental development, etc.

The diversity and inclusive spirit of the INNS are reflected in our resolve to develop and support special interest groups (SIG) and regional chapters, especially in developing regions of the world, and to foster collaborations with other professional societies.

The INNS actively collaborates with several professional organizations, such as the European Neural Network Society (ENNS), the Japanese Neural Network Society (JNNS), the IEEE Computational Intelligence Society (CIS), and the Asia-Pacific Neural Network Assembly (APNNA), to name a few.

With IEEE-CIS, we take turns in leading the organization process of the International Joint Conference on Neural Networks (IJCNN), which is the premier venue for our members.

With the ENNS and JNNS, we share our flagship journal, Neural Networks, which publishes top-quality results of studies in various areas of neural networks and natural intelligence. We also publish this magazine, Natural Intelligence, under a skillful leadership of Soo-Young Lee and his team.

Let me also use this opportunity to welcome new members of the INNS Executive Commitee, from this year, Kathy Kuehn – the INNS Managing Director, Ali Minai – VP for Conferences, Have Siegelmann – Secretary, as well as continuing members Irwin King – VP for Membership and Dave Casasent – Treasurer.

I would like to welcome three new members of our Board of Governors, Prof. Juergen Schmidhuber, Prof. De-Shuang Huang and Prof. Danilo Mandic, as well as Ron Sun, Ali Minai and Irwin King who were elected too to continue their important service to the INNS.

On behalf of the Board, I would also like to express deep appreciation to Prof. Ron Sun for his two years of excellent service as the INNS President. I would also like to thank Bruce Wheeler, our past managing director, for his many years of dedication to our society!

Last but not least, let me also thank the outgoing members of the Board, Profs. Stefan Schaal, Robert Kozma, Derong Liu, Carlo Morabito and Klaus Obermayer for their valuable service to the society!

The INNS presents its highest honors to outstanding researchers who made significant, ground-breaking contributions to the science and technology in our fields, including the Hebb Award, the Helmholtz Award, and the Gabor Award. For this year the awards will be presented at the IJCNN 2013 in Dallas, to these outstanding scientists: Robert Hecht-Nielsen for the Gabor Award, David Willshaw for the Hebb Award, and Juergen Schmidhuber for the Helmholtz Award. And one more award called INNS Young Investigator Award is to be presented to Jan Peters. Congratulations to all of INNS Awardees!

I welcome you to join the INNS to help us improve our collective understanding of the human brain/mind and create more powerful intelligent machines for addressing complex problems of the world. ■

# An International Doctorate in Natural Intelligence: a Proposal for Trisociety Collaboration

**Francesco Carlo Morabito**
Co-Editor, Natural Intelligence: the INNS Magazine

The doctorate degree was historically introduced by von Humboldt at Berlin University in 1810. In the USA, the first student that received the Ph.D. was at Yale University in 1862.

The university students that conclude the study program (typically in three years) are awarded the Doctor of Philosophy (Ph. D) degree. The doctorate has the main objective of forming researchers through training in relevant centres and under the supervision of an experienced, somehow recognized, supervisor. Recently, the possibility of having a pool of supervisors has emerged somewhere, in particular, in the USA. This is the purpose of my editorial writing.

BACKGROUND: Recently, the European Universities Association (EUA) fixed the core competencies to be acquired at the 3rd cycle level through the Salzburg recommendations, in order to avoid reducing the standard of the Ph.D. and to refocus on the importance of research. At an European level, there was an impetus for increasing the Ph.D. education till thinking at a model of doctoral training incorporated as a third level course within the programs of many Universities.

Among the relevant points to be achieved, there was the indication of some principles of innovative doctoral training. A significant emphasis was given to interdisciplinary for reaching research excellence. The Ph.D. students should follow an articulate path that includes opportunities to work in industry during the training as well as to develop generic, transferable skills. Ultimately, the training of doctorates should be finalized to improve the ability to cope with emerging economy's requirements. To be competitive there is a need for international networking also with the aim of recognizing modern and varied perspectives of different countries. However, the network of excellence is mainly intended at a "regional" level (say, European, or US-based Universities).

OPPORTUNITY: As more and more scientific, medical and engineering applications and innovative researches involve natural intelligence, and since ***our Society is "naturally" networked in the world (through INNS, ENNS, and JNNS)***, I think it is now the right time to think and, possibly, design an ***International Doctorate in Natural Intelligence***. Of course, the programme should lead to transform excellent qualified students from all over the world in successful researchers. Thus, they should spend part of their training time involved in researches which are carried out in different laboratories in the world, with supervisors proposed and selected by, for example, our societies' governors.

MOU: An international Memorandum of Understanding (MOU) agreement should be signed among different involved parties in order to define a commonly accepted plan of lectures and research topics. The students should be involved in preparing and being first authors of internationally peer-reviewed papers within the Ph.D. course, under the directions of a pool of supervisors.

SCOPE: Once assessed the involvement of different international research centres and guaranteed the mobility through the centres with mutual recognition of the training activities, with regards to the content of the course, a detailed study should define the core competencies to be acquired within the wide range of **Natural Intelligence** coverage. They should certainly include basic symbolic and computational intelligence (natural language processing, search, agents, knowledge based systems, social agents and signal processing); machine learning (neural networks, evolutionary computation, swarm intelligence, kernel machines, hybrid signal processing); software and hardware implementation of bio-(brain-)inspired circuits from design to realization (e.g., olfactory, vision, tactile-recognition systems); automation and robotics; neuromorphic engineering; but also some high-level topics of modern mathematics, statistics, physics, management/economic/financial matters, and perhaps psychology and certainly Brain-like computational tools.

GOVERNESS: The applications and the research topics to be developed as individual assignments/projects should be proposed by the Steering Committee of the Ph.D. Course that would include outstanding scientists from our societies covering all the world by focusing on different aspects of the various economies/cultural models involved. The students will be encouraged to meet at the IJCNN conferences, each year, to discuss the advancement of their activities, to present written reports and to collectively analyse their achievements. INNS/ENNS/JNNS could grant a special awards to the best reports. This could help in growing a group of future leading researchers within our Societies that finally will guarantee the survival of our associations through the years to come.

FINANCE: It would not be difficult to find the grants coverage for a limited number of excellent Ph.D. students within the research projects we individually carry out at our Universities. The clear advantage for our own organizations will be multiple recognitions at world level. Our NI Magazine could be a privileged publication site for the advancements of doctorates' activity.

ACTION: We shall continue the discussion to sketch a DALLAS MOU while enjoying a great IJCNN 2013, Dallas, TX, USA! ■

# Machine Consciousness: A Modern Approach

**Riccardo Manzotti**
IULM University, Italy
*corresponding author*: Riccardo.manzotti@iulm.it

## Abstract

The purpose of the tutorial is to offer an overview of the theoretical and empirical issues in artificial consciousness. Is it possible to devise, project and build a conscious machine? What are the theoretical challenges? What are the technical difficulties? What is the relation between cognition and consciousness? The most promising models will be sketched and contrasted: Global Work Space, Tononi's information integration, Embodied Cognition, Externalist approaches, bio-inspired cognitive architectures. Questions to be discussed include: What advantage does consciousness provide? Is conscious experience the hallmark of a special style of information processing? Would conscious machines be able to outperform intelligent machines?

## 1. Is consciousness Relevant for AI?

Since 1949 – when Shannon and Weaver cast the foundation for the forthcoming information age (Shannon and Weaver 1949) – computer science, cognitive science, AI and engineering have aimed to replicate the cognitive and mental capabilities of biological beings. To this purpose, various strategies have been envisaged. By and large, we may distinguish various approaches: the symbolic and logical approach of classic AI (Haugeland 1985a; Russell and Norvig 2003), the sensori-motor approach (Pfeifer 1999), neural-network oriented design (Sporns 2011), the bioinspired strategy (Pfeifer, Lungarella et al. 2007b), and the classic AI approach (Russell and Norvig 2003). All these approaches share something – they focus mostly on the intelligent behavior showed by agents. They try to replicate the capability to react to the environment stimuli and to choose the appropriate course of actions. However, something may be missing. According to Russel & Norvig (2003) one of the main goal of AI has been that of designing system that think … "*machine with minds* in the full and literal sense" (Haugeland 1985b). A full-fledged mind inevitably raises the issue of consciousness.

If we take the human being as the target of our efforts, we are immediately struck by something that AI so far has not addressed properly, namely consciousness.

Human beings not only act and behave. They are conscious of what they do and perceive. Somehow, human beings *feel* what happens to them, a condition usually defined as *being conscious* or as having *consciousness*. There is something that it like to be a certain human being (Nagel 1974). Furthermore, there seems to be some strong dependence between autonomy and consciousness.

The problem of consciousness appears so difficult that it has been dubbed the *hard* problem (Chalmers 1996), to the extent that some scientists and philosophers have even argued that it may lie beyond our cognitive grasp (McGinn 1989; Harnad 2003).

For one, there is a crucial question of paramount importance in neuroscience and AI: does consciousness provide a better way to cope with the environment? Or, to put it differently, has consciousness any selective advantage?

At this point and very broadly, there are two conflicting positions. On the one hand, there are authors that set aside consciousness as a philosophical issue of no concern for AI, Cognitive Science and Neuroscience. As Ronald Arkin put it, "Most roboticists are more than happy to leave these debates on consciousness to those with more philosophical leanings" (Arkin 1998). Either because consciousness has no practical consequences or because it is a false problem, these group of authors prefer to focus on more defined issues (vision, problem solving, knowledge representation, planning, learning, language processing). For them, either consciousness is a free bonus at the end of the AI lunch, or is nothing but a by-product of biological/computational processes.

On the other hand, an increasing number of scientists are taking seriously into consideration the possibility that human beings' consciousness is more than an epiphenomenal by-product. Consciousness may be the expression of some fundamental architectural principle exploited by our brain. If this insight were true, it would mean that, in order to replicate human level of intelligence, we ought to tackle with consciousness too.

In support of pro-consciousness group, there is the fact that we have a first-person experience of being conscious, which is not deniable by any amount of theoretical reasoning. In other words, when I feel a pain in my arm, there is something more than the triggering of some appropriate behavioral response. If this feeling had no practical consequences, it would follow that consciousness is epiphenomenal – namely that it has no practical consequences whatsoever. More bluntly, it would follow that consciousness is a useless phenomenon. Such a conclusion would contradict the principle of natural selection – it does not seem likely. Furthermore, in the animal kingdom, there seems to be a correlation between highly adaptable cognitive systems (such as human beings, primates, and mammals) and consciousness. Insects, worms, arthropods, and the like that are usually considered devoid of consciousness are much less adaptable (they are adaptable as a species but not very much as individuals).

As a result, many scientists are now looking for something that explicitly addresses the issue of machine consciousness (Buttazzo 2001; Holland 2003; Holland 2004; Adami 2006; Chella and Manzotti 2007; Aleksander 2008; Aleksander, Awret et al. 2008a; Buttazzo 2008; Chrisley 2008; Manzotti and Tagliasco 2008). So far, there is no accepted consensus as to what consciousness may be. There are several and often conflicting hypotheses. According to some authors, consciousness is the result of a special kind of information process related with information integration (Tononi 2004b; Tononi 2008). According to another group depend on goal generation and development (Manzotti and Tagliasco 2005b), or embodiment (Holland 2004), or a certain kind of information processing akin to the global workspace (Shanahan 2005a; Shanahan 2010), or the replication of imagination and synthetic phenomenology (Aleksander, Awret et al. 2008b; Chrisley 2009b), or emotions (Ziemke 2008a), and so forth.

Furthermore, consciousness is a not only a technical challenge but also a theoretical feat. In this paper, I would like to address two lines of enquiry. On the one hand, I would like to consider and to list a series of fundamental scientific problems that consciousness research cannot set aside. On the other hand, I would like to consider a series of approaches and I will briefly evaluate their pros and cons.

Among the main scientific issues, I would list:

- Cognitive unity
- Intentionality
- Representation
- Freedom
- Temporal integration
- Feeling vs. functing

These issues are paramount both because they are correlated with conscious experience and because they poses a formidable obstacle to our scientific understanding of the nature of consciousness.

In short, the issue of consciousness is still controversial and full of obstacles. We do not yet know how to tackle with it nor how to measure our success. On this regard, Jaegwon Kim stated that (Kim 1998)

> we are not capable of designing, through theoretical reasoning, a wholly new kind of structure that we can predict will be conscious; I don't think we even know how to begin; or indeed how to measure our success.

Yet it may be a necessary step in devising and building a true autonomous and efficient intelligence machine – a machine with a mind. After all, the lack of a formal definition is not necessarily an obstacle that prevents any progress (Koch 2004):

> Historically, significant scientific progress has commonly been achieved in the absence of formal definitions. For instance, the phenomenological laws of electrical current flow were formulated by Ohm, Ampère, and Volta well before the discovery of the electron in 1892 by Thompson. For the time being, therefore, I adopt [a] working definition of consciousness and will see how far I can get with it.

## 2. Strong and Weak Machine Consciousness

The recent upsurge of interest and optimism as to the possibility of modeling and implementing conscious machines or conscious agents (Buttazzo 2001; Holland 2003; Holland 2004; Adami 2006; Chella and Manzotti 2007; Aleksander 2008; Aleksander, Awret et al. 2008a; Buttazzo 2008; Chrisley 2008; Manzotti and Tagliasco 2008) should not lead anyone to underestimate the many critical issues lurking in the background.

Machine consciousness is not simply a technological issue, but rather a field that poses old unanswered questions such as the relation between information and meaning, the nature of teleology, the unity of the self, the nature of phenomenal experience, and many others. Like psychology, it can be observed that machine consciousness has a long past and a very brief history (Ebbinghaus 1908). Although the term is fairly recent (first time Nemes 1962), the problem has been addressed since Leibniz's mill. Machine consciousness offers the opportunity to deal with the hard problem of consciousness from a different perspective – a fact already clear 40 years ago when Hilary Putnam wrote that (Putnam 1964, p. 669)

> What I hope to persuade you is that the problem of the Minds of Machines will prove, at least for a while, to afford an exciting new way to approach quite traditional issues in the philosophy of mind. Whether, and under what conditions, a robot could be conscious is a question that cannot be discussed without at once impinging on the topics that have been treated under the headings Mind-Body Problem and Problem of Other Minds.

Machine consciousness is a promising field of enquiry for at least two reasons. First, it assumes that consciousness is a real phenomenon affecting behavior (Jennings 2000; Koch 2004; Miller 2005; Seth, Dienes et al. 2008). Secondly, it suggests the possibility to reproduce, by means of machines, the most intimate aspect of our mind – namely conscious experience. Although many argued against the possibility of machine consciousness mostly because of a priori assumptions ("no machine will ever be like a man"), no one has conclusively argued against such a possibility so far. Biological chauvinism does not seem move from convincing arguments.

Besides, any argument that seems to deny the possibility of machine consciousness is faulty insofar as the same argument would deny the very possibility of human consciousness. For instance, a naïve adversary of machine consciousness may argue that since CPUs and computer memory do not seem to be the right kind of stuff to harbor phenomenal experience, a computer will never be conscious. And yet, borrowing Lycan's words if such (Lycan 1981, p. 37-38)

> pejorative intuition were sound, an exactly similar intuition would impugn brain matter in just the same way […]: 'A neuron is just a simple little piece of insensate stuff that does nothing but let electrical current pass through it from one point in space to another; by merely stuffing an empty brainpan with neurons, you couldn't produce *qualia-immediate phenomenal feels*!' – But I could and would produce feels, if I knew how to string the neurons together in the right way; the intuition expressed here, despite its evoking a perfectly appropriate sense of the eeriness of the mental, is just wrong.

Contrary to classic AI and functionalism, machine consciousness enthusiasts seem to consider that the classic functional view of the mind in terms of either functions or modules (a la Dennett, so to speak) is insufficient to grasp the full scope and capacity of a conscious agent. Therefore, the traditional arguments against strong AI – for instance, Searle's Chinese Room or Block's Chinese nation argument – loose some of their strength. A machine is not necessarily a Turing machine. In fact, although most available machines are instantiations of von Neumann's blue print, other architectures are becoming available. There is no a priori reason why a machine has to be an instantiation of a Turing machine. Views – such as embodiment, situatedness, and externalism – challenge the classic AI disembodied view of a syntactical symbol-crunching machine (Chrisley 1995; Hirose 2002; Shanahan 2005b; Pfeifer, Lungarella et al. 2007a).

Roughly speaking, machines consciousness lies in the middle between biological chauvinism (only brains are conscious) and liberal functionalism (any functional systems behaviorally equivalent is conscious). Its proponents maintain that biological chauvinism could be too narrow and yet they concede that some kind of physical constraints is unavoidable (no multiple realizability).

Recently, many authors emphasized the alleged behavioural role of consciousness (Baars 1988; Aleksander and Dunmall 2003; Sanz 2005; Shanahan 2005b) in an attempt to avoid the problem of phenomenal experience.

Owen Holland suggested that it is possible to distinguish Weak Artificial Consciousness from Strong Artificial Consciousness (Holland 2003). The former approach deals with agents that behave as if they were conscious, at least in some respects. Such view does not need any commitment to the hard problem of consciousness. On the contrary, the latter approach deals with the possibility of designing and implementing agents capable of real conscious feelings.

Although the distinction between weak and strong artificial consciousness sets a useful temporary working ground, it may suggest a misleading view. Setting aside the crucial feature of the human mind – namely phenomenal consciousness – may divert from the understanding of the cognitive structure of a conscious machine. Skipping the so-called "hard problem" could not be a viable option in the business of making conscious machines.

The distinction between weak and strong artificial consciousness is questionable since it is not matched by a mirror dichotomy between true conscious agents and "as if" conscious agents. Yet, human beings are conscious and there is evidence that most animals exhibiting behavioural signs of consciousness are phenomenally conscious. It is a fact that human beings have phenomenal consciousness. They have phenomenal experiences of pains, pleasures, colors, shapes, sounds, and many more other phenomena. They feel emotions, feelings of various sort, bodily and visceral sensations. Arguably, they also have phenomenal experiences of thoughts and of some cognitive processes. Finally, they experience being a self with a certain degree of unity. Human consciousness entails phenomenal consciousness at all levels.

In sum, as mentioned above, it would be very bizarre whether natural selection had selected consciousness without any selective advantage. Thus, we cannot but wonder whether it could be possible to design a conscious machine without dealing squarely with the hard problem of phenomenal consciousness. If natural selection went for it, we strongly doubt that engineers could avoid doing the same. Hence it is possible that the dichotomy between phenomenal and access consciousness – and symmetrically the separation between weak and strong artificial consciousness – is eventually fictitious.

While some authors adopted an open approach that does not rule out the possibility of actual phenomenal states in current or future artificial agents (Chella and Manzotti 2007; Aleksander, Awret et al. 2008a), other authors (Manzotti 2007; Koch and Tononi 2008) maintained that a conscious machine is necessarily a phenomenally conscious machine. For them to be conscious is necessarily having phenomenal experiences or having P-consciousness (Block 1995). For instance, Giulio Tononi suggested that the kind of information integration necessary to exhibit a human level of cognitive autonomy is associated to the emergence of consciousness (Tononi 2004a).

## 3. Scientific Issues

This paragraph will sketch the scientific, theoretical and philosophical issues at the roots of machine consciousness (indeed often of consciousness itself). Too often researchers accept assumptions that are very far from being justified either empirically or theoretically. As a result, many years have been wasted in pursuing goals on the basis on unwarranted premises.

For one, there is no reason why consciousness should be related to biology. So far, no one has ever been able to suggest any kind of necessary link between the carbon-based molecules featured by living organisms and consciousness. For instance, at a meeting sponsored in 2001 at the Cold Spring Harbour Laboratories addressing the question 'Could Machines Be Conscious?', the participants agreed on the fact that there is no known law of nature that forbids the existence of subjective feelings in artifacts designed or evolved by humans. And yet machine consciousness poses many scientific issues that are worth of attention. I will briefly consider each of them.

### *A. Embodiment*

A much heralded crucial aspect of agency has been embodied cognition (Varela, Thompson et al. 1991/1993; Clark 1997; Ziemke and Sharkey 2001; Pfeifer and Bongard 2006). It cannot be in any way underestimated the importance of the interface between an agent and its environment, as well as the importance of an efficient body, it is far from clear whether this aspect is intrinsically necessary to the occurrence of consciousness.

Although we believe that a body is indeed a necessary condition, we wonder whether there had been any clear understanding of embodiment.

Apart from intuitive cases, when is an agent truly embedded? In some sense, there is no such a thing as a not embodied agent, since even the classic AI algorithm has to be implemented as a physical set of instructions running inside a physical device. In some other sense, even a complex robot such as ASIMO is not embodied. It has a very centralized inner controller computing everything. There are many examples of biological agents that would apparently score very well as to embodiment and that do not seem good candidate for consciousness. Take insects, for instance. They show impressive morphological structure that allows them to perform outstandingly well without a very sophisticated cognitive capability.

The notion of embodiment is probably a lot more complex than the simple idea of having a body and controlling actuators and sensors. It refers to the kind of development and causal processes engaged between an agent, its body, and its environment.

### *B. Situatedness*

Besides having a body, a conscious agent could need also being part of a real environment. Yet this is controversial. For instance, many authors argued that consciousness could be a purely virtual inner world created inside a system that, to all respects, could avoid any true contact with a real world (Lehar 2003; Metzinger 2003; Grush 2004). They seem to advocate the possibility of a conscious brain in a vat. Yet we have no empirical evidence that an isolated brain would ever be conscious. There are no known real cases. To this extent, the possibility of a pure virtual phenomenal experience is bizarre, and this bizarreness dims its appeal considerably.

If a consciousness requires embodiment and situatedness, a definition of situatedness would be necessary.

Usually, alleged embodied robots such as Brook's agents, Babybot, Passive walkers, and similar (Brooks, Breazeal et al. 1999; Collins, Wisse et al. 2001; Metta and Fitzpatrick 2003; Paul, Valero-Cuevas et al. 2006) are regarded as examples of integration with the environment since they outsource part of their cognitive processes to smart morphological arrangements that allow greater efficiency or simpler control. Yet this could be a unwarranted premise.

True situatedness may involve some kind of developmental integration with the environment such that the behavioral and teleological structure of the agent is the result of past interactions with the environment. A real integrated agent is an agent that changes in some non-trivial way (which has to be better understood) as a result of its tight coupling with the environment. The aforementioned artificial agents lack this kind of development: they remain more or less the same.

Another fruitful approach is represented by those implementations that outsource part of the cognitive processes to the environment (Brooks 1991). For instance, the field of epigenetic robotics is strongly interested in designing robots capable of developing accordingly with the environment (Metta, Sandini et al. 1999; Zlatev 2001; Bongard, Zykov et al. 2006).

### *C. Emotions and motivations*

It has been maintained that emotions are key to consciousness. For instance, Damasio suggested that there is a core consciousness supporting the higher forms of cognition (Damasio 1999). Although this is a fascinating hypothesis, it remains unclear how emotions should be implemented. Many roboticists draw inspiration from various emotional models (Manzotti 1998; Arkin 2003; Breazeal 2003; Fellous and Arbib 2003; Trappl, Petta et al. 2003; Arbib and Fellous 2004; Minsky 2006; Ziemke 2008b). However, in which case an architecture is really equipped with emotion? When are emotions more than labels on cognitive modules?

Furthermore, it may be the case that emotions depends on consciousness. Another misleading approach has been that offered by the ubiquitous Kismet often described as a robot with emotions (Breazeal 2003). Kismet has nothing to do with emotions apart mimicking them in front of their users. The robot does not contain any convincing model of emotions but only an efficacious hard-wired set of behaviors for its captivating robotic human-like facial features. In Kismet case, it is not altogether wrong saying that emotions are in the eye of the human beholder.

### *D. Unity and causal integration*

Consciousness seems to depend on the notion of unity. Yet what does it give unity to a collection of parts, being them events, parts, processes, computations, instructions? The ontological analysis has not gone very far (Simons 1987; Merrick 2001) and neuroscience wonders at the mystery of neural integration (Revonsuo 1999; Hurley 2003). Machine consciousness has to face the issue of unity. Would be enough to provide a robot with a series of capabilities for the emergence of a unified agent? Should we consider the necessity of a central locus of processing or the unity would stem out of further unexpected aspects?
Classic theories of consciousness are often vague as to what gives unity to a scattered collection of processes. For instance, would the Pandemonium like community of software demons championed by Dennett (Dennett 1991)

become a whole? Has software unity out of its programmer's head? Would embodiment and situatedness be helpful?

A novel approach to the problem of unity is the notion of integrated information introduced by Giulio Tononi (Tononi 2004a). According to him, certain ways of processing information are intrinsically integrated because they are going to be implemented in such a way that the corresponding causal processes are entangled together. Although still in its initial stage, Tononi's approach may cast a new light on the notion of unity in an agent.

### *E. Time, duration or present*

Conscious experience is located in time. Human beings experience the flow of time in a characteristic way that is both continuous and discrete. On one hand, there is the flow of time in which we float seamlessly. On the other hand, our cognitive processes require time to produce conscious experience. Surprisingly, there is evidence that half a second of continuous nervous activity is necessary in order to be visually aware of something (Libet 2004).

Furthermore, the classic Newtonian time fits very loosely with our experience of time. According to Newton, only the instantaneous present is real. Everything had to fit in such Euclidean temporal point. Such a present has no duration. For instance, speed is nothing more than the value of a derivative and can be defined at every instant. We assume to occupy only an ever-shifting width-less temporal point. The Einstein-Minkowsky space-time model expresses this view (Minkowsky 1908) – time is a geometrical dimension in which the present is a point with no width. Such an instantaneous present cannot accommodate the long-lasting and content-rich conscious experience of present.

Neuroscience faces similar problems. According to most neuroscientists, every conscious process is instantiated by patterns of neural activity extended in time. This apparently innocuous hypothesis hides a possible problem. If a neural activity spans in time (as it has to do so since neural activity consists in trains of temporally distributed spikes), something that takes place in different instants of time has to belong to the same cognitive or conscious process. For instance, what glue together the first and the last spike of a neural activity underpinning the perception of a face? Simply suggesting that they occur inside the same window of neural activity is like explaining a mystery with another mystery. What is a temporal window? And how does it fit with our physical picture of time? Indeed, it seems to be at odds with the instantaneous present of physics.

In the case of machines, this issue is extremely counterintuitive. For instance, let us suppose that a certain computation is identical with a given conscious experience. What would happen if we would purposefully slow down the speed of such a computation? Certainly, we may envisage an artificial environment where the same computation runs at an altered time (for instance, we may slow down the internal clock of such a machine). Would the alleged conscious machine have a slowed but otherwise identical conscious experience?

A related issue is the problem of the present. As in the case of brains, what does define a temporal window? Why are certain states part of the present? Does it depend on certain causal connections with behavior or is it the effect of some intrinsic properties of computations?

Machine consciousness may require a change in our basic notion of time.

### *F. Will, freedom, and mental causation*

Another issue, which does not math with the standard scientific picture of reality, is the fact that a conscious subject seems capable of a unified and free will. The topic is as huge as a topic can be (for a comprehensive review see Kane 2001). The problem connects with the so-called problem of mental causation and top-down causation. If a subject is nothing more than the micro-particles constituting it (and their state also), all causal powers are drained by the smallest constituents. In other words, you and I can't have a will different from what all the particles constituting us do (Kim 1998). If this were true, there will be no space left for any level apart from the lowest one. All reality would be causally reduced to what happens at the micro-particles level. No top-down causation would be possible and no space would remain for the will.

Yet, we have a strong (although possibly wrong) intuition that human beings are capable of influencing their behavior and thus that conscious will makes a difference in the course of events. Many philosophers defended conscious will efficacy (Searle 1992).

Another threat for free will comes from Benjamin Libet's famous studies that showed that awareness of one's own choices follows neural activity by roughly 300 ms (Libet 1985). Although Libet left open the possibility that our consciousness can veto brain deliberations, there is still a lot of controversy about the best interpretation of his experimental results.

In short, a huge open problem is whether a system *as a whole* can have any kind of causal power over its constituents. Since consciousness seems to depend on the system as a whole, a theory of consciousness should be able to address the relation between wholes and parts.

As to machines, the aforementioned issue is quite difficult. The classic mechanistic approach and several respected design strategies (from the traditional *divide et impera* rule of thumb, to sophisticated object-oriented programming languages) suggested to conceive machines as made of separate and autonomous modules. As a result, machines are expected to be cases of physical systems whereas the parts completely drain the causal power of the system as a whole. From this point of view, machines are completely unsuited to endorse a conscious will.

However, two possible approaches can provide a viable escape route out of this blind alley.

The first approach consists in recent connectionist approaches stressing the kind of connectivity between elementary computational units. According to such approaches, it could be possible to implement network whose behavior would stem out of the integrated information of the system as a whole (Tononi 2004a). In short, machines would not have to be mechanistic, after all.

The other approach stresses the teleological roles of certain feedback loops that could do more than classic control feedbacks. Here, the idea is to implement machines capable of modifying their teleological structure in such a way as to pursue new goals by means of a tight coupling with their environment. Thus, the behavior of the agent would be the result of all its history as a whole. There would not be separate modules dictating what the agent has to do, but rather the past history as a whole would reflect in every choice (Manzotti and Tagliasco 2005a).

### *G. Representation*

One of the most controversial problem in philosophy of mind is that of representation. How is it possible that something represent something else? We face an apparent insurmountable problem. If the physical world were made only of extensional entities that do not refer to anything, the physical world could not possess any semantics. In fact, nobody knows why subject may have semantics in a physical world. The classic Searle's argument suggests that machines could not have intrinsic intentionality and thus are devoid of semantics. If this were true, machines will never be conscious since they will be only syntactic engines. Unfortunately, at the best of our knowledge, the same arguments would rule out brain semantics, too. Why are brains different from machines? Searle's suggestion that brains have special causal powers has never been too persuasive.

Since it is a fact that we have a conscious representation of the world, it conceivable that we need to reframe our view about the physical world in order to accommodate the apparently impossible fact of representation. All attempts to naturalize semantics, intentionality, and representations (with all the known differences among these terms) either failed or did not succeed enough (Millikan 1984; Dretske 1995; Fodor 1998; Tye 2002). How can symbols been grounded with other facts in the world (Harnad 1990; Harnad 1995)?

It is curious that neuroscience is tempted by the metaphors introduced by computer science in order to provide (incomplete) explanations of the activity of the brain (Bennett and Hacker 2003). The current debate about the existence of a neural code or about mental imagery are deeply indebted with the computer science view of the mind. Why should there be a code in the brain and why should a code provide any justification of brain semantics? In short, I am suspicious of any argument that seems to apply different criteria in biological and in artificial contexts.

In sum, to address the issue of conscious machines, we need to address the issue of representation avoiding any circularity. What does it change a physical process (or state) into a representation of another physical process (or state)?

### *H. Feeling vs functing, or quantitative vs qualitative*

Finally, the allegedly most conspicuous problem – namely how can a physical system produce subjective qualitative phenomenal content? At sunset, we receive boring light rays on our retinas and we experience glorious symphony of colors. We swallow molecules of various kinds and, as a result, we feel the flavour of a delightful Brunello di Montalcino:

> Consciousness is feeling, and the problem of consciousness is the problem of explaining how and why some of the functions underlying some of our performance capacities are felt rather than just "functed." (Harnad and Scherzer 2008)

Famously, Galileo Galilei suggested that smells, tastes, colors, and sounds are nothing without the body of a conscious subject (Galilei 1623). The subject body allegedly creates phenomenal content in some unknown way. A very deep-rooted assumption is the separation between the domain of subjective phenomenal content and the domain of objective physical events. Such assumption deeply intertwines with the deepest epistemological roots of science itself. It is a dogma that a quantitative third-person perspective oblivious of any qualitative aspect can adequately describe physical reality. Yet, many scientists and philosophers alike questioned the soundness of such a distinction as well as our true understanding of the nature of the physical (Mach 1886; James 1905; Eddington 1929/1935; Bohm 1990; Manzotti 2006; Strawson 2006).

Whether the mental world is a special construct concocted by some irreproducible feature of most mammals is still an open question. There is neither empirical evidence nor theoretical arguments supporting such a view. In the lack of a better theory, many scholars wonder whether would not be wiser to take into consideration the rather surprising idea that the physical world comprehends also those features that we usually attribute to the mental domain (Skrbina 2009). In short, many suspects that some form either of panpsychism or of pan-experientialism ought to be seriously considered.

In the case of machines, how is it possible to take over the so called *functing* vs. *feeling* divide (Lycan 1981; Harnad and Scherzer 2008)? As far as we know, a machine is nothing more than a collection of interconnected modules each functioning in a certain way. Why the functional activity should transfigure in the feeling of a conscious experience? Yet, as it happened for other issues, the same question may be asked about the activity of neurons. Each neuron, taken by itself, does not score a lot better than a software module or a silicon chip as to the emergence of feelings. So one possibility remains: it is not a problem of the physical world but rather of our picture of the physical world. We may discount a too simplistic view of the physical world. Machines are part of the same physical world that produced conscious human subjects, thus they could take advantage of the same relevant properties and features.

### *I. Other issues*

It is clear that there is a very long list of correlated issued, which I couldn't adequately address here – 1$^{st}$ person vs 3$^{rd}$ person perspectives, intentionality, qualia, relation between phenomenal content and knowledge, special physical phenomena (usually described by quantum laws), mental imagery, meaning, symbol grounding, and so on. It is also true that, while some of these issues partially overlap with the above mentioned topics, some have their own specificity. In general, all problems share a similar structure with respect to machine consciousness: as long as something seems preventing a machine from being conscious, the same condition would deny a brain to be so. Yet, human beings are conscious and thus we should conclude that there must be some mistake in our assumptions about that conditions that apparently deny the very possibility of a conscious physical system.

## 4. Current Approaches to Machine Consciousness

Although the field is still in its infancy, a few attempts are worth of some consideration. This chapter does not pretend to provide an exhaustive description of these efforts. However, it will be sufficient to overview the ongoing projects.

### *A. Autonomy and resilience*

A conscious agent is a highly autonomous agent. It is capable of self development, learning, self-observation. Is the opposite true?

According to Sanz, there are three motivations to pursue artificial consciousness (Sanz 2005): 1) implementing and designing machines resembling human beings (cognitive robotics); 2) understanding the nature of consciousness (cognitive science); 3) implementing and designing more efficient control systems. The third goal overlaps with the issue of autonomy. A conscious system has to be able to take choices in total autonomy as to its survival and the achievements of its goals. Many authors believe that consciousness endorses a more robust autonomy, a higher resilience, a more general problem solving capability, reflexivity, and self-awareness.

A conscious agent is thus characterized by a strong autonomy that often leads also to resilience to an often huge range of disturbances and unexpected stimuli. Many authors addressed these aspects trying to focus on the importance of consciousness as a control system. Taylor stressed the relation between attention and consciousness (Taylor 2002; Taylor 2007; Taylor 2009) that will be sketched at greater length below. Sanz *et al.* aims to develop a full-fledged functional account of consciousness (Sanz 2005; Sanz, Lopez et al. 2007; Hernandez, Lopez et al. 2009). According to their view, consciousness necessarily emerges from certain, not excessively complex, circumstances in the dwelling of cognitive agents. Finally, it must be quoted Bongard who is trying to implement resilient machines able to recreate their internal model of themselves (Bongard, Zykov et al. 2006). Though he does not stress the link with consciousness, it has been observer that a self-modeling artificial agents has many common traits with a self-conscious mind (Adami 2006).

### *B. Phenomenal experience in machines*

What about explicitly addressing phenomenal experience in machines? There are two approaches, apparently very different: the first approach tries to mimic the functional structure of a phenomenal space (usually vision). The advantage is that it is possible to build robots that exploit the phenomenal space of human beings. For instance, Chrisley is heralding the notion of synthetic phenomenology as an attempt "either to characterize the phenomenal states possessed, or modeled by, an artifact (such as a robot); or 2) any attempt to use an artifact to help specify phenomenal states". (Chrisley 2009a, p.53) Admittedly, Chrisley does not challenge the hard problem. Rather his theory focuses on the sensori-motor structure of phenomenology. Not so differently, Igor Alexander defended various versions of depictive phenomenology (Aleksander and Dunmall 2003; Aleksander and Morton 2007) that suggest the possibility to tackle from a functional point of view the space of qualia.

Another interesting and related approach is that pursued by Antonio Chella who developed a series of robots aiming to exploit sensorimotor contingencies and externalist inspired frameworks (Chella, Gaglio et al. 2001; Chella, Frixione et al. 2008). An interesting architectural feature is the implementation of a generalized closed loop based on the perceptual space as a whole. In other words, in classic feedback only a few parameters are used to control robot behavior (position, speed, etc.). The idea behind the robot is to match a global prediction of the future perceptual state (for instance by a rendering of the visual image) with the incoming data. The goal is to achieve a tight coupling between robot and environment. According to these models and implementations, the physical correlate of robot phenomenology would not lie in the images internally generated but rather in the causal processes engaged between the robot and the environment (Chella and Manzotti 2009).

### *C. Self motivations*

It is a fact that artificial agents do not develop their own goals and thus it is fair to suspect that there is a strong link between being conscious and developing new goals. Up to now there was a lot of interest as to how to learn achieving a goal in the best possible way, but not too much interest as to how develop a new goal. For instance, in their seminal book on neural network learning processes Richard S. Sutton and Andrew G. Barto stresses that they design agent in order to "learn what to do – how to map situations to actions – so as to maximize a numerical reward signal [the goal] […] All learning agents have explicit goals" (Sutton and Barto 1998, p.3-5). In other words, learning deals with situations in which the agent seeks "how" to achieve a goal despite uncertainty about its environment. Yet the goal is fixed at design time. Nevertheless, there are many situations in which it could be extremely useful to allow the agent to look for "what" has to be achieved – namely, choosing new goals and developing corresponding new motivations. In most robots, goals are defined elsewhere at design time (McFarland and Bosser 1993; Arkin 1998) but, at least, behavior changes according to the interaction with the environment.

Interestingly enough, in recent years various researchers tried to design agents capable of developing new motivations and new goals (Manzotti and Tagliasco 2005; Bongard, Zykov et al. 2006; Pfeifer, Lungarella et al. 2007) and their efforts were often related with machine consciousness.

*D. Information integration*

A possible and novel approach to this problem is the notion of integrated information introduced by Tononi (Tononi 2004). According to him, certain ways of processing information are intrinsically integrated because they are going to be implemented in such a way that the corresponding causal processes get entangled together. Although still in its final stage, Tononi's approach could cast a new light on the notion of unity in an agent. Tononi suggested that the kind of information integration necessary to exhibit the kind of behavioural unity and autonomy of a conscious being is also associated to certain intrinsic causal and computational properties which could be responsible for having phenomenal experience (Tononi 2004).

*E. Attention*

If consciousness has to play a role in controlling the behaviour of an agent, a mechanism that cannot be overlooked is attention control. Attention seems to play a crucial role in singling out to which part of the world to attend. However, it is yet unclear what is the exact relation between attention and consciousness. Though it seems that there cannot be consciousness without attention (Mack and Rock 1998; Simons 2000), there is not sufficient evidence to support the thesis of the sufficiency of attention to bestow consciousness. However, implementing a model of attention is fruitful since introduces many aspects from control theory that could help in figuring out what are the functional advantages of consciousness. This is of the utmost importance since any explanation of consciousness should be tied down to suitable functional ground truth. A satisfying attention control mechanism could satisfy many of the abovementioned goals of consciousness such as autonomy, information integration, perhaps intentionality.

A promising available model of attention is the CODAM neural network control model of consciousness whose main is to provide a functional account (Taylor and Rogers 2002; Taylor 2003; Taylor 2007). Such model has several advantages since it suggests various ways to speed up the response and the accuracy of the agent.

A main advantage of the CODAM neural network control model is that it provides suggestions as to how the brain could implement it. The central idea is that the functional role of the attention copy signal is endorsed by the corollary discharge of attention movement (which is the reason of the name of the model). The possible neural basis of the CODAM has been addressed at length by Taylor (Taylor 2000; Taylor and Rogers 2002; Taylor 2003; Taylor 2007).

*F. Global workspace and other cognitive models*

A huge area is represented by cognitive models based on some kind of central control structure – often based on the Global Workspace model (Baars 1984) and other likewise cognitive structure. A well-known examples is is Stan Franklin's IDA whose goal is to mimic many high-level behaviors (mostly cognitive in the symbolic sense) gathering together several functional modules. In IDA's top-down architecture, high-level cognitive functions are explicitly modeled (Franklin 1995; Franklin 2003). They aim at a full functional integration between competing software agencies. However, IDA is essentially a functionalist effort. We maintain that consciousness is something more than information processing – it involves embodiment, situatedness and physical continuity with the environment in a proper causal entanglement.

Consider now Gerard Baars' Global Workspace as it has been implemented by Murray Shanahan (Shanahan and Baars 2005; Shanahan 2006). Shanahan's model addresses explicitly several aspects of conscious experience such as imagination and emotion. Moreover, it addresses the issue of sensory integration and the problem of how information is processed in a centralized workspace. It is an approach that, on the one hand, suggests a specific way to deal with information, on the other hand, endorses internalism to the extent that consciousness is seen as the result of internal organization. Consciousness, in short, is a style of information processing (the bidirectional transfer of information from/to the global workspace) achieved through different means – "conscious information processing is cognitively efficacious because it integrates the results of the brain's massively parallel computational resources" (Shanahan 2006, p. 434). He focuses on implementing a hybrid architecture mixing together the more classic cognitive structure of global workspace with largely not symbolic neural networks.

## 5. What is AI still missing?

Although AI achieved impressive results (Russell and Norvig 2003), it is always astonishing the degree of overvaluation that many non-experts seem to stick to. In 1985 (!), addressing the Americal Philosophical Association, Fred Drestke was sure that "even the simple robots designed for home amusement talk, see, remember and learn" (Dretske 1985, p. 23). It is not unusual to hear that robots are capable of feeling emotions or taking autonomous and even moral choices (Wallach and Allen 2009). It is a questionable habit that survives and that conveys false hopes about the status of AI research. For instance, in a discussion about machine consciousness, it has been claimed that not even research-grade robots, but rather Legobots used in first-year undergraduate robot instruction should be able to develop new motivations (Aleksander, Awret et al. 2008a, p. 102-103). If this were true, why do not autonomous machines developing their own agenda in order to deal with their environment surround us?

Such approximate misevaluation of the real status of AI hinders new researchers from addressing objectives allegedly but mistakenly assumed as already achieved. Due to various motivations not all of strict scientific nature, in the past, many AI researchers made bold claims about their achievements so to endorse a false feeling about the effective level of AI research.

In AI, various misunderstandings hamper most approaches to machine consciousness. I list here the possible methodological mistakes that are specific to the field of machine consciousness.

*A. "False goals"*

Due to its vagueness and intrinsic difficulty, the issue of consciousness has often downgraded to some more tractable aspect. This is an example of the mereological fallacy that consists in confusing a problem with a part of it. For instance, it is true that often a conscious agent is also an autonomous agent. However, are we sure that an autonomous agent is necessarily a conscious one? Similar arguments suggest a more cautious approach for other capacities and aspects presented as more or less sufficient for conscious experience: autonomy, embodiment, situatedness, resilience, and so on.

Whether or not consciousness can be reduced to certain capacities or features that are often correlated with the existence of a conscious agent is, to say the least, rather obscure. Along these lines, Giulio Tononi and Cristof Koch argued that consciousness does not require many of the skills that AI researchers strive to emulate in machines (Koch and Tononi 2008, p. 50)

> Remarkably, consciousness does not seem to require many of the things we associate most deeply with being human: emotions, memory, self-reflection, language, sensing the world, and acting in it.

The issue is still controversial. Most machine consciousness enthusiasts would probably argue against such view – more prominently those that associate conscious agency with the capacity either to integrate cognitive skills (Baars 1988; Haikonen 2003; Shanahan 2005b) or to be autonomous, resilient, and embodied (Sanz 2005; Bongard, Zykov et al. 2006).

*B. Labeling*

Very often cognitive scientists, roboticists and AI researchers shows their architecture labeling their boxes with intriguing and suggestive names: "emotional module", "memory", "pain center", "neural network", and so on. Unfortunately, labels on boxes in architecture models constitute empirical and theoretical claims that must be justified elsewhere. To use Dennett's terminology they are "explanatory debts that have yet to be discharged" (Dennett 1978).

Even an uncontroversial term such as "neural network" is loaded with vague references to biological assumptions. The very choice of the name endorses a series of expectations. Probably, if neural networks had been introduced under the sober name of "not linear functional approximator", their explanatory power would not have been the target of high expectations.
Similarly, a frequent, and often reasonable, complaint from machine consciousness skeptics addresses the liberal use of not always justified labels.

*C. Confusion between ontological and explanatory levels*

It is easy to accept the existence of multiple levels of reality co-existent in the same physical system. Why should we not talk of bits or numbers or even images and sounds when referring to computer memories? Yet the explanatory power of multiple levels ought not to be confused with their reality. It is well known that such use could be a powerful source of confusion (Bennett and Hacker 2003). The use of language is not innocent.

For instance, are images really *inside* a computer memory? Are values inside computers really symbols or characters or whatever we take them to be? From a physical perspective, there are different levels of tensions in small capacitors. From another perspective, there are logical values in logical gates. Getting higher and higher, we obtain bits, numbers, array, RGB triplets, and even music and images. We could get even higher and consider the existence of images having a certain content. Yet, are all these levels real or are they just different epistemic perspectives on the same phenomenon?

The trouble is that most of these levels – bits, logical values, numbers, RGB triplets - are properties off a way of thinking about what takes place in our computer; they are not properties of the computer as such. What we think for quite naturally as two *pixels* in an image are nothing but two tensions causally related with what happens on a computer screen. On this Zenon Pylyshyn wrote

> The point here is not that a matrix representation is wrong. It's just that it is neutral with respect to the question of whether it models intrinsic (i.e. architectural) properties of mental images. (Pylyshyn 2003, p. 365)

In short all these levels may be akin to a center of mass, insofar as centers of mass do not exist but are simply epistemic shortcuts to refer to complex mass distributions. In the case of machine consciousness, the problem cannot be postponed since there is, at least, one level that should be real: the level of conscious experience. Yet, why is it real? It is not easy to resist to the reductionist pull draining out from every level except the physical one.

*D. Inside and outside*

Finally, where is the mind and its content located? Inside or outside the body of the agent? So far, both options are not entirely satisfactory and thus the debate keeps going on.
On one side, it would be very simple if we could locate consciousness inside the body of the agent and thus inside future conscious machines. However, such view is not convincing since most mental states (very broadly speaking) are about something that appears to be external to the body. Therefore, mental states should somehow address external

states of affairs (Putnam 1975; Gertler 2007; Lenay and Steiner 2010) – whether they are concepts, thoughts, percepts, objects, events. Unfortunately, there are no available theories explaining how the arrow of the mind could hit the external world and, consequently, many authors opted for a completely internal view of the mind. Since the world cannot get in, either the mental world must be inside the agent from the beginning or it must be concocted inside (Fodor 1983; Metzinger 2003). All these positions can broadly be labeled as cases of internalism.

On the other hand, consciousness refer to the external world that could be constitutive either as content or as vehicle. Maybe, it is so difficult to bring content inside the mind because it remains outside. So we should reframe our model of the agent such as to include the external world (Honderich 2006; Manzotti 2006; Clark 2008). Not only the content of our experience would lie outside our body, but also the vehicles responsible for consciousness may be totally or partially external to the agent's body. Such a twist in our perspective about the limit of the agent endorses those views that consider embodiment and situatedness as relevant factors for a conscious machine.

## 6. Conclusion

I tried to outline the present and foreseeable future state of machine consciousness studies. As it should be clear, machine consciousness is a broad field that stretches and enlarges significantly the traditional ground for mind-body problem discussions. It is both a technological and a theoretical field since it addresses old and new problems using a different approach. Machine consciousness will push many researchers to reconsider some threads left loose by classic AI and cognitive science. It may also be that machine consciousness will succeed in shedding a new light on the thorny issue of consciousness.

## References


Adami, C., (2006), "What Do Robots Dreams Of?" in *Science*, 314(58): 1093-1094.

Aleksander, I., (2008), "Machine consciousness" in *Scholarpedia*, 3(2): 4162.

Aleksander, I., U. Awret, et al., (2008a), "Assessing Artificial Consciousness" in *Journal of Consciousness Studies*, 15(7): 95-110.

Aleksander, I., U. Awret, et al., (2008b), "Assessing Artificial Consciousness" in *Journal of Consciousness Studies*, 15: 95-110.

Aleksander, I. and B. Dunmall, (2003), "Axioms and Tests for the Presence of Minimal Consciousness in Agents" in *Journal of Consciousness Studies*, 10: 7-18.

Aleksander, I. and H. Morton, (2007), "Depictive Architectures for Synthetic Phenomenology" in A. Chella and R. Manzotti, Eds, *Artificial Consciousness*, Exeter, Imprint Academic(30-45).

Arbib, M. A. and J. M. Fellous, (2004), "Emotions: from brain to robot" in *Trends in Cognitive Sciences*, 8 (12): 554-561.

Arkin, R. C., (1998), *Behavior-Based Robotics*, Cambridge (Mass), MIT Press.

Arkin, R. C., (2003), "Moving Up the Food Chain: Motivation and Emotion in Behavior-Based Robots" in J. M. Fellous and M. A. Arbib, Eds, *Who needs emotions? The brain meets the robots*, Oxford, Oxford University Press**:** 35-84.

Baars, B. J., (1988), *A Cognitive Theory of Consciousness*, Cambridge, Cambridge University Press.

Bennett, M. R. and P. M. S. Hacker, (2003), *Philosophical Foundations of Neuroscience*, Malden (Mass), Blackwell.

Block, N., (1995), "On a Confusion about a Function of Consciousness" in *Behavioral and Brain Sciences*, 18: 227-287.

Bohm, D., (1990), "A new theory of the relationship of mind and matter" in *Philosophical Psychology*, 3(2): 271-286.

Bongard, J., v. Zykov, et al., (2006), "Resilient Machines Through Continuous Self-Modeling" in *Science*, 314(5802): 1118-1121.

Breazeal, C., (2003), "Emotion and Sociable Humanoid Robots" in *International Journal of Human Computer Studies*, 59.

Brooks, R. A., (1991), "New Approaches to Robotics" in *Science*, 253: 1227-1232.

Brooks, R. A., C. Breazeal, et al., (1999), "The Cog Project: Building a Humanoid Robot" in C. Nehaniv, Ed., *Computation for Metaphors, Analogy, and Agents*, Berlin, Springer-Verlag(1562)**:** 52-87.

Buttazzo, G., (2001), "Artificial Consciousness: Utopia or Real Possibility" in *Spectrum IEEE Computer*, 34(7): 24-30.

Buttazzo, G., (2008), "Artificial Consciousness: Hazardous Questions" in *Journal of Artificial Intelligence and Medicine*(Special Issue on Artificial Consciousness).

Chalmers, D. J., (1996), *The Conscious Mind: In Search of a Fundamental Theory*, New York, Oxford University Press.

Chella, A., M. Frixione, et al., (2008), "A Cognitive Architecture for Robot Self-Consciousness" in *Artificial Intelligence in Medicine*(Special Issue of Artificial Consciousness).

Chella, A., S. Gaglio, et al., (2001), "Conceptual representations of actions for autonomous robots" in *Robotics and Autonomous Systems*, 34(4): 251-264.

Chella, A. and R. Manzotti, (2007), *Artificial Consciousness*, Exeter (UK), Imprint Academic.

Chella, A. and R. Manzotti, (2009), "Machine Consciousness: A Manifesto for Robotics" in *International Journal of Machine Consciousness*, 1(1): 33-51.

Chrisley, R., (1995), "Non-conceptual Content and Robotics: Taking Embodiment Seriously" in K. Ford, C. Glymour and P. Hayes, Eds, *Android Epistemology*, Cambridge, AAAI/MIT Press**:** 141-166.

Chrisley, R., (2008), "The philosophical foundations of Artificial Consciousness" in *Journal of Artificial Intelligence and Medicine*(Special Issue on Artificial Consciousness).

Chrisley, R., (2009a), "Synthetic Phenomenology" in *International Journal of Machine Consciousness*, 1(1): 53-70.

Chrisley, R., (2009b), "Synthetic Phenomenology" in *International Journal of Machine Consciousness*, 1: 53-70.

Clark, A., (1997), *Being there: putting brain, body and world together again*, Cambridge (Mass), MIT Press.

Clark, A., (2008), *Supersizing the Mind*, Oxford, Oxford University Press.

Collins, S., M. Wisse, et al., (2001), "A Three-dimensional Passive-dynamic Walking Robot with Two Legs and Knees" in *The International Journal of Robotics Research*, 20(7): 607-615.

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

Dennett, D. C., (1978), *Brainstorms: philosophical essays on mind and psychology*, Montgomery, Bradford Books.

Dennett, D. C., (1991), *Consciousness explained*, Boston, Little Brown and Co.

Dretske, F., (1985), "Machines and the Mental" in *Proceedings and Addresses of the American Philosophical Association* 59(1): 23-33.

Dretske, F., (1995), *Naturalizing the Mind*, Cambridge (Mass), MIT Press.

Ebbinghaus, H., (1908), *Abriss der Psychologie*, Leipzig, Von Veit.

Eddington, A. S., (1929/1935), *The nature of the physical world*, New York, MacMillan.

Fellous, J. M. and M. A. Arbib, (2003), *Who needs emotions? The brain meets the robots*, Oxford, Oxford University Press.

Fodor, J. A., (1983), *The Modularity of Mind. An Essay on Faculty Psychology*, Cambridge (Mass), MIT Press.

Fodor, J. A., (1998), *Concepts. Where Cognitive Science went Wrong*, Oxford, Oxford University Press.

Galilei, G., (1623), *The Assayer*.

Gertler, B., (2007), "Content Externalism and the Epistemic Conception of the Self" in *Philosophical Issues*, 17: 37-56.

Grush, R., (2004), "The emulation theory of representation: Motor control, imagery, and perception" in *Behavioral and Brain Sciences*, 27: 377-442.

Haikonen, P. O., (2003), *The Cognitive Approach to Conscious Machine*, London, Imprint Academic.

Harnad, S., (1990), "The Symbol Grounding Problem" in *Physica*, D(42): 335-346.

Harnad, S., (1995), "Grounding symbolic capacity in robotic capacity" in L. Steels and R. A. Brooks, Eds, *"Artificial Route" to "Artificial Intelligence": Building Situated Embodied Agents*, New York, Erlbaum.

Harnad, S., (2003), "Can a machine be conscious? How?" in *Journal of Consciousness Studies*.

Harnad, S. and P. Scherzer, (2008), "First, Scale Up to the Robotic Turing Test, Then Worry About Feem" in *Journal of Artificial Intelligence and Medicine*(Special Issue on Artificial Consciousness).

Haugeland, J., (1985a), "Artificial Intelligence: The very Idea" in, *Mind Design II*, Cambridge (Mass), MIT Press.

Haugeland, J., (1985b), *Semantic Engines: An introduction to mind design*, Cambridge (Mass), MIT Press.

Hernandez, C., I. Lopez, et al., (2009), "The Operative mind: A Functional, Computational and Modeling Approach to Machine Consciousness" in *International Journal of Machine Consciousness*, 1(1): 83-98.

Hirose, N., (2002), "An ecological approach to embodiment and cognition" in *Cognitive Systems Research*, 3: 289-299.

Holland, O., Ed. (2003), *Machine consciousness*, New York, Imprint Academic.

Holland, O., (2004), "The Future of Embodied Artificial Intelligence: Machine Consciousness?" in F. Iida, Ed., *Embodied Artificial Intelligence*, Berlin, Springer**:** 37-53.

Honderich, T., (2006), "Radical Externalism" in *Journal of Consciousness Studies*, 13(7-8): 3-13.

Hurley, S. L., (2003), "Action, the Unity of Consciousness, and Vehicle Externalism" in A. Cleeremans, Ed., *The Unity of Consciousness: Binding, Integration, and Dissociation*, Oxford, Oxford University Press.

James, W., (1905), "A world of pure experience" in *Journal of Philosophy*, 1: 533-561.

Jennings, C., (2000), "In Search of Consciousness" in *Nature Neuroscience*, 3(8): 1.

Kane, R., Ed. (2001), *The Oxford Handbook of Free Will*, New York, Oxford University Press.

Kim, J., (1998), *Mind in a Physical World*, Cambridge (Mass), MIT Press.

Koch, C., (2004), *The Quest for Consciousness: A Neurobiological Approach*, Englewood (Col), Roberts & Company Publishers.

Koch, C. and G. Tononi, (2008), "Can Machines be Conscious?" in *IEEE Spectrum*: 47-51.

Lehar, S., (2003), "Gestalt Isomorphism and the Primacy of Subjective Conscious Experience: A Gestalt Bubble Model" in *Behavioral and Brain Sciences*, 26(4): 375-444.

Lenay, C. and P. Steiner, (2010), "Beyond the internalism/externalism debate: the constitution of the space of perception" in *Consciousness and Cognition*, 19(4): 938-52.

Libet, B., (1985), "Unconscious Cerebral Initiative and the Role of Conscious Will in Volutary Action" in *Behavioral and Brain Sciences*, VIII: 529-566.

Libet, B., (2004), *Mind Time. The Temporal Factor in Consciousness*, Cambridge (Mass), Harward University Press.

Lycan, W. G., (1981), "Form, Function, and Feel" in *The Journal of Philosophy*, 78(1): 24-50.

Mach, E., (1886), *The Analysis of the Sensations*, New York, Dover Publications.

Manzotti, R., (1998), "Emotions and learning in a developing robot", in *Emotions, Qualia and Consciousness*, Casamicciola, Napoli (Italy), World Scientific.

Manzotti, R., (2006), "An alternative process view of conscious perception" in *Journal of Consciousness Studies*, 13(6): 45-79.

Manzotti, R., (2007), "Towards Artificial Consciousness" in *APA Newsletter on Philosophy and Computers*, 07(1): 12-15.

Manzotti, R. and V. Tagliasco, (2005a), "From "behaviour-based" robots to "motivations-based" robots" in *Robotics and Autonomous Systems*, 51(2-3): 175-190.

Manzotti, R. and V. Tagliasco, (2005b), "From behaviour-based robots to motivation-based robots" in *Robotics and Autonomous Systems*, 51: 175-190.

Manzotti, R. and V. Tagliasco, (2008), "Artificial Consciousness: A Discipline Between Technological and Theoretical Obstacles" in *Journal of Artificial Intelligence and Medicine*(Special Issue on Artificial Consciousness).

McGinn, C., (1989), "Can we Solve the Mind Body Problem?" in *Mind*, 98: 349-366.

Merrick, T., (2001), *Objects and Persons*, Oxford, Oxford Clarendon Press.

Metta, G. and P. Fitzpatrick, (2003), "Early integration of vision and manupulation" in *Adaptive Behavior*, 11(2): 109-128.

Metta, G., G. Sandini, et al., (1999), "A developmental approach to visually guided reaching in artificial systems" in *Neural Networks*, 12: 1413-1427.

Metzinger, T., (2003), *Being no one: the self-model theory of subjectivity*, Cambridge (Mass), MIT Press.

Miller, G., (2005), "What is the Biological Basis of Consciousness?" in *Science*, 309: 79.

Millikan, R. G., (1984), *Language, Thought, and other Biological Categories: New Foundations for Realism*, Cambridge (Mass), MIT Press.

Minkowsky, H., (1908), "Raum und Zeit", in *Versammlung Deutscher Naturforscher*, Köln Vortrag.

Minsky, M., (2006), *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind* New York, Simon & Schuster.

Nagel, T., (1974), "What is it like to be a Bat?" in *The Philosophical Review*, 4: 435-450.

Nemes, T., (1962), *Kibernetic Gépek*, Budapest, Akadémiai Kiadò.

Paul, C., F. J. Valero-Cuevas, et al., (2006), "Design and Control of tensegrity Robots" in *IEEE Transactions on Robotics*, 22(5): 944-957.

Pfeifer, R., (1999), *Understanding Intelligence*, Cambridge (Mass), MIT Press.

Pfeifer, R. and J. Bongard, (2006), *How the Body Shapes the Way We Think: A New View of Intelligence (Bradford Books)* New York, Bradford Books.

Pfeifer, R., M. Lungarella, et al., (2007a), "Self-Organization, Embodiment, and Biologically Inspired Robotics" in *Science*, 5853(318): 1088 - 1093.

Pfeifer, R., M. Lungarella, et al., (2007b), "Self-Organization, Embodiment, and Biologically Inspired Robotics" in *Science*, 5853: 1088-1093.

Putnam, H., (1964), "Robots: Machines or Artificially Created Life?" in *The Journal of Philosophy*, 61(21): 668-691.

Putnam, H., (1975), *Mind, language, and reality*, Cambridge, Cambridge University Press.

Pylyshyn, Z. W., (2003), *Seeing and Visualizing. It's Not What You Think*, Cambridge (Mass), MIT Press.

Revonsuo, A., (1999), "Binding and the phenomenal unity of consciousness" in *Consciousness and Cognition*, 8: 173-85.

Russell, S. and P. Norvig, (2003), *Artificial Intelligence. A Modern Approach*, New York, Prentice Hall.

Sanz, R., (2005), "Design and Implementation of an Artificial Conscious Machine", in *IWAC2005*, Agrigento.

Sanz, R., I. Lopez, et al., (2007), "Principles for consciousness in integrated cognitive control" in *Neural Networks*, 20: 938-946.

Searle, J. R., (1992), *The rediscovery of the mind*, Cambridge (Mass), MIT Press.

Seth, A., Z. Dienes, et al., (2008), "Measuring consciousness: relating behavioural out neurophysiological approaches" in *Trends in Cognitive Sciences*, 12(8): 314-321.

Shanahan, M., (2005a), "Global Access, Embodiment, and the Conscious Subject" in *Journal of Consciousness Studies*, 12: 46-66.

Shanahan, M., (2010), *Embodiment and the Inner Life. Cognition and Consciousness in the Space of Possible Minds*, Oxford, Oxford University Press.

Shanahan, M. P., (2005b), "Global Access, Embodiment, and the Conscious Subject" in *Journal of Consciousness Studies*, 12(12): 46-66.

Shannon, C. E. and W. Weaver, (1949), *The Mathematical Theory of Communication*, Urbana, University of Illinois Press.

Simons, P. M., (1987), *Parts. A Study in Ontology*, Oxford, Clarendon Press.

Skrbina, D., Ed. (2009), *Mind that abides*, Dordrecht, John Benjamins Pub.

Sporns, O., (2011), *Networks of the Brain*, Cambridge (Mass), MIT Press.

Strawson, G., (2006), "Does physicalism entail panpsychism?" in *Journal of Consciousness Studies*, 13(10-11): 3-31.

Taylor, J. G., (2002), "Paying attention to consciousness" in *TRENDS in Cognitive Sciences*, 6(5): 206-210.

Taylor, J. G., (2007), "CODAM: A neural network model of consciousness" in *Neural Networks*, 20: 983-992.

Taylor, J. G., (2009), "Beyond Consciousness?" in *International Journal of Machine Consciousness*, 1(1): 11-22.

Tononi, G., (2004a), "An information integration theory of consciousness" in *BMC Neuroscience*, 5(42): 1-22.

Tononi, G., (2004b), "An information integration theory of consciousness" in *BMC Neuroscience*, 5: 1-22.

Tononi, G., (2008), "Consciousness as integrated information: a provisional manifesto." in *Biological Bulletin*, 215: 216-42.

Trappl, R., P. Petta, et al., (2003), *Emotions in Humans and Artifacts*, Cambridge (Mass), MIT Press.

Tye, M., (2002), "Representationalism an the Transparency of Experience" in *Nous*, 36(1): 137-151.

Varela, F. J., E. Thompson, et al., (1991/1993), *The Embodied Mind: Cognitive Science and Human Experience*, Cambridge (Mass), MIT Press.

Wallach, W. and C. Allen, (2009), *Moral Machines. Teaching Robots Right from Wrong*, New York, Oxford University Press.

Ziemke, T., (2008a), "On the role of emotion in biological and robotic autonomy" in *BioSystems*, 91: 401-408.

Ziemke, T., (2008b), "On the role of emotion in biological and robotic autonomy" in *BiosSystems*, 91: 401-408.

Ziemke, T. and N. Sharkey, (2001), "A stroll through the worlds of robots and animals: applying Jakob von Uexküll's theory of meaning to adaptive robots and artificial life" in *Semiotica*, 134 (1/4): 701/46.

Zlatev, J., (2001), "The Epigenesis of Meaning in Human Beings, and Possibly in Robots" in *Minds and Machines*, 11: 155-195.

# Thermodynamic Learning Rule

**Harold Szu**
Catholic University of America, USA
*Corresponding author*: szuharoldh@gmail.com

### Abstract

The 2nd law of thermodynamic governs an open dynamical system at an isothermal equilibrium. Helmholtz proved that such a system operates at the Minimum Free-Helmholtz Energy (MFE) min $H \equiv E - T_o S,$ much like an efficient car engine. An MFE engine has internal combustion energy $E$, with an exhaust entropy $S$ operating at an optimum engine temperature $T_o$. We propose to model the human brain's learning process according to the isothermal equilibrium, assigning an MFE cost function to associated input vector time series', with unknown output features. We further examine the implication of modeling synaptic ion currents among neurons of various inter-connection sizes for 'grey matter boxes' of arbitrary emissivity. We compare these with the normal modes of a single back box Planck black-body of an ideal emissivity radiation curve. As a result, we derive a *Hebbian learning rule* that is consistent with Donald Hebb's original observation in neurophysiology a half century ago. Such an Artificial Neural Network (*ANN)* enjoys a self-referenced *unsupervised learning* process known as regularized *Lagrange Constraint Neural Network (LCNN).* We rigorously solve the space-variant, ill-posed inverse imaging problem called Blind Separation *of Equivalent Planck Source's (EPS's).*

## 1. Introduction

On Earth, the animals which can learn by experience seem to be equipped with: (1) warm blooded brains that provide steady kinetic transport for efficient cellular operations, and (2) the 'power of paired sensors' (pops) which gather vector time series data $\vec{X}_{s_i}(t)$ for self-referenced unsupervised learning. Likewise, humans have symmetrical vector time series sensors (ears, eyes, nostrils, olfactory bulbs, taste buds, limbs extremities) which communicate with each other through the nervous system. The nervous system and the brain must be kept at isothermal equilibrium, $T_o = 37^o C$. These are necessary but not sufficient conditions for intelligent beings. A higher temperature does not necessarily imply smarter or quicker learning. For example, the chicken's brain is in equilibrium at $40^o C$ but they lack the hands and tools necessary to be smarter than humans.

Almost all imaging at a distance produces imagery having an unknown mixture of Equivalent Planck Sources (EPSs) of arbitrary emissivity [*proved in Sect. 1.2*]. Solving the inverse imaging problem requires thermodynamic learning rules.

It turns out that physics dealt with efficient measurement that took into account of robust and reliable results. The linear programming used in compressive sensing [1,2,3] turns out to be a linear approximation of the MFE cost function [*proved in Remark #2*]:

$$min. \quad H = E - T_o S, \tag{1}$$

The MFE is like an efficient car engine which has an internal combustion energy $E$, an exhaust entropy $S$, which operates at the optimum engine temperature $T_o$.

Eq.(1) included CRT&D CS as a special case when (1) the negentropy, –S, was the convex hull L1 minimization $-T_o S \approx T_o \sum_i^k < log s_i >_{s_i}$ , because the class entropy must be real positive ($-log s_i \geq 0$; $s_i \leq 1$); and (2) the internal energy $E$ is analytic function of which the first order Taylor series expansion assumes the linear error slope:

$$\mu_j(\vec{s}^{(o)}) \equiv \left.\frac{\partial E}{\partial s_j}\right]_{\vec{s}=\vec{s}^{(o)}} \approx \left(s_j - s_j^{(o)}\right), \tag{2a}$$

which becomes the LMS L2 similarity.

$$\begin{aligned} E &= E_o + \textstyle\sum_{i=1}^k \frac{\partial E}{\partial s_i}(s_i - s_i^{(o)}) \\ &= E_o + \textstyle\sum_{i=1}^k (s_i - s_i^{(o)})(s_i - s_i^{(o)}) \\ &= E_o + ||s_i - s_i^{(o)}||^2 \end{aligned} \tag{2b}$$

In this paper, we have generalized Eq.(2) with a non-linear 2nd order Lagrange Constrained Neural Network (LCNN).

### 1.1 Application of Planck Law to the"Brain in Box"

Planck discovered a unique irradiance distribution which peaked at a unique wavelength for each isothermal temperature of a perfectly emissive black body, at unit emissivity $\varepsilon = 1$. We compare Planck's fixed black body resonator cavity at arbitrarily constant temperature with our brain. Our brain is kept at a fixed isothermal equilibrium at $37^o C$ , but consists of different sized grey-matter interconnection boxes having arbitrary emissivity $0 < \varepsilon \leq 1$. Despite our grey matter box is not an ideal black body, we suspected this unique peak might be true as long as an arbitrary grey matter body box has a unknown but fixed emissivity [proved in Theorem 1]. In addition to Planck, our inspiration comes from the 'big-bang' perspective of the universe; expanding incessantly without an outside boundary to reflect back any outgoing electromagnetic waves. The Chinese philosopher Lao-tze once said that the largest has no outside while the smallest has no inside. Therefore, the universe cannot be ideal black body but at best a grey body. Nobel Laureates Arno Penzias and David Wilson of Bell Lab measured the comic background radiation and found it is indeed peaked at the twelve hundred microns wavelength (1.2 mm at 160 GHz) having the apparent brightness temperature at $\varepsilon T_K = 0.91x3^o = 2.73^o\ Kelvin$ at the non-ideal emissivity $\varepsilon\sim 0.91$[Remark #4] Our EEG brainwaves are mediated by synaptic ion currents within different sizes of functional grey matter boxes of arbitrary emissivity filled with neurons. The largest box is the Left logical & Right artistic Hemispheres of the size of 7 cm; associative memory Hippocampuses at 3 cm; the 'emotional' Amygdala at 2 cm; and the smallest may be the Pituitary gland grey matter box clock cycles at 1cm. Consequently, our brains support mixture of different modes: delta (0-4 Hz); theta (4-7 Hz); alpha (8-12Hz); beta (13-30 Hz). Another interesting observation is that all *EEG* waves have peak frequencies which are separated equally by 4 Hz intervals. This could correspond to fundamental topological structures among sub-grey-matter boxes.

Our mental activity is thought to be an unknown mixture of *EEG* waves which co-exist at an isothermal equilibrium that might generate like *EPS's* with various peaks mixture along the full *em* spectrum. Human brains can now be non-invasively monitored by a wireless baseball hat, equipped with dry nano-electrodes which utilize a compressive sensing algorithm of sparse linear combinations of *EEG* signals[4]. However, no one has systematically analyzed brainwaves from the inter-nested, "Pandora's brain" made of arbitrarily sized grey matter and fixed emissivity viewpoints. Thus, the myth of telepathy as a 'super-resonance' among different "Pandora's brains" remains. The simple physics is given for the collaboration with neurophysiologists.

*Planck*'s law described a quantized set of simple normal modes of *electromagnetic (em)* waves that oscillate within a black-body resonator cavity, realizing the vanishing *em*-amplitudes at a constant wall temperature $T_K$, kept by an outside large heat reservoir. The Planck heat source is an ideal black box resonator. It supports all positive integer numbers $n = I_+$ of harmonic wavelengths $\lambda = \frac{c_o}{n\nu}$ like a violin string vibrating at the fundamental frequency $\nu$ and the constant speed $c_o$: $\left\{\lambda = \frac{c_o}{n\nu}\middle| n = I_+\right\}$. Use was made of *Einstein* *n*-photon energy formula: $E \Longrightarrow E_n = nh\nu$ to compute *Maxwell-Boltzmann* probability: $\exp\left(-\frac{E}{K_B T_K}\right) \equiv z^n$, resulted in an infinite geometric series of $z \equiv \exp(-\frac{h\nu}{K_B T_K})$:

$$z^1 + z^2 + z^3 + \cdots = z(1 + z + z^2 + ..) = \frac{z}{1-z} = \frac{1}{z^{-1}-1} \quad (1)$$

where the fluctuating vacuum state $n = 0$ was intentionally excluded to result in $z^{-1}$ other than $z$ in Eq.(1) [cf. Remark 6]. Multiplying the density of states $\frac{2h\nu^3}{c_o^2}$, *Planck* derived rigorously and reproduced early laws:

$$I_\nu(T_K) = \frac{2h\nu^3}{c_o^2}\frac{1}{\exp\left(+\frac{h\nu}{K_B T_K}\right) - 1}$$

$$= \begin{cases} \frac{2h\nu^3}{c_o^2}\exp\left(-\frac{h\nu}{K_B T_K}\right); h\nu \gg K_B T_K;\ Wien\ law \\ 2\nu^2 K_B T_K;\ h\nu \ll K_B T_K;\ Rayleigh - Jeans\ Law \end{cases} \quad (2)$$

(*UV catastrophe Ehrenfest*) Q.E.D.

where *Planck*'s constant $h$ =4.1 × 10^-15^ *eV sec;* $K_B T_K = \frac{1}{40} eV$ at room $T_K = 300$ and thus Boltzmann $K_B \approx 0.1\ meV$.

## Theorem 1: Image Processing by Equivalent Planck Sources (EPRs)

If we define the observed apparent irradiance in terms of an apparent brightness temperature $T_{B_\nu}$, then the apparent irradiance $I_{B_\nu}$ is a percentage of true blackbody irradiance that is reduced by incomplete thermal accommodation of absorbed and re-emitted radiation. This causes a non-zero boundary condition and non-ideal emissivity $0 < \varepsilon_\nu \leq 1$:

$$I_{B_\nu}(T_{B_\nu}) \equiv \varepsilon_\nu I_\nu(T_K) \quad (3)$$

Given that Planck sources have unit emissivity, then grey bodies have an arbitrary emissive source which is uniquely defined by the *EPS of apparent brightness temperature in* proportional to the grey body emissivity multiplied with the ideal Blackbody Temperature:

$$T_{B_{\nu_o}} \cong \varepsilon_{\nu_o} T_K \quad (4)$$

*PROOF:*

Use was made of Planck's law Eq.(2) and the definition of apparent irradiance for $0 < \varepsilon_{\nu_o} \leq 1$ Eq.(3),

$$\varepsilon_{\nu_o}\left(\exp\left(\frac{h\nu_o}{K_B T_{B_{\nu_o}}}\right) - 1\right) = (\exp(\frac{h\nu_o}{K_B T_K}) - 1);$$

$$\frac{T_{B_{\nu_o}} K_B}{h\nu_o} = \frac{1}{Log\left[1 + \frac{\exp\left(\frac{h\nu_o}{K_B T_K}\right) - 1}{\varepsilon_{\nu_o}}\right]}.$$

Fig.1 plotted the apparent brightness temperature $T_{B_{\nu_o}}$ against the black body Kelvin temperature $T_K$ at the peak emissivity $\varepsilon_{\nu_o}$ in 10% increments. Q.E.D.

## 1.2 Mixing Matrix

We measured day color triplets per pixel $\vec{X}_s$=(*Blue: 0.2~0.3 μm; Green*: 0.3~0.5*μm* and *Red:* 0.6~0.8 *μm)* or night infrared (IR) triplets $\vec{X}_s$=(*MWIR*: 3~5 *μm*, *LWIR I: 8~10 μm* and *LWIR II: 10~12 μm).* We derived the unknown mixture of discrete heat sources associated with discrete temperatures in the percentage vector $S \Rightarrow S_j \Longrightarrow \vec{\boldsymbol{S}}$; $j = 1,2,3,$ whose probability is related to the Boltzmann entropy formula derived in R,G,B color entities [*Remark #3*]. Typically, we had illuminating source $S_1$, object body

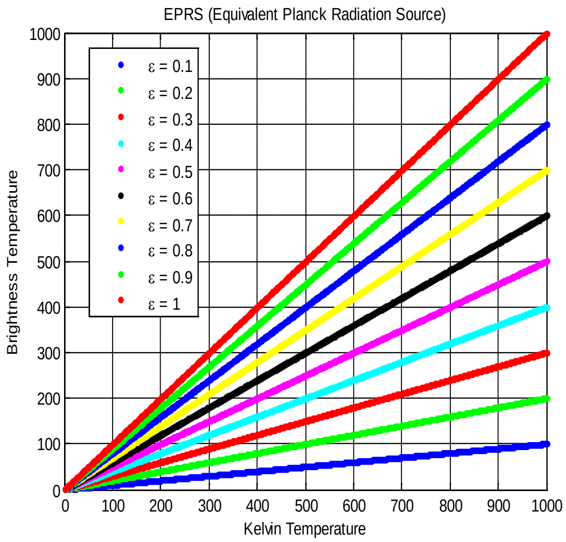



**Fig.1** Defining EPS, we plotted the Brightness temperature $T_{B_{\nu_o}}$ versus the Kelvin Temperature $T_K$. Setting unit $h\nu_o/K_B$=1, we plot: $y = 1/Log(1+\frac{1}{\varepsilon_{\nu_o}}(\exp(\frac{1}{x})-1))$ from 1 to 1000, as we step emissivity $\varepsilon_{\nu_o} \le 1$ in 0.1 increments.

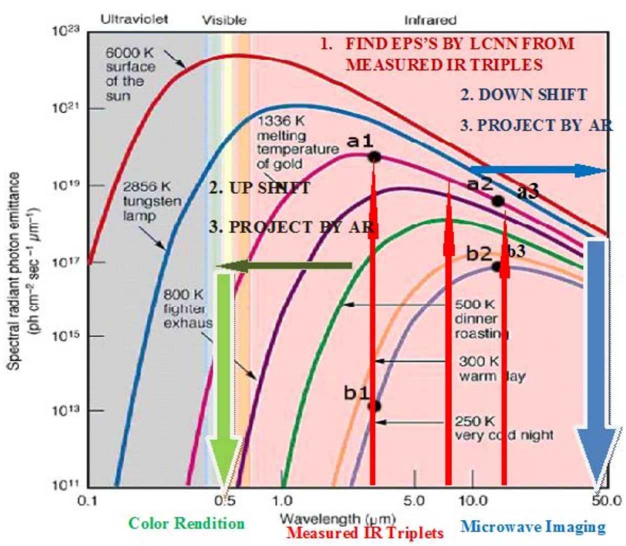



**Fig.2 Compressive Sensing** by means of Equivalent Planck Sources (EPSs): These virtual upper-down spectral extrapolations were possible because the peak wavelength associated one, and only one Kelvin Temperature in a single monotonic curve, where the left-shoulders are up-left-shifted to a shorter sub-micron wavelength toward X-rays. Wilhelm Wien's displacement law ~T^4 followed due to monotonic single peaks per Kelvin temperature, according to the dimensionality analysis of the integrated Eq.(2a).

hot spot source $S_2$, and other sensor coolant source $S_3$, etc. in terms of Boltzmann entropy probability normalization 1=$S_1+S_2+S_3$. The mixing matrix mapped j-temperature sources to spectral *i*-components:

$$\vec{X}_{s_i} = [A_{i,j}]\vec{\boldsymbol{S}} = S_1\,\vec{a}(T_1) + S_2\,\vec{b}(T_2) + S_3\vec{c}(T_3) \qquad (5)$$

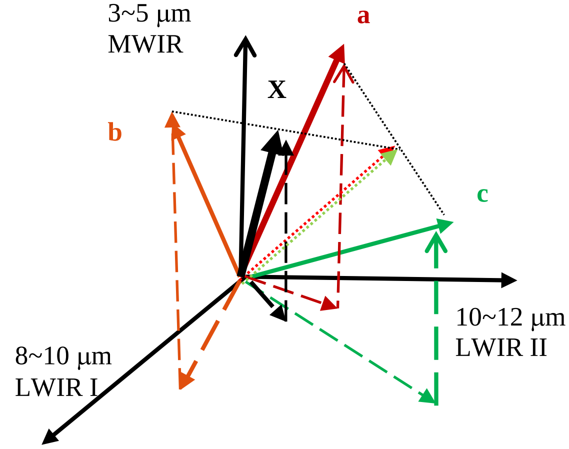



**Fig.3** Three column vectors corresponding to three EPS associated equivalent brightness temperatures.

The *EPS* temperature sources of mixing matrix [A] are usually not known for the generation of ground spectral data. The *IR triplets* per pixel as transposed (*Tr*) row vector $\vec{X}^{Tr} \equiv (X_1 \quad X_2 \quad X_3)$ were measured at the center of each band value, i.e., 4 μm, 9 μm, and 11 μm, respectively, in **Fig.2.** The mixing matrix had three column vectors $\vec{a}(T_1), \vec{b}(T_2)$, and $\vec{c}(T_3)$, corresponding to three EPS associated equivalent brightness temperatures $\vec{a}(330^oK), \vec{b}(265^oK)$, and $\vec{c}(200^oK)$ shown in **Fig.3**. For example,

$$\mathbf{X}(\text{pixel}) \equiv \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = [\boldsymbol{a}(T_1) \quad \boldsymbol{b}(T_2) \quad \boldsymbol{c}(T_3)]\boldsymbol{S}(pixel)$$

$$\equiv \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix} \begin{bmatrix} S_1 \\ S_2 \\ S_3 \end{bmatrix}$$

$$= 10^{-5} \begin{pmatrix} 67.46 & 4.65 & 0.05 \\ 502.82 & 152.39 & 21.41 \\ 449.86 & 168.13 & 33.61 \end{pmatrix} \begin{bmatrix} S_1 \\ S_2 \\ S_3 \end{bmatrix}.$$

Each source $S_j$ might have different percentage values for each pixel such as $\vec{S}^{Tr} = (S_1, S_2, S_3) = (30\%, 50\%, 20\%)$. The space-variant propagation defines an unknown mixing matrix that can vary in space from a group of pixels to others known as different isoplanar or Komogorov regions in astronomy imaging; while the unknown object sources could vary from sub-pixel to pixel for a high definition picture.

Mathematically, Blind EPS Separation (BSS) is challenging because the mixing matrix [*A*] in Eq.(5), mapping the temperature sources to the spectral band values, was unknown for the case of remote sensing, and the inverse weight matrix is typically ill-conditioned:

c1

$$\vec{S}_j = [W_{j,i}]\vec{X}_{s_i}\text{, where } [W_{j,i}] = [A_{i,j}]^{-1} \qquad (6)$$

The following theorems met the challenge.

## 2. Thermodynamic Learning Rule

The key enabler of our new approach is resolving the 'Mexican Standoff' slowdown, or in critical slow down during thermodynamic phase transition phenomena.

**LEMMA:** Due to unsupervised learning, the energy cost function is unknown for image processing at remote sensing. The first order Taylor series becomes 2nd order in the smallness, requiring the second order Taylor series expansion, curvature $C^k$, to determine the Lagrange error slope vector $\mu_\alpha$ together with the estimation error which converges self-consistently.

$$\boldsymbol{E} \cong \boldsymbol{E}^{(o)} + \mu_\alpha([W_{\alpha,\beta}]X_\beta - S_\alpha^{(o)}) + \tfrac{1}{2}C^k|[W_{i,\beta}]X_\beta - S_i^{(o)}|^2 \Rightarrow O(0)^2; \qquad (7)$$

### 2.1 Critical Slowdown Phenomena

When the state function is known, then the 1st order Taylor series derivative is of 1st order smallness. However, Lagrange knew that the critical slowdown phenomena in thermodynamic phase transition occurred because the state function is unknown which required other expansion to determine it, causing the 1st order to become 2nd order in smallness. For example in BSS, the minimum of Helmholtz free energy, $min. H = E - T_o S$, is expanded in a Taylor series with respect to unknown number of heat sources in terms of associated class entropy per pixel, $\boldsymbol{S} \Rightarrow \{\boldsymbol{S}_i \leftrightarrow \vec{\boldsymbol{S}} = [W]\vec{\boldsymbol{X}}_s\}$. We applied the 1st order Taylor series of MFE of the unknown cost function $E$

$$(\frac{\partial \boldsymbol{E}}{\partial \boldsymbol{s}_\alpha})(\boldsymbol{S}_\alpha - \boldsymbol{S}_\alpha^{(o)}) \equiv \boldsymbol{\mu}_\alpha \boldsymbol{c}_\alpha(\vec{\boldsymbol{S}}) \Rightarrow O(0)^2, \qquad (8)$$

where the inner product involves the 2nd order smallness and it is difficult to determine the iteration involving both the data and the slope of data. Simultaneously, the *unknown Lagrange energy slope vector* flattens near the minimum with a zero slope

$$\boldsymbol{\mu}_\alpha(\vec{\boldsymbol{s}}^{(o)}) \equiv \frac{\partial \boldsymbol{E}}{\partial s_\alpha^{(o)}} \Rightarrow O(0)$$

together with the weighted learning of the $\boldsymbol{s}$-spectral vector measurement data $\vec{\boldsymbol{X}}_s$ generating an learning error vector.

$$c_i(\vec{\boldsymbol{s}}^{(o)}) \equiv (\boldsymbol{s}_i - \boldsymbol{s}_i^{(o)}) = [W_{\alpha,\beta}]\boldsymbol{X}_\beta - \boldsymbol{s}_\alpha^{(o)} \Rightarrow O(0). \quad (9)$$

Consequently, the 1st order of MFE involves 2nd order smallness in a product which becomes too small to determine whether it is the cause or the consequence about which one of the product to take the next step that approaches zero. Such double loops of iterations will suffer slow convergence, known in unsupervised learning of the artificial neural network (ANN) community with the nickname 'Mexican standoff', from wild western Hollywood movies. Taylor expansion of the *MFE* Eq.(7) involved unknown internal energy $E$ defines the linear slope as the Lagrange vector constraint of the error slope $\vec{\mu}_j$ and the second order of smallness curvature as the Karush-like penalty function:

$$E = E_o + \sum_{i=1}^{K} \frac{\partial E}{\partial \boldsymbol{s}_i^{(o)}}(\boldsymbol{s}_i - \boldsymbol{s}_i^{(o)}) + \frac{1}{2}\frac{\partial^2 E}{\partial \boldsymbol{s}_\alpha^{(o)} \partial \boldsymbol{s}_\beta^{(o)}}(\boldsymbol{s}_\alpha - s_\alpha^{(o)})(\boldsymbol{s}_\beta - \boldsymbol{s}_\beta^{(o)})$$

$$\cong E_o + \boldsymbol{\mu}_\alpha(\vec{\boldsymbol{s}}^{(o)})(s_\alpha - \boldsymbol{s}_\alpha^{(o)}) + \tfrac{1}{2}C^k|\vec{\boldsymbol{s}} - \vec{\boldsymbol{s}}^{(o)}|^2 ,$$

$$\frac{\partial^2 E}{\partial s_i^{(o)} \partial s_j^{(o)}} \cong C^k \delta_{i,j}; \quad C^k = \beta_o C^{k-1};$$

$$\beta_0 > 0; k = 1,2,3, etc. \qquad (10)$$

**Theorem 2: Hebb learning Rule of Neural Network Weight Matrix**

Given a measured $\boldsymbol{s}$-spectral band vector per pixel location, $\vec{\boldsymbol{X}}_s$: we solve unknown heat sources $\vec{\boldsymbol{S}} = [W]\vec{\boldsymbol{X}}_s$ by Artificial Neural Network unsupervised leaning weight matrix update.

$$\vec{S} = [A]^{-1}\vec{X} \equiv [W]\vec{X}. \qquad (11a)$$

$$[W_{i,j}]^{k+1} = [W_{i,j}]^k - \frac{1}{C^k}\frac{<\vec{\mu}_i^k \vec{X}_j>}{<[\vec{X}\vec{X}^T]>} \qquad (11b)$$

PROOF:

$$H^{(2)} = E_o + \frac{\partial E}{\partial \boldsymbol{s}_\alpha^{(o)}}(\boldsymbol{s}_\alpha - \boldsymbol{s}_\alpha^{(o)}) + \frac{1}{2}\frac{\partial^2 E}{\partial \boldsymbol{s}_\alpha^{(o)} \partial \boldsymbol{s}_\beta^{(o)}}(\boldsymbol{s}_\alpha - \boldsymbol{s}_\alpha^{(o)})(s_\beta - s_\beta^{(o)}) + T_o K_B \sum_{i=1}^{k} \boldsymbol{s}_i \log \boldsymbol{s}_i + (\mu_o - T_o K_B)(\sum_{i=1}^{k} \boldsymbol{s}_i - 1).$$

We may consistently vary $H^{(2)}$ w.r.t. Artificial Neural Network learning weight matrix:

$$\frac{\delta H^{(2)}}{\delta W_{j,i}} = \mu_j X_i + C^k\{[W_{j,\alpha}]X_\alpha - s_j\}X_i = 0$$

We assume the Ergodic hypothesis that the temporal average of high frequency turbulent fluctuations is equivalent to the equilibrium ensemble average with the nearest neighbor 3x3 average, denoted with the angular brackets to make sure the outer product of pixel spectral vector to be full rank for non-singular inverse $<[\vec{X}\vec{X}^T]>^{-1}$.This was consistent with our image resolution assumption, a 3x3 neighborhood resolvable as a single space-invariant macro-pixel.

$$< \vec{\mu}\vec{X}^T > + C^k[W] < \vec{X}\vec{X}^T > - C^k < \vec{s}\vec{X}^T > = 0;$$

$$[W]^{k+1} = \left\{< \vec{s}\vec{X}^T > - \tfrac{1}{C^k} < \vec{\mu}^k\vec{X} >\right\} < [\vec{X}\vec{X}^T] >^{-1}, \quad (12)$$

Use is further made of the definition of weight matrix: $\vec{s} = [W]\vec{X}$, and we can simply the first term of Eq.(12) and derived the learning rule Eq.(11b). Q.E.D.

## 2.2 Remarks on Some Fundamental Concepts

**Hebbian Product Rule:** Donald Hebb discovered that the neuro-biological synaptic junction learning rule is similar to a pipeline flow, that is proportional to how much goes in and how much comes out. The Hebbian product learning rule:

$$W_{i,j} \propto \vec{X}_i \vec{\mu}_j.$$

We demanded a proper normalization, i.e., $\frac{<\vec{\mu}_i^k \vec{X}_j>}{<[\vec{X}\vec{X}^T]>} \to 1$, if $\vec{\mu}_i^k \sim \vec{X}_i$.

What is the thermodynamic learning rule? It's systematic way to guess the most probable inverse source solution by directly computing the maximum probability. By systematic trial and errors, we can de-mix the local mixtures by the MFE principle. There is a finite number of ways that the positive sum of a photon counts can be made. Among them, we choose the lowest energy cases, e.g., giving Beethoven first 3 notes " 5, 5, 1…." , we split the sum 5 = (0+5; 1+4; 2+3; 3+2; 4+1; 5+0) in the unit of energy at temperature $K_BT$=1/40$eV$ for $T$=300°; and find hidden source tones 2+3 and 3+2 occurring twice that have the highest canonical probability 2 $exp(-2/K_BT)exp(-3/K_BT)$. In MFE, we might wish to rule out the rare *high energy* cases (0+5 and 1+4) in favor with lower energy, but *higher chances* in equilibrium (twice 2+3) unless other summations involve also these specific pixels. Given a set of vector measurements of multiple spectral bands, we applied the thermodynamic equilibrium theory to find hidden object sources at MFE by LCNN.

**Helmholtz MFE:** Helmholtz assumed such an open dynamic sub-system within the heat reservoir closed system where Boltzmann heat death at maximum entropy was assumed.

$$\Delta H_{\ object} \equiv \Delta E_{object} - T_0 \Delta S_{object} \leq 0\,;$$

$$\text{min. } H \equiv E - T_o S \qquad (13)$$

PROOF:
Let $S_{Total}$ denote the total entropy of a closed system. Then $S_{Total}$ is the sum of entropy of reservoir and object,

$$S_{Total} = S_{Reservoir} + S_{object}.$$

If the object takes $\Delta E_{object}$ energy from its surroundings, the entropy change of $S_{Reservoir}$ will be $\Delta S_{Reservoir} = -\Delta E_{object}/T_0$ , and the total entropy change is

$$\begin{aligned}\Delta S_{Total} &= \Delta S_{Reservoir} + \Delta S_{object} \\ &= -\frac{\Delta E_{object}}{T_0} + \Delta S_{object} \\ &= -\frac{\Delta E_{object} - T_0 \Delta S_{object}}{T_0} = -\frac{\Delta H_{object}}{T_0} \qquad (14)\end{aligned}$$

where $\Delta H_{object} \equiv \Delta E_{object} - T_0 \Delta S_{object}$ is the change of the object's Helmholtz free energy, which is an analytic state function defined by $H = E - T_o S$. Note that $\Delta S_{Total} > 0$ since the total entropy of a closed system is always increasing, and $\Delta H_{object} \leq 0$ given a positive $T_0$. Q.E.D.

**Boltzmann Entropy:** Ludwig Boltzmann inscribed on his tomb headstone the entropy formula (cf. a picture of his Math Genealogy)

$$S = K_B \, Log\, W;$$

where W is the aforementioned phase space trajectory volume that represents all possibility which an identical macroscopic system can be prepared and realized; and $K_B$ is the Boltzmann constant, i.e., 0.1 $meV$. To be explicit for remote sensing, we considered 3 kinds of identical entities; $R$ denotes the number of red balls/molecules/photons, likewise $G$ & $B$ in a closed system. Thus, the chance of realizing total $N$ balls is $N!$ divided by identical colors $R!\,G!\,B!$, because of the over-counting of permutations with identical particles.

$$W = \frac{N!}{R!G!B!}.$$

Sterling approximation of logarithmic factorial was valid when $N > 10$.

$$Log\, N! \cong N\, Log\, N - N;$$

$$\frac{N}{N} = \frac{R}{N} + \frac{G}{N} + \frac{B}{N};$$

$$1 = S_1 + S_2 + S_3.$$

Then, Boltzmann discrete entropy formula follows:

$$S = -K_B \sum_{i=1}^{K} S_i Log\, S_i - \text{const.}\left(\sum_{i=1}^{K} S_i - 1\right);$$

$$const. = \frac{\mu_o}{T_o} - K_B$$

where the minus sign was derived due to $Log\, S_i \leq 0$ for $S_i \leq 1$ and the Lagrange scalar constraint of the probability norm $\sum_{i=1}^{K} S_i - 1 = 0$ was chosen to be $\left(\frac{\mu_o}{T_o} - K_B\right)$ that insured the normalization and a simple slope:

$$\begin{aligned}\frac{\partial S}{\partial S_i} &= -K_B\left(1 + \sum_{i=1}^{K} \log S_i\right) - \left(\frac{\mu_o}{T_o} - K_B\right) \\ &= -K_B \log S_i - \frac{\mu_o}{T_o} \qquad \text{Q.E.D.}\end{aligned}$$

The maximum entropy in a closed system corresponds to the minimum free energy in open sub-systems.

**Grey-Body Planck Law:** Planck's law was the triumph of modern quantum physics during 1900~1919. Max Planck received the Nobel Prize in Physics in 1918. A remarkable result which we exploited theoretically in this paper was that the spectral irradiance $I_\nu(T_K)$, leaking out of a small hole of the black body cavity's opaque walls kept at a constant temperature, peaked at a single wavelength monotonically $\lambda_o = c_o/\nu_o$ and that uniquely determines the associated Kelvin temperature $T_K$ once and only once in Fig.2. We were not the first one either. Astronomers applied the apparent measured brightness temperature $T_B$ related by the unique peak spectrum value to an equivalent black body Kelvin temperature $T_K$. It turned out roughly $T_B \sim \varepsilon T_K$. Cosmic Background Radiation is not a blackbody for a large expanding universe having no outside, yet the estimation of an approximated reflectivity $\gamma \sim 0.1$ suggests by the conservation of energy, the equivalent emissivity is about $\varepsilon \sim 0.9$. Nevertheless, the universe was cooled down after the Big Bang happened at 13B years; 380K years ago, the universe reached 3000K~0.25 eV, which is the time of hydrogen atoms were formed with 13.6 eV ionization energy (5% light mater, 27% dark mater, 68% dark energy). Thus, the background light did not have enough energy to become de-coupled from the hydrogen matter, and the universe became transparent. As a result, the Cosmic Microwave Background Radiation (CMBR) can be observed and may be called the "time of last (inelastic) scattering." The decoupled photons from matter are continuously cooled down 1000 times at now 2.7 $^oK$ ~0.23meV corresponding to microwave range frequency of 160.2 GHz(1.9 mm wavelength). Robert Dicke, George Gamow, Ralph Alpher, and Robert Herman conjectured that CMBR was the inflationary Big Bang theory. 1978 Nobel Laureates Arno Penzias and David Wilson applied Dicke radiometry of 15 meter horn antenna to measure the peak at $T_B \sim 2.73^o K \sim 0.9x3^0 K$ . Subsequently, NASA's Cosmic Background Explorer (COBE) satellite using differential microwave instruments confirmed an anisotropic CMBR (George Smoot & John Mather, Nobel Prize 2006).

**Quantum Statistics**: G.E. Uhlenbeck & S. Goudsmidt discovered in 1910 in the Stern & Gerlach experiment an electron beam split under an inhomogeneous magnetic field into 2 beans: spin up or spin down: that the spin fine structure constant $2s+1=2$ that implies the e-spin quantum number s=1/2. It suggested the e-wave function with the phase factor $e^{i(n-1)\pi}=(-1)^{n-1}$ generates alternation signs in the Fermi-Dirac distribution function

$$z-z^2+z^3-z^4+\cdots=z(1-z+z^2-z^3+\cdots)$$
$$=\frac{z}{1+z}=\frac{1}{z^{-1}+1}=\frac{1}{\exp\left(+\frac{E}{K_BT}\right)+1}\leq 1$$

for odd integer electron spin Fermion the Pauli's exclusion principle, one pigeon per hole, and led to a finite Fermi surfaces as the Band gap phenomena in the semiconductors. Bose-Einstein condensation of integer spin Bosons is due to the friendship principle: the condensed ground state becomes divergent $1/[\exp(+E/K_BT)-1]\to\infty$ , where $E/K_BT\to 0$. BCS theory of superconductor of electron spin ½ was due to the lattice vibration of bounding two Fermions together called Cooper pairs, electrons or positrons which become a spin-1 Boson. Paul Chu et al. discovered higher temperature superconductor made of ceramic $Y_1Ba_2Cu_3O_x$ material whose lattice defects, positron holes, enjoyed a larger internal pressure due to a larger replaced $Ba_2$ molecule, which bounded two Fermions together, by phonon exchange energy, into a spin-1 Boson and sustained a disruptive thermal noise $\gtrsim 77^oK$.

**Vacuum Fluctuation:** Paul Ehrenfest wrote the corresponding principle between classical mechanics and quantum mechanics. The Poison Bracket is related to Heisenberg uncertainty principle commutator between position $\hat{P}$ and momentum $\hat{Q}$ operators. It represented the effects of two different sequences of measurements. It helped us quantized the Hamiltonian $\hat{H}$ of simple harmonic oscillator in the quantum field theory in terms of 2nd quantization operators $\hat{a}^\dagger$: $\hat{H}=\hat{P}^2+\hat{Q}^2=(\hat{P}+i\hat{Q})(\hat{P}-i\hat{Q})=\hat{a}^\dagger\hat{a}$ that computes the vacuum fluctuation due to the commutator uncertainty principle generating the non-zero vacuum energy $\hat{H}\Big|0>=\frac{1}{2}h\nu\Big|0>$. This zero-point vacuum fluctuation existed everywhere helped Higgs, following the Anderson phase transition model: $\hat{\phi}^4 \approx(\hat{\phi}$-$\phi_+)^2(\hat{\phi}-\phi_-)^2$ (having a lower potential well at symmetric ground state $|\phi_+>$ associated with a non-zero order parameter) condensing the energy into the mass $m=E/c_o^2$ . CERN experiments seemed to have verified the Higgs boson phase transition mechanism.

## 3. Nonlinearly Regularized Lagrange Constraint Neural Network (LCNN)

The Mexican standoff will be regularized by **Karush, Kuhn-Tucker (KKT) 2nd order** penalty by steepening an isotropic sphere. CRT&D CS assumed that the Lagrange slope was the estimation error itself, and no longer is an unknown. However, when the cost function is unknown, we have demonstrated the double iteration at the linear order is at 2nd order smallness, and therefore cannot consistently determined at the 1st order LCNN. We identified the penalty as the 2nd order Taylor series of MFE. This generates a linearly decoupled closed set of 3 equations for solving sources from spectral vector data, as a fast LANCELOT algorithm, called regulated LCNN (Szu, Miao, Qi, SPIE 2007). Given input s-spectral vector data $\boldsymbol{X}_s$ per pixel we give double iteration superscript index $k=I_+\equiv\{1,2,3,etc.\}$.

**Theorem 3: NL Regularized Lagrange Constraint Neural Network (LCNN)** is a fast LANCELOT algorithm of nonlinear optimization. Given lemma on ANN learning matrix $[W_{j,i}]^k$ and the j-component $\boldsymbol{c_j}$ of the EPS sources estimation error vector together, we determine the slope j-component $\boldsymbol{\mu}_j^k$ of the Lagrange error energy by iteration as;

Hebbian rule:

$$[\mathrm{W}_{i,j}]^{k+1}=[\mathrm{W}_{i,j}]^{k}-\frac{1}{\mathrm{C}^k}\frac{<\vec{\mu}_i^k\vec{\mathrm{X}}_j>}{<[\vec{\mathrm{X}}\vec{\mathrm{X}}^{\mathrm{T}}]>}; \quad (15)$$

Lagrange error slope rule:

$$\vec{\boldsymbol{\mu}}_j^{k+1}=\vec{\boldsymbol{\mu}}_j^{k}+C^k\{[W_{j,\alpha}^{k+1}]\vec{\boldsymbol{X}}_\alpha-\vec{\boldsymbol{S}}_j^{k+1}\}\ ; \quad (16)$$

Unknown object sources:

$$T_o\,K_B\log\vec{\mathrm{S}}_j^{\ k+1}+C^k\vec{\mathrm{S}}_j^{\ k+1}=C^k[\mathrm{W}_{j,\alpha}{}^k]\vec{\mathbf{X}}_\alpha+\vec{\boldsymbol{\mu}}_j{}^k-\mu_0{}^k; \quad (17)$$

Curvature Penalty:

$$C^k = \beta_o C^{k-1};\ \beta_0 > 0; k = 1,2,3, etc. \quad (18)$$

PROOF:

The tradeoff between minimum energy and maximum entropy for the most probable configuration requires the 1st and 2nd order Taylor series expansions:

$$H^{(1)} = E_o + \frac{\partial E}{\partial \boldsymbol{s}_\alpha^{(o)}}\left(s_\alpha - \boldsymbol{s}_\alpha^{(o)}\right)$$

$$+T_o K_B \sum_{i=1}^{k} \boldsymbol{s}_i \log \boldsymbol{s}_i + (\mu_o - T_o K_B)(\textstyle\sum_{i=1}^{k} \boldsymbol{s}_i - 1) \quad (19a)$$

$$H^{(2)} = E_o + \frac{\partial E}{\partial \boldsymbol{s}_\alpha^{(o)}}\left(\boldsymbol{s}_\alpha - \boldsymbol{s}_\alpha^{(o)}\right)$$

$$+\frac{1}{2}\frac{\partial^2 E}{\partial \boldsymbol{s}_\alpha^{(o)} \partial \boldsymbol{s}_\beta^{(o)}}\left(\boldsymbol{s}_\alpha - \boldsymbol{s}_\alpha^{(o)}\right)\left(s_\beta - s_\beta^{(o)}\right)$$

$$+T_o K_B \sum_{i=1}^{k} \boldsymbol{s}_i \log \boldsymbol{s}_i + (\mu_o - T_o K_B)(\textstyle\sum_{i=1}^{k} \boldsymbol{s}_i - 1) \quad (19b)$$

We take variation calculus to set flattening extremes to be zero:

$$\delta E^{(1)} = \frac{\delta E^{(1)}}{\delta s_\beta{}^{(o)}} \boldsymbol{c}_\beta\left(\vec{\boldsymbol{s}}^{(o)}\right) \equiv \boldsymbol{\mu}_\beta\{[W_{\beta,\alpha}]\boldsymbol{X}_\alpha - \boldsymbol{s}_\beta{}^{(o)}\} \quad (19c)$$

$$\delta H^{(1)} = \frac{\delta H^{(1)}}{\delta s_j} = -\boldsymbol{\mu}_\beta \frac{\delta c_\beta}{\delta s_j} + T_o K_B \log \boldsymbol{s}_j + \mu_o \quad (19d)$$

$$\delta H^{(2)} = \frac{\delta H^{(2)}}{\delta s_j} = -(\mu_\beta + C^k\{\mathrm{W}_{\beta,\alpha} X_\alpha - s_\beta\}\frac{\delta c_\beta}{\delta s_j}$$

$$+T_o K_B \log s_j + \mu_o \quad (19e)$$

where use is made of isotropic curvature in an increasing constant penalty term for k+1 iteration: $C^{k+1} = \beta_o C^k$ ; $\beta_o \approx 4$, and the simple component selection: $\frac{\delta c_i}{\delta s_j} = -\delta_{i,j}$ .

From the variation of 2nd order (19e) :

$$\delta \mathrm{H}^{(2)} = 0:$$

$$\mathrm{T_o}\,\mathrm{K_B} \log \mathbf{s}_\mathrm{j} + \mathrm{C^k}\mathbf{s}_\mathrm{j} = \mathrm{C^k}[\mathrm{W}_{\mathrm{j},\alpha}]\mathbf{X}_\alpha + \boldsymbol{\mu}_\mathrm{j} - \mu_0 \quad (20a)$$

From the variation of 1st order (19d):

$$\delta \mathrm{H}^{(1)} = 0:$$

$$\boldsymbol{\mu}_\mathrm{j} + \mathrm{T_o}\,\mathrm{K_B} \log \mathbf{s}_\mathrm{j} + \mu_\mathrm{o} = 0 \quad (20b)$$

We obtain the consistency condition, after cancelling the entropy related common terms: $\mathrm{T_o}\,\mathrm{K_B} \log \mathbf{s}_\mathrm{j} + \mu_\mathrm{o}$, the next iteration of the 1st order Lagrange error energy slope $\boldsymbol{\mu}_\mathrm{j}^{\mathrm{k}+1}$ is determined by the 2nd order variation of MFE:

$$\boldsymbol{\mu}_\mathrm{j}^{\mathrm{k}+1} = \boldsymbol{\mu}_\mathrm{j}{}^\mathrm{k} + \mathrm{C^k}\mathbf{c}_\mathrm{j}^{\mathrm{k}+1}; \mathbf{c}_\mathrm{j}^{\mathrm{k}+1} = [\mathrm{W}_{\mathrm{j},\alpha}^{\mathrm{k}+1}]\mathbf{X}_\alpha - \mathbf{s}_\mathrm{j}^{\mathrm{k}+1} \quad (20c)$$

This linear decoupled Lagrange error slope equation is historically called LANCELOT in FORTRAN massive database optimization. The blind sources estimation error provided the next gradient descent of Lagrange energy error slope vector $\boldsymbol{\mu}_\mathrm{j}{}^\mathrm{k}$. De-mixing weight matrix Eq.(15) was proved in Theorem 2. Q.E.D.

## 4. Conclusion

This thermodynamics learning rule may be a paradigm shift for dealing with spectral image processing with thermodynamics. Various applications have been developed and reported in different journals. It might allow us to consider virtually crossing the full electromagnetic spectrum. Compressive modeling and simulation based on NL LCNN will be published in Optical Engineering (Krapels, Cha, Espinola, Szu). IR triplets for seeing through hot fire and cold dust will be published in IEEE Tran IT (Cha, Abbott, Szu). Thermodynamics physics laws and modern applications will be published in Journal of Modern Physics (Szu, Willey, Cha, Espinola, Krapels). Lots more can happen with your participation in Appendix A and Appendix B. MATLAB pseudo source code is given with benchmarked results.

### References


[1].E. J. Candes, J. Romberg, and T. Tao, “Robust Uncertainty Principle: Exact Signal Reconstruction from Highly Incomplete Frequency Information,” *IEEE Trans IT*, 52(2), pp.489-509, Feb. 2006.

[2].E. J. Candes, and T. Tao, “Near-Optimal Signal Recovery from Random Projections: Universal Encoding Strategies,” *IEEE Trans IT* 52(12), pp.5406-5425, 2006.

[3].D. Donoho, “Compressive Sensing,” *IEEE Trans IT*, 52(4), pp. 1289-1306, 2006.

[4].H. Szu, C. Hsu, G.Moon, T. Yamakawa, B.Q. Tran, T.P. Jung, J. Landa, “Smartphone household wireless Electroencephalo-gram hat,” *Applied Computational Intelligence and Soft Computing*, 2012 (Hindawi Internet Publ.).

[5].H. Szu, L. Miao, & H. Qi, “Unsupervised learning with minimum free energy,” Proc. of SPIE, Vol. 6576, IN: *Independent Component Analyses, Wavelets, Unsupervised Nano-Biomimetic Sensors, and Neural Networks V* (edited by Harold H. Szu, Jack Agee), pp. 657605, 2007.

[6].H. Szu & C. Hsu, “Landsat spectral unmixing a la super-resolution of blind matrix inversion by constraint maxent neural nets,” *Proc. of SPIE*, Vol. 3078, pp. 147–160, 1997.

[7].US PTO Patent, Harold Szu, 7355182 “Infrared multi-spectral camera and process of using infrared multi-spectral camera” April 8, 2008; US PTO 7366564, Harold H. Szu et al; “Nonlinear blind demixing of single pixel underlying radiation sources and digital spectrum local thermometer,” April 29, 2008.

[8].H. Szu, C. Hsu, J. Jenkins, J. Willey, J. Landa, “Capture Significant Events with Neural Networks,” *Neural Networks* 29-30(2012) pp. 1-7.

[9].R. Appleby, P. Coward, J. N. Sanders-Reed, “Evaluation of a Passive Millimeter Wave Imaging for wide detection in degraded visual condition,” *SPIE Proc. 7309*, April 2009; P. Coward & R. Appleby, “Comparison of passive millimeter-wave and IR imagery in a nautical environment,” *SPIE Proc. 7309*,” pp. 04-1, 04-8. 2009.

[10].J.H. Cha, A. L. Abbott, and H.H. Szu, “Passive Ranging Redundancy Reduction in Diurnal Weather Conditions,” *SPIE Proc. 8750*, “ICA, Compressive Sampling, Wavelets etc.” (ed. Szu & Dai ) May 1~3, 2013 (Baltimore).

[11].C. Hsu, M.K. Hsu, J. Cha, T. Iwamura, J. Landa, C. Nguyen, H.Szu," Adaptive Compressive Sensing Camera, *SPIE Proc. 8750*, "ICA, Compressive Sampling, Wavelets etc." (ed. Szu & Dai ) May 1~3, 2013 (Baltimore).

[11] S. Vahidinia, "Regolith Radiative Transfer Model; Applications to Saturn's Icy Rings," Ph.D. 2010, UC Santa Cruz.

[12]. J. F. Mustardm, J. E. Hays, " Effects of hyperfine particles on reflectance spectra from 0.3 to 25 microns," *ICARUS 125*, pp, 145-163 (1997).

[13] V. Payan and A. Royer, "Analysis of temperature emissivity separation (TES) algorithm applicability and sensitivity," *Int. J. Remote Sensing V.*25, pp. 15-27, Jan 2004.

[14] P. Vernazza, B. Carry, J. Emery, J.L. Hora, D. Cruikshank, R.P. Binzel, J. Jackson, J. Helbert, A. Maturilli, "Mid-IR spectral variability for compositionally similar asteroids; implications of asteroid particle size distribution," *ICARUS 207*, pp. 800-809 (2010).

[15] B. Hapke, "Bi-direction reflectance spectroscopy I. Theory," *Geophysical Rev.* Vol. 86, pp. 3039-3054,1981.

[16] H. Ren, H. Szu, and J. Buss," Lagrange Constraint Neural Network for Fully constrained Subpixel Classification in Hyperspectral imagery," *Proc. SPIE 4738*, pp. 184-190, 2002.

[17] I. Kopriva, and H. Szu, "Fast LCNN ica for Unsupervised Hyperspectral Image Classifier," *Proc. SPIE 4738*, pp. 169-183, 2002.

[18] H. Ren, and C.I. Chang," target-constrained interference-minimized approach to subpixel target detection for hyperspectral images," *Opt. Eng.* V.39, pp. 3138-3145, 2000.

[19] C.I. Chang and D.C. Heinz, "Constrained subpixel target detection for remotely sensed imagery," *IEEE Trans. Geosci. Remote Sesnign*, Vol. 38, pp. 1144-1158, 2000.


## Appendix A: BSS by Engineering Filter Approach: Pixel Parallelism (at Maximum Output Entropy )

Bell, Sejnowski, Amari & Oja (BSAO) have systematically formulated an *unsupervised learning of ANN algorithm for unknown but identical for space-invariant mixing* by varying the unknown de-mixing weight matrix $[W_{i.j}]$ until nothing but the Max Entropy $S(y_i)$ of the output $y_i = [W_{i,\alpha}]x_\alpha$, where $x_j = [A_{j,\alpha}]s_\alpha$ with measured $x_j$ and unknown $[A_{j,\alpha}]$ and $s_\alpha$. Here, the repeated Greek indices represent the summation. ANN model used a monotonically sigmoid-squashed threshold output

$$y_i = \sigma(x_i) \equiv \{1 + \exp(-[W_{i,\alpha}]x_\alpha)\}^{-1}.$$

that is nonlinear analytic solution of Riccati equation $\frac{dy}{dx} = \mathrm{y}(\mathrm{y}-1)$, for asymptotically binary logic $y = 0$ *or* 1 for no or yes. Since a single neuron learning rule turns out to be massively parallel to N neurons in tensor index notion, for simplicity, we derived for a single neuron to point out why the engineering filter does not follow Hebb's synaptic weight updates. A bona fide unsupervised learning did not have a desirable specific output entropy S(y) became maximized, the de-mixing filtering $[W_{i.j}]$ becoming the inverse of unknown mixing matrix $[A_{j,\alpha}]$. Thus, the filter weight adjustment is defined as:

$$\frac{\delta w}{\delta t} = \frac{\partial S(y)}{\partial w}, \quad \mathrm{S}(y) = -\int f(y) \log f(y) dy$$

$$\Longrightarrow \delta w = \frac{\partial H(y)}{\partial w}\delta t = \{|w|^{-1} + (1-2y)x\}\delta t.$$

<u>Derivation</u>**:** From the normalized probability definitions:

$$\int f(y)dy = \int g(x)dx = 1\,; f(y) = \frac{g(x)}{\left|\frac{dy}{dx}\right|};$$

$$H(y) \equiv -< \log f(y) >_f \ ,$$

we expressed the output pdf in terms of the input pdf with changing Jacobian variables. We exchanged the orders of operation of the ensemble average brackets and the derivatives to compute

$$\frac{\partial H(y)}{\partial w} = \frac{\partial < \log\left|\frac{dy}{dx}\right| >_f}{\partial w} \cong \left|\frac{dy}{dx}\right|^{-1} \frac{\partial\left|\frac{dy}{dx}\right|}{\partial w};$$

Since Riccati equation was satisfied by the sigmoid: $y = [1+\exp(-wx)]^{-1}$; $\frac{dy}{d(wx)} = y(1-y)$, we readily derived the following results by chain rule

$$\frac{dy}{dx} = wy(1-y); \frac{dy}{dw} = xy(1-y).$$

Substituting these results into Max Ent learning rule, one obtains the Bell-Sejnowski equation

$$\frac{\partial H(y)}{\partial w} = [W]^{-1} - (2y-1). \qquad \text{Q.E.D.}$$

The first term computing the inverse matrix $|w|^{-1}$ is not scalable with increasing *N* nodes, while the second term satisfied the Hebbian product rule between bipolar output 2*y*-1 and input *x*. S. Amari et al. at RIKEN assumed the identity $[\delta_{i,k}] = [W_{i,j}][W_{j,k}]^{-1}$ and multiplied the identity through both sides of original non-biological algorithm

$$\frac{dH}{dW_{i,j}}[\delta_{i,k}] = \{[\delta_{i,j}] - (2\vec{y}-1)\vec{y}^T]\}[W_{i,j}]^{-1},$$

where use was made of $y_i = [W_{i,\alpha}]x_\alpha$ to change the input $x_j$to the synaptic gap by its weighted output $y_i$. Amari et al. derived a natural gradient ascend as the final BSAO algorithm in information geometry,

$$\frac{dH}{dW_{i,j}}[W_{i,j}] = \{[\delta_{i,j}] - (2\vec{y}-1)\vec{y}^T]\},$$

which was not in the original gradient direction $dH/dW_{i,j}$ and enjoyed a faster update without the inverse .

**Fast ICA**: Erkki Oja began his ANN learning of nonlinear PCA for pattern recognition in his Ph.D. study 1982.

$$< \vec{x}\vec{x}^T > \hat{e} \ = \lambda\hat{e};$$

$$w' - w = \vec{x}\sigma(\vec{x}^T\vec{w}) \cong < \vec{x}\vec{x}^T > \vec{w};$$

$$\frac{d\vec{w}}{dt} = < \vec{x}\vec{x}^T > \vec{w} \cong \sigma(\vec{x}^T\vec{w})\vec{x} \cong \frac{dK(u_i)}{du_i}\frac{\mathrm{d}u_i}{\mathrm{d}w_i} \equiv \mathrm{k}(\vec{x}^T\vec{w})\vec{x};$$

where Oja changed the unary logic to bipolar hyperbolic tangent logic as $v_i = \sigma(u_i) \approx u_i - \frac{2}{3}u_i^3 \cong \frac{dK(u_i)}{du_i}$; $u_i = w_{i,\alpha}x_\alpha$.

By derivation we obtain therefore the **BSAO** unsupervised learning collectively in a termination condition: It becomes similar to a Kurtosis slope, which suggested to Oja a new contrast function K. The following is the geometric basis of a stopping criterion of unsupervised learning. Taylor expansion of the normalization, and set $|\vec{w}|^2 = 1$:

$$|\vec{w}'|^{-1} = [(\vec{w} + \epsilon\vec{x}k(\vec{w}^T\vec{x}))^T(\vec{w} + \epsilon\vec{x}k(\vec{w}^T\vec{x}))]^{-\frac{1}{2}}$$
$$= 1 - \frac{\epsilon}{2}\, k(\vec{w}^T\vec{x})(\vec{x}^T\vec{w} + \vec{w}^T\vec{x}) + O(\epsilon^2).$$

$$\vec{w}" \equiv \vec{w}'|\vec{w}'|^{-1}$$
$$= \left(\vec{w} + \epsilon\vec{x}k(\vec{w}^T\vec{x})\right)\left(1 - \frac{\epsilon}{2}\, k(\vec{w}^T\vec{x})(\vec{x}^T\vec{w} + \vec{w}^T\vec{x})\right)$$
$$\Delta\vec{w}"=\vec{w}" - \vec{w} = \epsilon\left[\delta_{\alpha,\beta} - w"_{\alpha}w"^{T}_{\beta}\right]\vec{x}_{\alpha}\,\frac{dK(u_{\beta})}{dw"_{i}}.$$

The ICA algorithm lets the joint probability density function be factorized equally according to Max Entropy requirement of white noise. Unfortunately, such a filter approach cannot work for the remote sensing imageries, because the atmospheric turbulence changes spatially rather quickly due to a large pixel footprint which is about a tenth of a squared kilometers (Landsat 30 km$^2$ per pixel). If one were insisting to apply ICA to mix together all spatial (x,y) ***s***-spectral vectors data $\vec{X}_s(x,y,t)$ in parallel, the assumption of spatial invariant mixing matrix will produce inaccurate sources at the **maximum entropy (Max Ent)**. Spatial invariant assumption will not work for a close-up dual infrared spectral band image for screening for cancer, because of the localization of cancer has a strongly space-variant physiology with and without the cancer. Solving the inverse of space-variant impulse response Green's function or optical point spread function (psf) we need new MFE approach to BSS. To improve it, we have explored space invariant imaging (in terms of 3x3 macro-pixels) for macro-pixel joint density factorization ICA method; (Du, Kopriva, Szu, by Non-negative Matrix Factorization IEEE 2005;& by JADE & Fast ICA Op Eng. 2006). Then, the comparison experience help us reformulated in 2007 a nonlinear LCNN, ala KKT penalty, to regularize the linear LCNN for assuming space-invariant result within 3x3 nearest-neighbor macro-pixels per tumor cluster cells, but space-varying among macro-pixels with or without tumor [4]. In remote sensing, the Mexican standoff product challenge was overcome by the 2$^{nd}$ order energy expansion curvature known as KKT nonlinear optimization penalty. Typical 80 spectral band images among 158 Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) (Kopriva, Szu, Fast ICA, SPIE ICA etc.2002) and corresponding BSS out of applying LCNN to 158 channels data sources maps where the color blue means no class, i.e. low probability while red means high probability.

## Appendix B: BSS by Physics Source Approach: Pixel sequentially (Min. Free-Helmholtz Energy (MFE) Nonlinearly regularized LCNN Matlab Code)

### Pseudo-code of LCNN for 3-source model

Given multispectral data vector **x** = $(x_1\ x_2\ \ x_3)^T$ at a single pixel,

Initialize $\mu^0$ = [0 0 0]$^T$, C$^0$ = 1, $\beta_o$ = 4, and W= $A^{-1}$.

Set iteration index *k* = 1, and maximum iteration number ITM

While *k* ≤ *ITM*

  Calculate sources $s^{k+1}$ subject to 0 ≤ s ≤ 1 by solving Eq.(20a):

$$T_0 K_B \ln(s^{k+1}) + C^k s^{k+1} = C^k W^k x + \mu^k - \mu^k_0$$

  If $s^{k+1}$ converges, return. Otherwise continue;

  Update de-mixing matrix Eq.(15):

$$W^{k+1} = W^k - (1/C^k) < \mu^k x^T > / < x x^T >$$

  Update Eq.(20c):

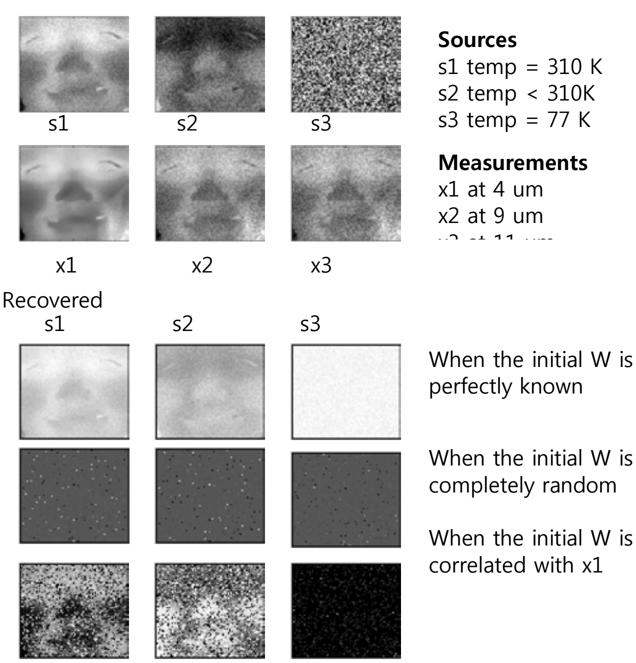
$$\mu^{k+1} = \mu^k + C^k (W^{k+1} x - s^{k+1})$$

  Update $C^{k+1} = \beta_o C^k$

  $k \leftarrow k + 1$

end

From multiple LWIR bands images we can find EPS from the ground truth data generated from original human face radiology map in the RHS assuming 2 temperature sources per pixel at 29% and 70%, and the third sensor coolant source 1% distributed spatial randomly. Then compute the Planck curve RHS projected on 3 spectral bands in the mean value 4, 9, 11 microns.



# Probabilistic Adaptive Learning Mapper

**Luca Marchese**
Syn@ptics, Italy
*Corresponding author*: luca.marchese@synaptics.org

## Abstract

This tutorial describes well known machine learning principles in a new neural network model called PALM (Probabilistic Adaptive Learning Mapper). PALM learns using a supervised algorithm inspired by the Adaptive Resonance Theory (ART), Restricted Coulomb Energy (RCE) and Learning Vector Quantization (LVQ) principles. During the recognition phase, PALM adapts itself to a «changing world», by performing an unsupervised learning. A supervised learning phase based on a poor data set can be completed during the recognition phase. The new model is based on twins neurons: in each couple there is one neuron, called "static", whose synapses are modified by supervised learning, and another neuron, called "dynamic", which inherits “knowledge” from the static one and modifies this knowledge in order to adapt to the evolving stimuli. In this tutorial, there is not a new learning theory, but just the idea that existing learning theories can be experimented in new architectural frameworks with interesting results.

## 1. Introduction

A new neural network model is described based on a Radial Basis Function architecture with adaptive and probabilistic behavior during the recognition phase. The network learns using a supervised algorithm inspired by the Adaptive Resonance Theory (ART), Restricted Coulomb Energy (RCE) and Learning Vector Quantization (LVQ) principles. During the recognition phase, the network, called PALM (Probabilistic Adaptive Learning Mapper), adapts itself to recognize new patterns using pattern probability distribution criteria. Here "mapper" should be intended as a general capability of mapping variables space in categories (mapping system). It does not have the special meaning of "visual map", as in "Kohonen maps". PALM can adapt itself to a changing world, but can also complete a supervised learning phase based on a poor data set, performing unsupervised learning during recognition.

During the supervised learning phase, two twin neurons will be created when a new prototype is required. One of these twin neurons is *static*, the other is *dynamic*, and both have the same prototype and NIF (Neuron Influence Field).

During the recognition phase, the dynamic neurons move their prototypes toward the pattern that they or their relative static twins recognized. When the prototype of a dynamic neuron goes outside its static twin's influence field, it becomes static and a dynamic copy of it is created. Any neuron, static or dynamic, is associated with an NIC (Neuron Identifications Counter) that is normalized on the category and behaves as a parameter to evaluate the reliability of recognition performed on an uncertainty region.

NIC works as it does in a PNN (Probabilistic Neural Network), but it is also incremented in the recognition phase. During the learning phase, NIC is incremented only if the recognition is correct (supervised probability update), while NIC is always incremented (unsupervised probability reinforcement) during the recognition phase.

A recognition performed on an «uncertainty region» is evaluated probabilistically, using the NIR (Neuron Identification Reliability) of the neurons (owned by different categories) that identify the input pattern. In this case, the NIF of the not-winning neurons can be reduced to exclude the “incorrectly” identified pattern.

The architecture is suitable for digital VLSI implementation. Due to its adaptive behavior during recognition, PALM can grow (dynamic neurons becomes twins of static-dynamic neurons), and a resource optimization mechanism can be investigated. The first investigated method is deleting the dynamic neurons with prototypes too close to the prototypes of their corresponding static copies.

The second method investigated and used here is deleting the dynamic or static neuron with the lowest relative NIC in the same category (if NIC < k (parameter) or searching other categories with the same criteria).

It is also possible to execute a learning phase after a recognition phase, inheriting all knowledge from the first learning phase and the adaptive behavior of the recognition phase.

A test on the “circle in the square” problem demonstrates how PALM, trained with few patterns on the boundary regions, can build a complete knowledge base of the areas related to the two shapes during recognition.

## 2. Architecture

PALM contains three fully interconnected layers, as shown in Fig.1.

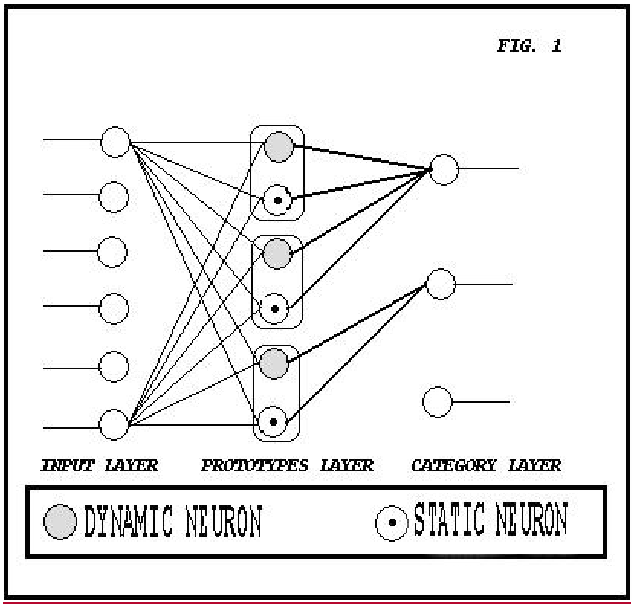



Fig.1 PALM architecture

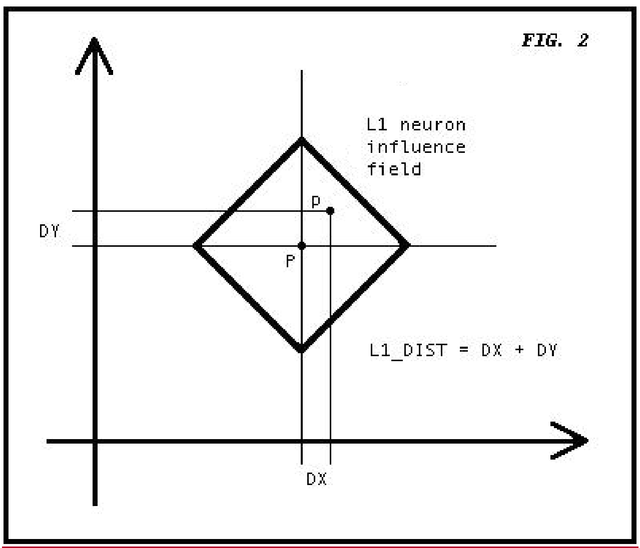



Fig.2 L1 distance

The input layer receives the input pattern and does not have an active function.

The middle layer is the prototype layer: each neuron is associated with a prototype. The output layer is the category layer: each neuron is associated with a category. The PALM size grows during the learning and recognition phases, due to an adaptive behavior based on the probability distribution of the recognized patterns. PALM works as an adaptive classifier based on the patterns' L1 distance calculation (Fig.2).

## 3. Initialization

The PALM neural network must be initialized with the following parameters:

MIF (Minimum Influence Field)
MAF (Maximum Influence Field)
Epsilon (prototype approaching coefficient)
STME (Short Term Memory Edge)
DeletionMode (flag 1/0)

MIF is the minimum value that can assume the NIF (Neuron Influence Field). This parameter works when a new prototype is created or when a category mismatch is verified and the NIF of the wrong identifier neuron must be reduced.

MAF is the maximum value that can assume the NIF (Neuron Influence Field). This parameter works only when a new prototype is created. The valid condition is 0 < MIF < MAF < 16000 in this simulation.

Epsilon is the value that tunes the speed of a prototype to approach identified patterns during the learning (supervised) and recognition phases. Epsilon can assume a value between 0.1 and 1.0.

STME is a threshold value between 0 and 100, used to decide whether a neuron is not useful. When a network-full event is verified and a new prototype must be created, a not-useful prototype to delete is sought with the condition NIC[neuron] < STME. If the DeletionMode flag value is 1(IN_CATEGORY), the condition NIC[neuron] < STME is sought in the cluster of neurons matching the category of the neuron to be created. If the DeletionMode flag value is 0(ALL_CATEGORY), the condition NIC[neuron] < STME is sought in the entire network.

## 4. Learning algorithm

Table 1 explains the supervised learning algorithm. When a new pattern is presented to PALM with an associated category, the following situations can be found:

(1) The pattern falls in the NIF (Neuron Influence Field) of one or more existing neurons. If the category associated with the pattern is the same as the category associated with the winning neuron, the prototype of the winning neuron approaches the pattern with the rule:

$$\forall(0 \le k < n-1) P_k = P_k \times (1-\varepsilon) + p_k \times \varepsilon \qquad (1)$$

Or, it may be presented in meta-language as

```
FOR k=0 TO pattern dimension – 1
{
 P[k] = (P[k]*(1-Epsilon) + p[k]*Epsilon) }
```

Here, P[k] = $P_k$ = the $k$-th element of prototype, p[k] = $p_k$ = the $k$-th element of pattern, and Epsilon = $\varepsilon$ = an appropriate constant (0.0, 1.0).

If the approach causes the NIF to include an existing prototype, the approach is aborted.

When a pattern is recognized, the NIC (Neuron Identification Counter) associated with the identifier neuron is incremented. When the NIC reaches the maximum value 100, it is divided by 2, along with the NICs of all neurons

Table 1. Learning Algorithm

```
PROCEDURE LEARNING (VECTOR, CAT)
 FOR N = 1 TO N = LAST_ACTIVE_NEURON                       #for any committed neuron
  L1_D = L1_DISTANCE(VECTOR, PROTOTYPE[N])                 #L1 distance calculation
  IF(L1_D < NIF[N])                                        #neuron is firing
   IF(CAT[N] = CAT)                                        #category is correct
    ID_FLAG = TRUE                                         #at least one correct id
    NIC[N] = NIC[N] + 1                                    #increasing NIC
    IF(NIC[N] = 100)                                       #maximum value of NIC
     FOR M = 1 TO M = LAST_ACTIVE_NEURON                   #normalizing NIC…
      IF (CAT[N] = CAT)                                    #…on neurons with same cat
       NIC[N] = NIC[N] / 2
      ENDIF
     ENDFOR
    ENDIF
    APPROACH(N, VECTOR)                                    #approaching prototype
   ELSEIF                                                  #category is wrong…
    NIF[N] = L1_D                                          #…reducing NIF
   ENDIF
  ENDIF
  IF(L1_D < MIN_L1_D)                                      #searching min distance
   MIN_L1_D = L1_D
  ENDIF
 ENDFOR
 IF(ID_FLAG = FALSE)                                       #no neurons fired correctly
  IF((POS = TEST_RESOURCES()) > 0)                         #test memory resources
   CREATE_NEW_PROTOTYPE(POS, CAT, VECTOR,MAX_NIF, BOTH)    #new prototype creation
  ENDIF
 ENDIF
END PROCEDURE
```

associated with the same category (NIC normalization inside a category).

If, at the same time, the pattern falls in the influence field of one or more neurons associated with a different category, the NIF of these neurons is reduced (Fig.3) to exclude the pattern, accordingly with MIF:

$$\eta_j = (\sum_{k=0}^{n-1}(P_{jk} - p_k)) \vee MIF \qquad (2)$$

Or, in meta-language as:

New_NIF[j] = max( L1_dist(P[j],p), MIF )

where P = prototype, p = pattern, j = the index of the identifier neuron, k = the index of the vector component, NIF = $\eta$ = L1 influence field and

$$\vee = fuzzy - OR, (i.e: 0.2 \vee 0.5 = 0.5).$$

(2) The pattern is not identified, and one new neuron must be created (Fig.4).

The new neuron has the pattern as a prototype and an NIF equal to the L1 distance between the new prototype and the nearest existing prototype:

$$\eta_j = (\sum_{k=0}^{n-1}(P_{jk} - P\_n_k)) \wedge MAF \qquad (3)$$

Or, in meta-language as:

NIF[new_neuron]
=min(L1_dist(P[new_neuron],P_n), MAF )

where P = prototype, and P_n = nearest existing prototype, j = the index of the new neuron, k = the index of the vector component, NIF = $\eta$ = L1 influence field and

$$\wedge = fuzzy - AND, (i.e: 0.2 \wedge 0.5 = 0.2).$$

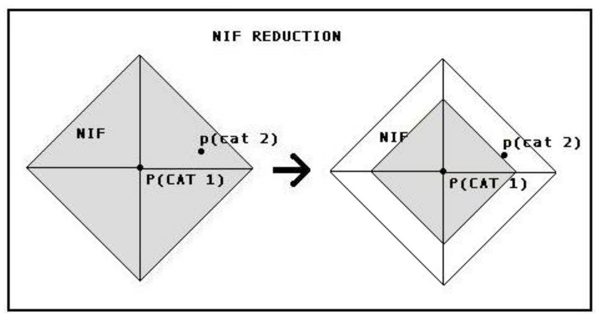



Fig.3 NIF reduction

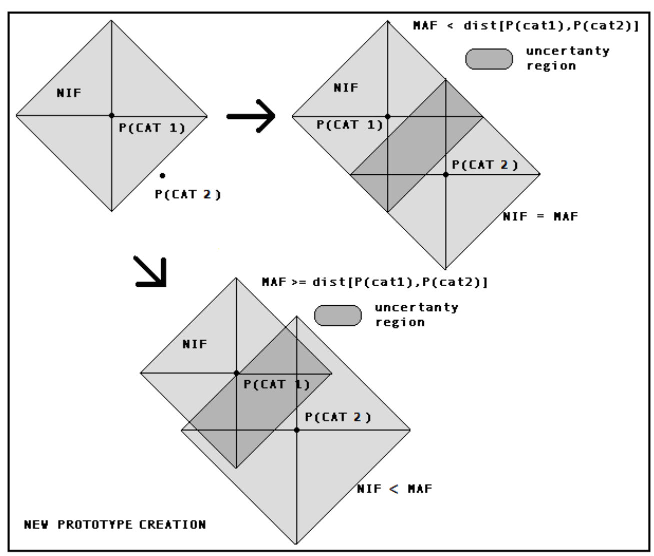



Fig.4 New prototype creation

(3) Only neurons with different categories identify the pattern. The NIF of all identifying neurons must be reduced (Fig.3), and a new prototype must be created (Fig.4).

## 5. Recognition Phase

Table 2 explains the recognition phase behavior. Any pattern input into the neural network is evaluated on its L1 distance from all existing prototypes. All neurons with an L1 distance from the pattern less than the relative NIF are identifiers. Three situations are possible:

(1) Only neurons associated with the same category identify the pattern.

In this case, the identification is performed (ID), and the recognition confidence is set to its maximum value.

The prototype of the dynamic copy of the winning neuron (least L1 distance) also approaches the input pattern in the recognition phase:

$$\forall(0 \le k < n-1) D_k = D_k \times (1-\varepsilon) + p_k \times \varepsilon \quad (4)$$

Or, in meta-language as:

```
FOR k=0 TO pattern dimension – 1
 {
  D[k] = (D[k]*(1-Epsilon) + p[k]*Epsilon)
 }
```

where D[k] = the *k*-th element of dynamic prototype, and p[k] = the *k*-th element of pattern (Fig.5).

If the approach causes the NIF to include an existing prototype, the approach is aborted. This operation probabilistically adapts the dynamic prototype. This adaptive behavior moves the dynamic prototype's center to the maximum probability density region inside the NIF of the static twin.

A new identification region is automatically created because the dynamic neuron has its own independent recognition capability (Fig.5). The dynamic neuron can be a winner when recognizing of a new input pattern and its prototype again moves toward the pattern. When the prototype center of a dynamic neuron exceeds the NIF of its static twin (Fig.6), two actions are possible:

a) Fusion: only if the condition to perform the operation exists, described in Fig.7. The condition is that the new NIF must not include an existing prototype.

b) Mutation: the static neuron loses the dynamic twin that becomes a static neuron with an associated dynamic twin (Fig.8). When a mutation event happens, the neural network grows from a single neuron. Considering a software or hardware PALM implementation, a mechanism to manage the full use resource (full network event) must be investigated. For PALM, I have provided for the deletion of a not-useful prototype. This operation follows the rule:

$$\because ((\forall(k \ne j) NIC_j < NIC_k) \,\&\, (NIC_j < STME)) \therefore n_j = \varnothing \quad (5)$$

Or, in meta-language as:

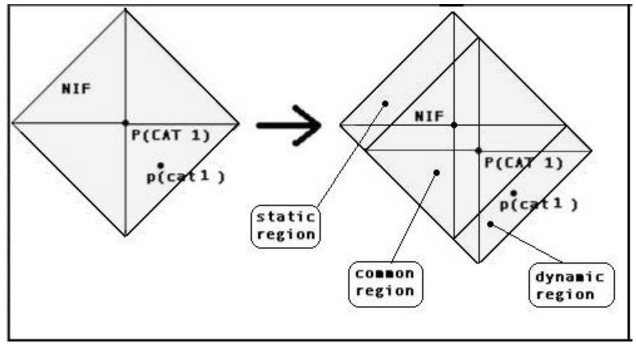



Fig.5 Dynamic prototype approaching

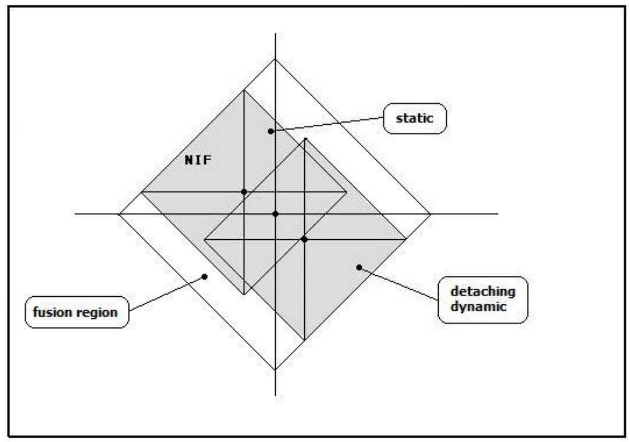



Fig.6 NIF exceed condition

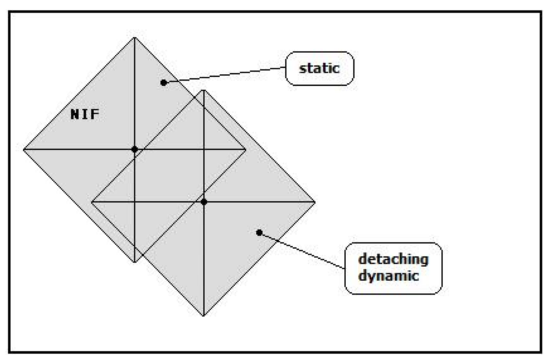



Fig.7 Fusion

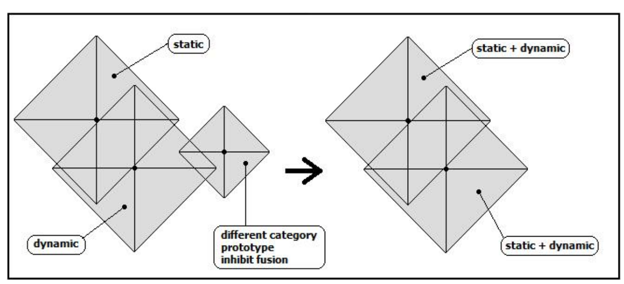



Fig.8 Mutation

```
if(min(NIC[0],NIC[1],…,…,NIC[k-1],NIC[k])< STME)
then DELETE neuron
```

Table 2. Recognition Algorithm

```
PROCEDURE RECOGNITION (VECTOR)
 IDENTIFIERS_NIR = 0                              #initialization of variables
 IDENTIFIERS_COUNTER = 0
 CATEGORIES_COUNTER = 0
 FOR N = 1 TO MAX_CAT_NUMBER
  IDENTIFIERS_NIR_CAT[N] = 0
 ENDFOR

 FOR N = 1  TO N = LAST_ACTIVE_NEURON             #for any active neuron…
  L1_D = L1_DISTANCE(VECTOR, PROTOTYPE[N])        #L1 distance with vector is calculated
  IF (L1_D < NIF[N])                              #if the neuron fires…
   IDENTIFIERS_COUNTER = IDENTIFIERS_COUNTER + 1  #a counter of firing neurons is incremented
   IF(IDENTIFIERS_NIR_CAT[CAT[N]] = 0)            #if this is the first firing neuron for category
    CAT_LIST[CATEGORIES_COUNTER] = CAT[N]         #the new category is added to the list
    CATEGORIES_COUNTER = CATEGORIES_COUNTER + 1   #the number of categories is incremented
   ENDIF
   NIR[N] = NIC[N] / ((L1_D + 1) * (NIF[N] + 1))  #NIR of the firing neuron is calculated
   IDENTIFIERS_NIR_CAT[CAT[N]] = IDENTIFIERS_NIR_CAT[CAT[N]] + NIR[N]
                                                  #the NIR of the neuron is added to the cumulative
                                                  #NIR of category
   IDENTIFIERS_NIR = IDENTIFIERS_NIR + NIR[N]     #the NIR of the neuron is added to the global
                                                  #cumulative NIR

   IF(STATUS[N] = DYNAMIC)                        #if the neuron is dynamic…
    APPROACH(N, VECTOR)                           #it's prototype is approached to vector
    IF(L1_DISTANCE(PROTOTYPE[N], PROTOTYPE[TWIN[N]]) > NIF[TWIN[N]])
                                                  #if it's prototype go outside the NIF of static
                                                  #twin…

      IF (PERFORM_FUSION(N) = FALSE)              #if the fusion doesn't include other prototypes is
                                                  #performed and and the function return TRUE
      IF((POS = TEST_RESOURCES()) > 0)            #fusion includes other prototypes and was aborted
       STATUS[N] = STATIC                         #N becomes static
       TWIN[TWIN[N]] = NO_TWIN                    #TWIN[N] has lost the twin
       CREATE_NEW_PROTOTYPE(POS, CAT[N], PROTOTYPE[N], NIF[N], DYNAMIC)
      ENDIF
     ENDIF
    ENDIF

   ELSEIF(STATUS[N] = STATIC)     #the neuron is static
     IF(N HAS A DYNAMIC TWIN)     #the static neuron could have already lost it's dynamic twin…
     APPROACH(TWIN[N], VECTOR)    #the prototype of the dynamic twin is approached to vector

     IF(L1_DISTANCE(PROTOTYPE[N], PROTOTYPE[TWIN[N]]) > NIF[N])
                                  #if it's prototype go outside the NIF of it's static twin…

       IF (PERFORM_FUSION(N) = FALSE)      #fusion includes other prototypes and then has been aborted
       IF((POS = TEST_RESOURCES()) > 0)    #TWIN[N] becomes static and an identical dynamic twin is
                                           # created
        STATUS[TWIN[N]] = STATIC  #TWIN[N] becomes static
        TWIN[N] = NO_TWIN         #N has lost the twin
        CREATE_NEW_PROTOTYPE(POS, CAT[N], PROTOTYPE[TWIN[N]], NIF[N], DYNAMIC)
       ENDIF
      ENDIF
     ENDIF
    ENDIF
   ENDIF
  ENDIF
 ENDFOR                                    #the cycle on committed neurons is ended

 FOR INDEX = 1 TO CATEGORIES_COUNTER       #a cycle on categories of firing neurons starts
  CAT = CAT_LIST[INDEX]                    #get category
  CONFIDENCE[CAT] = IDENTIFIERS_NIR_CAT[CAT] * 100 / IDENTIFIERS_NIR
                                           #confidence of recognition for any category is computed
  IF (CONFIDENCE[CAT] > MAX_CONFIDENCE)
   CONFIDENCE[CAT] = MAX_CONFIDENCE
  ENDIF
 ENDFOR
END PROCEDURE
```

where NIC = Neuron Identification Counter, NIC[0] = Neuron Identification Counter of neurons indexed 0 in the cluster, STME = Short Term Memory Edge, $\eta_j$ = neuron indexed by $j$, and & = logical AND.

The cluster may contain neurons associated with the same category as the new prototype or all neurons in the network, depending on the "DeletionMode" initialization parameter settings. STME (Short Term Memory Edge) is a threshold indicating the maximum NIC acceptable to locate the not useful neuron in the LTM (Long Term Memory) context.

All neurons identifying the pattern have their NIC incremented. If one reaches the maximum value 100, the NIC values of all neurons associated with the category are divided by 2 (NIC normalization inside a category).

(2) Neurons associated with different categories are pattern identifiers.

There is uncertainty in the identification (UNC), and the confidence values for any category present in the identifiers list is calculated as explained in the following steps:

$$NIR_j = NIC_j \div ((NIF_j + 1) \times (\sum_{k=0}^{m} (P_{jk} - p_k) + 1)) \qquad (6)$$

Or, in meta-language as:

NIR[neuron]

= NIC[neuron]/(( NIF[neuron]+1)*(L1_DISTANCE+1))

where NIR = Neuron Identification Reliability of the neuron, NIC = Neuron Identification Counter of the neuron, NIF = Neuron influence Field of the neuron, L1_DISTANCE = L1 distance between input vector and prototype, $P_{jk}$ = the $k$-th component of prototype indexed by $j$, $p_k$ = the $k$-th component of input vector, and $m$ = input vector dimension.

$$NIRS_{cat} = \sum_{k=0}^{m} NIR_j \ , \quad j \ni \sum_{k=0}^{m} Pjk \ - p_k < NIF_j \qquad (7)$$

Or, in meta-language as:

NIRS[cat] = NIR[A1] + NIR[A2] + … + NIR[Am]

NIRS = NIR[B1] + NIR[B2] + … + NIR[Bm]

where NIRS = Neuron Identification Reliability Sum, cat = category of interest, A1, A2, …, Am = firing neurons owned by category cat, and B1, B2, …, Bm = all firing neurons.

$$C_{cat} = \max\_conf \wedge (NIRS_{cat} \times 100 \div NIRS) \qquad (8)$$

Or, in meta-language as:

C[cat]=min(MAX_CONFIDENCE,(NIRS[cat]*100/IRS))

Where C[cat] = recognition confidence for category cat, MAX_CONFIDENCE = maximum value of admitted confidence, and $\wedge = fuzzy\_AND$.

(3) No neurons are pattern identifiers. No action is performed.

## 6. The Circle in the Square Test

Recognizing points inside and outside the circle is useful for visually understanding the behavior of a neural network in a pattern recognition task. The program, written in Visual Basic, calls the PALM functions in PALM.dll, written in C. The program can show the neural network prototypes with their NIFs or the ensemble of input points with a different color for any category (circle, square, not identified) during the recognition phase. In the grey-scale pictures, different colors can be distinguished by different gray levels.

Fig.9(a) shows the neural network after 10000 learning points on the global surface. Fig.9(b) shows the neural network after 2000 learning points on the circumference and perimeter of the circle and square.

The recognition phase steps verify the probabilistic adaptive behavior trends for building knowledge to expand a category region in the direction, where no other categories exist, that has the highest probability pattern density. Figs.9(c) to 10f show the growth of the neural network knowledge based on unsupervised learning. After 600,000 input patterns, the neural network shows knowledge similar to that reached after full surface learning (Fig.9(a)) and seems to be more optimized (Fig.9(g) vs. Fig.9(h)). Considering that full surface learning has been executed using the setting MAF=30, the contour learning has been executed using the setting MAF=10 to obtain contour precision.

The deceit of this test seems to be the lack of points outside the square, thus inhibiting the compensation for prototype movements around the perimeter. Pattern recognition problems in the real world are not so far: category spaces are, often, contained in larger spaces, where the probability of finding patterns tends toward zero. The clearest example is recognizing flying objects: jets and helicopters can never assume positions with similar shapes. Another common rule is that a correct feature extraction technique should also create a situation where category spaces are as distinct as possible.

## 7. Network Evolution during Recognition

It is interesting to view how the neural network evolves, both after learning and during the recognition phase. The above prototype images of the steps of the recognition process allow this, but other network views can better show the structural transformation and growth. Fig.10(a) shows the links between static neurons and dynamic twins at the end of recognition step 4. Only the links between neurons that are not perfectly superimposed (identical prototypes) are shown because this is a view in the prototypes space.

Figs.10(b) to 10(g) show the structural evolution of the network (these steps do not exactly coincide with the above steps). In this image sequence, blue spots are neurons (static or dynamic) with a twin (connected with a black line), and red spots are neurons that have lost their twin. The sequence shows that any new dynamic neuron is created when a static

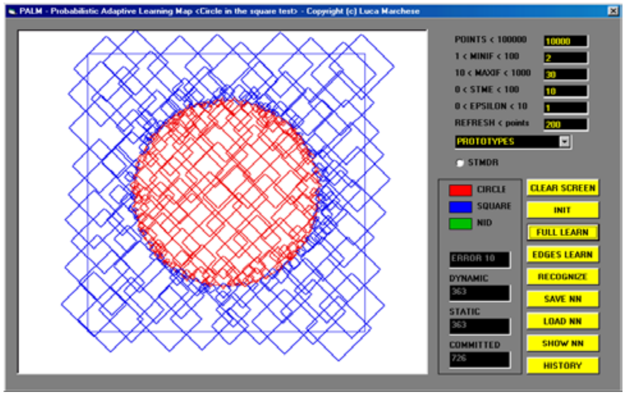


Fig.9(a) Full learning

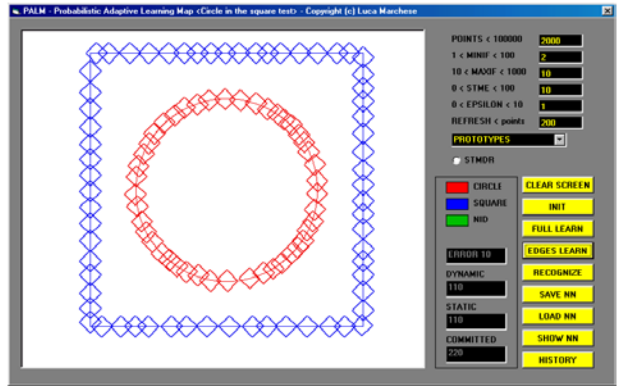


Fig.9(b) Contours learning

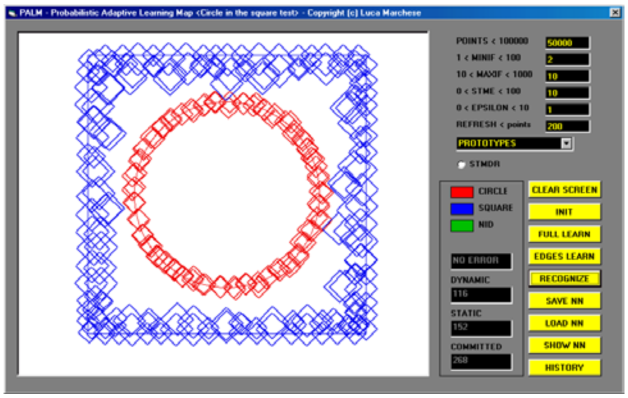


Fig.9(c) Recognition step 1 – prototype view

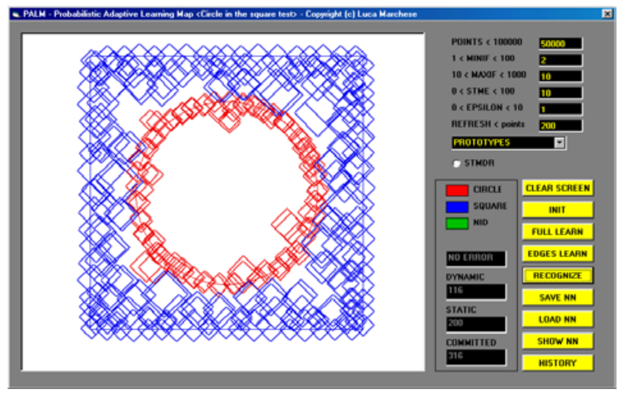


Fig.9(d) Recognition step 2 – prototype view

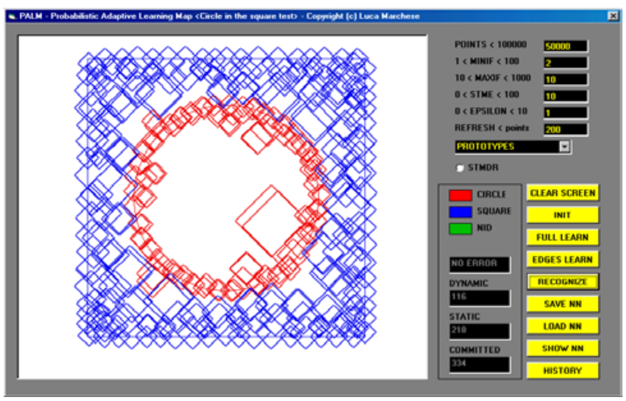


Fig.9(e) Recognition step 3 – prototype view

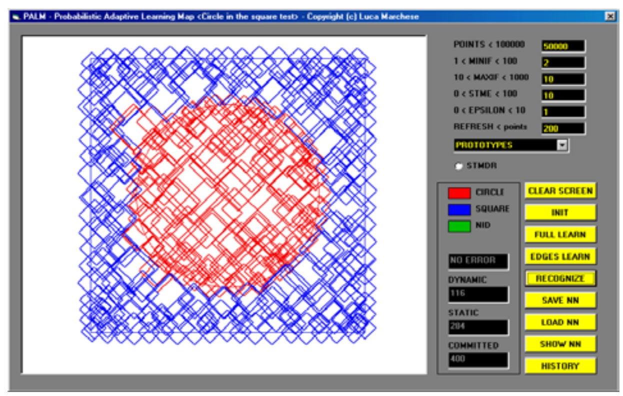


Fig.9(f) Recognition step 4 – prototype view

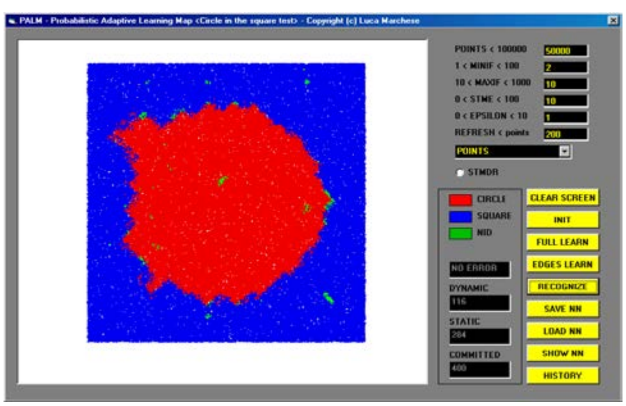


Fig.9(g) Recognition step 4 – recognition view

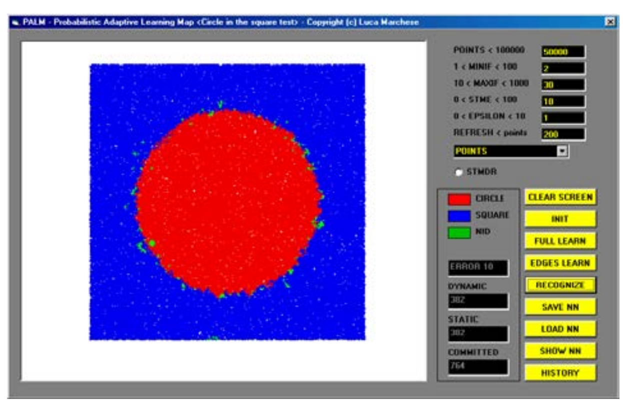


Fig.9(h) Recognition after full supervised learning – recognition view

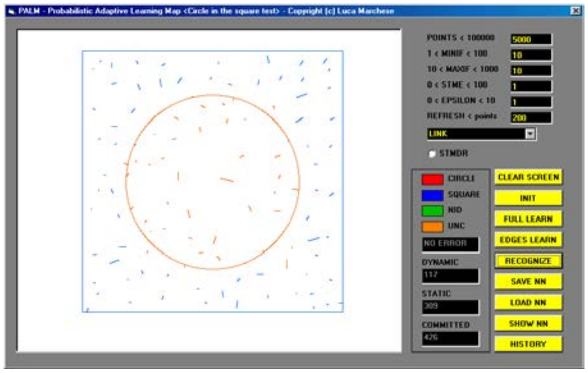


Fig.10(a) View of the static-dynamic links in category space

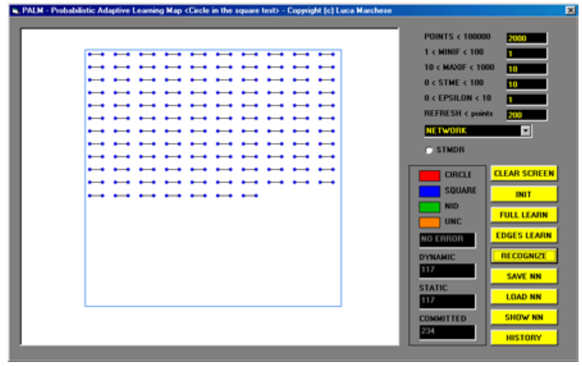


Fig.10(b) Structural view of the network after supervised learning: blue spots are neurons and black lines are links between static and dynamic neuron

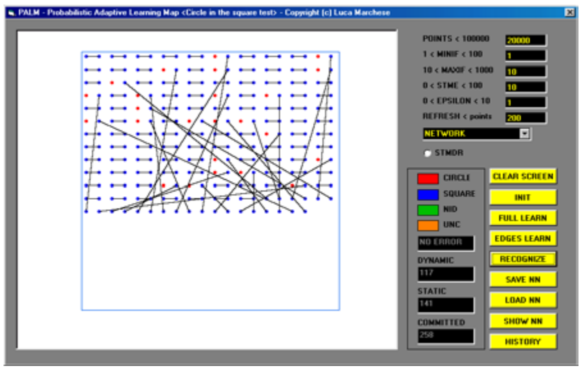


Fig.10(c) Structural view of the network after the first recognition step: red spots are static neurons that have lost their dynamic twins

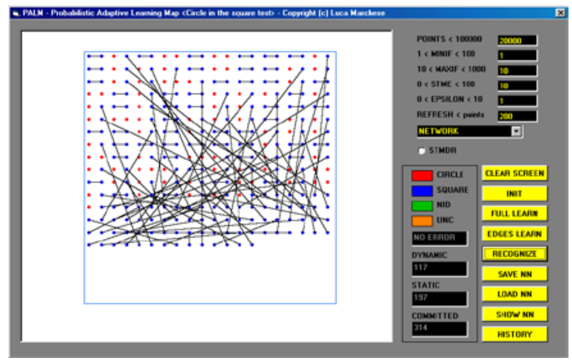


Fig.10(d) Structural view of the network after the second recognition step

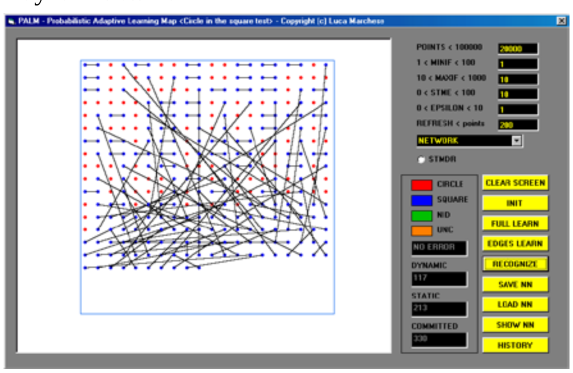


Fig.10(e) Structural view of the network after the third recognition step

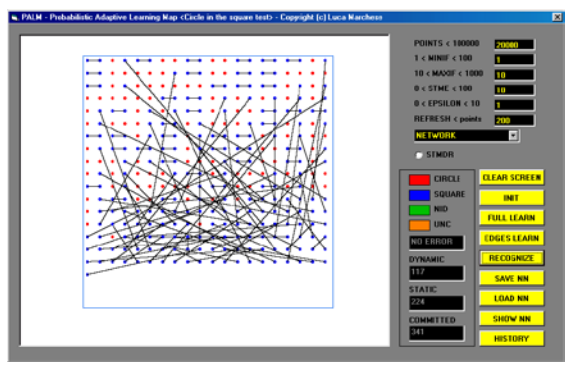


Fig.10(f) Structural view of the network after the fourth recognition step

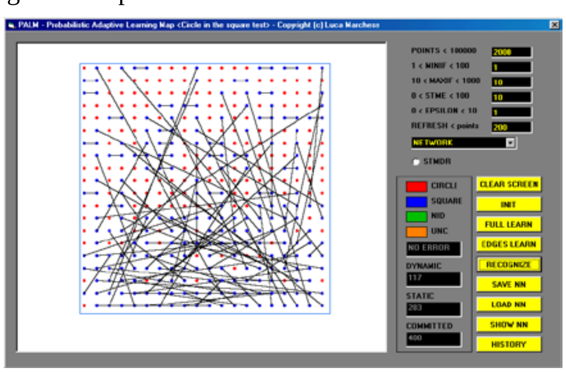


Fig.10(g) Structural view of the network after the fifth recognition step

neuron losses its dynamic twin (i.e., a blue spot becomes red). It becomes static, and a black line connects it to the new neuron.

## 8. Optimization for Implementation

In the above test, the neural network trained on the unsupervised algorithm during the recognition phase seems to be more optimized than the neural network fully trained on a supervised algorithm. This was true, with the following considerations:

The number of committed neurons in full supervised training are thus:

dynamic = 363, static = 363.

The number of committed neurons in edge-supervised training followed by unsupervised learning in the recognition phase are thus:

dynamic = 116, static = 284.

In the first case, 363 dynamic neurons are not useful because their prototype is perfectly superimposed with its static twin. If we consider a network without a dynamic copy, the number of committed neurons is 363, which is less than 116 + 284 = 400. In the first case, the network has been trained with MAF = 30, which enables a large (< 30) NIF neuron commitment, while, in the second case, MAF = 10 makes that committed neuron number NIF < 10.

In the second case, the choice was needed for further precision in the circumference/perimeter definitions, which constitute the only supervised data.

Following these considerations, we can say that unsupervised learning during the recognition phase grows with a self-optimizing mechanism (fusion when possible); in order to say that PALM is optimized, we must have resources committed only by useful neurons. Not-useful neurons are here the dynamic twins with exactly the same prototype as the static ones. The implementation must then consider dynamic neurons as "virtual units" that exist (resources commitment) only when their prototype is an amount "Kv" different from their static twins. The value of Kv is strictly related to the value of Epsilon.

Kv should be sufficiently small to warrant network flexibility during the recognition phase. Kv introduces a region around the static prototype in which a recognized pattern does not introduce a movement of the dynamic twin prototype. This behavior reduces the number of dynamic neurons enabling their creation only when a recognized pattern is sufficiently far (but obviously inside NIF) from the prototype of the winning static neuron. Following this reasoning, the Kv could be related to NIF, i.e., a percentage of it.

## 9. Image Sequence Test

This test is performed to verify the network's capability of adapting and completing its knowledge of a pattern whose shape is dynamically and continuously changing. As I am only interested in performing a test on adaptive behavior, I do not consider many aspects of real pattern recognition or pattern tracking tasks, simplifying the problem and focusing on my main target. The ROS (Region of Scanning) and ROI (Region of Interest) sizes have been fixed, and the former has a position independent of the last recognition. Pattern scale invariance is not considered here, as it is outside the scope of this experiment. The feature extracted is simply the composite profile (after thresholding) of the image in the ROI (Fig.11), and scanning is performed using a 4 pixel step.

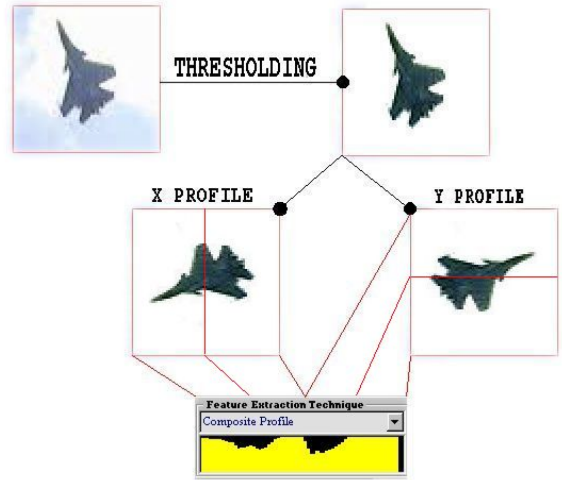
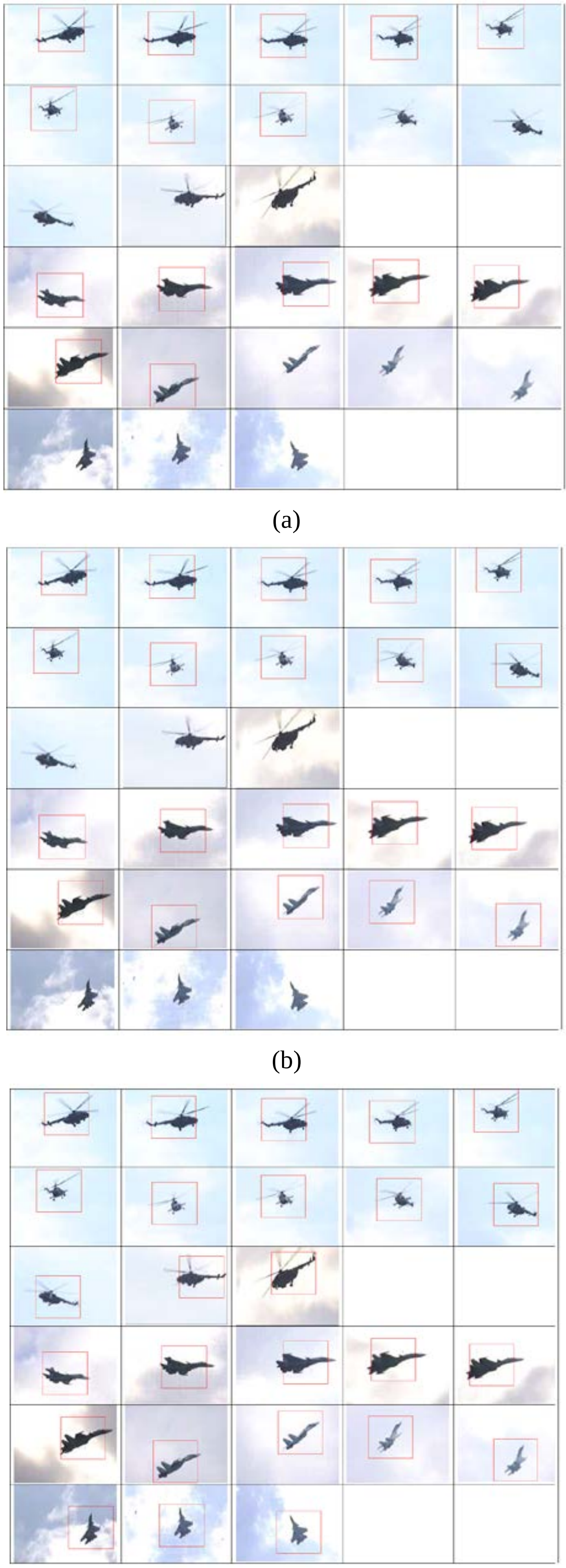



Fig.11 Thresholding and composite profiles

I have used two image sequences, showing a helicopter or a jet. Category 1 is associated with the helicopter, and category 2 is associated with the jet. The supervised learning uses only the first image of each sequence. After learning, both complete image sequences are input into the network in the recognition phase.

While receiving the input image sequences, the neural network adapts the dynamic prototypes to follow the continuously changing target shape. Figs.12(a)-(c) show the results after 100, 500, 1000 inputs, respectively, of each sequence. The target identification is extended during recognition to the most frequent shapes assumed. I have changed the number of frames in any position, simulating an increasing frame sampling rate from the first frame to the last. The images shown here are steps of the complete frame sequences that could not be fully represented. The experimental hypothesis is that the frame sampling rate increases from the start to the end of the "movie". When dynamic prototypes become static, new positions are learned in an unsupervised manner. The rectangle around the shape indicates pattern identification and coincides with the ROI. The two patterns are identified correctly in the corresponding categories if the new positions have been learned in an unsupervised manner without associating the category to the pattern. The experiment aims to show a wider applicability of PALM’s adaptive behavior: a specific pattern category present in the environment in multiple instances may gradually change its features in a growing percentage of its instances. Consider a smart traffic monitoring camera that learns vehicle shapes that change gradually over time due to new styles, sizes and volumes. Similarly, remember the facial recognition application: PALM could learn the features of the face of a 10-year-old child and adapt the synaptic connections when he is recognized. A frequently performed recognition allows future correct recognition even after the child becomes a man, without requiring additional neural network training. If the neural network uses all memory resources available during this adaptation process, a "network-full" event starts



(a)

(b)

(c)

Fig.12 Recognition after (a) 100, (b) 500, and (c) 1000 presentations of each sequence. Red box denotes detected ROI, while images without red box are not recognized.

a procedure of deleting not-useful neurons. In facial recognition, if the "network-full" event happens when the "child" is 18 years old, the prototypes related to age 10-15 should have a low NIC, due to the normalization process performed when a prototype reaches the maximum NIC value.

## 10. Adaptive and Probabilistic Behavior Work Together

The synaptic adaptive behavior and neuron probabilistic data seem to work with high synergy. Looking at PALM's behavior, one can easily understand the contribution of adaptive behavior to improving the reliability and meaningfulness of the neuron probabilistic data. Conversely, neuron probabilistic data are basic instruments for managing network resources, supplying criteria to evaluate prototype usefulness.

## 11. Prototype Tuning using Pattern Probability Distribution: Adaptive Behavior Helps Probabilistic Behavior

The adaptive behavior of neurons strongly influences using probabilistic recognition confidence. In a static environment, an adaptive behavior should not be required but the adaptation can work for tuning prototypes on the probability distribution of patterns falling into their influence fields. During the recognition phase, this characteristic can optimize poor data training, increasing the global probability of pattern identification in the field. Furthermore, as a secondary effect of this behavior, the probabilistic confidence value reliability, associated with any prototype, grows during the recognition task. This growing effect occurs not only because the number of input patterns grows but also because any prototype is tuned to optimize its position in the pattern probability distribution (Fig.13).

## 12. Prototype Usefulness and Short Term Memory Evaluation: Probabilistic Behavior Helps Adaptive Behavior

The adaptation mechanism described above implicitly produces a neural network growth that must be managed to enable real PALM implementation with limited resources (hardware or software). The method used here is deleting prototypes considered not useful or, more exactly, rarely used. This is a large matter where answers should be evaluated in any specific context. The relative number of patterns that a prototype identifies during its life is, in PALM, the value that describes the usefulness of the prototype itself. This value should be considered a probabilistic value relative to the owner's class space of the prototype. The normalization process performed on the NIC (Neuron Identification Counter) inside the class controls the relativity. The ensemble of NIC values inside a class

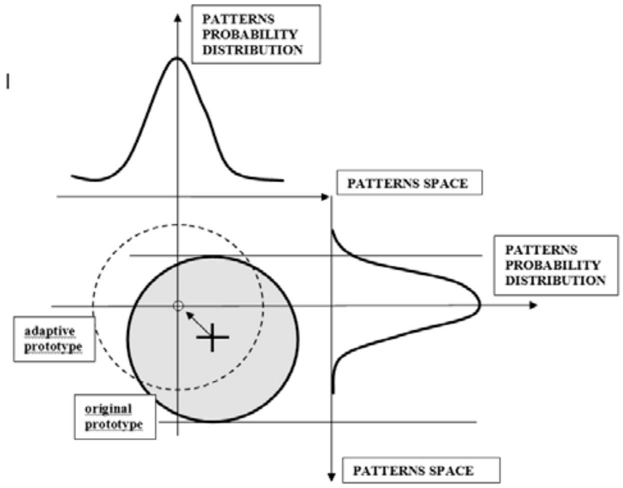



Fig.13 Two-dimensional view of adaptive prototype behavior in pattern probability distribution. The adaptive prototype shown can optimize itself using pattern probability distribution.

contains all data on the pattern probability distribution inside the class itself. In this work, the threshold that enables deleting neurons is called STME (Short Term Memory Edge) because this concept better describes the role of neurons with lower NIC values inside the class. It is not correct to speak about the "usefulness" of neurons related to the probability of patterns represented by an NIC. It is more correct to say that a neuron with a low NIC value is used "rarely", and "to delete" in this context means "to forget". This is a restricted STM vision that should also be correlated with other parameters: information can be more or less strongly learned depending on the context in which it has been received. This could be a future study in PALM applications.

## 13. Conclusions

This paper has presented a new neural network paradigm based on supervised learning and unsupervised probabilistic adaptive behavior during recognition. The tests performed have shown interesting properties. PALM shows the capability to interpolate the category space in an unsupervised manner when training data are poor and not uniformly distributed. During recognition, PALM can adapt the prototypes representing a category to be closest to the highest probability pattern density in that category space. This adaptive behavior allows PALM to adapt itself to gradually changing patterns. Introducing a probabilistic parameter inside neurons allows PALM to remove not-useful neurons when new neurons must be created. This architecture is based on simple algorithms and is suitable for fast hardware implementation on a Field-Programmable Gate Array. Listings are written in a meta-language to be clear. Software (PALM library, “circle in the square test” and source code) can be requested to the author (luca.marchese@synaptics.org)

## References


[1] Anagnostopulos, G,C., & Georgiopoulos,M.(2001). Ellipsoid ART and ARTMAP for incremental unsupervised and supervised learning. Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks(Vol.2,pp.1221-1226).Washington, DC:IEEE,inns,enns.

[2] Bezdek, J. C.,Reichherzer, T., Lim, G. S., & Attikiouzel, Y. (1998). Multiple-prototype classifier design. IEEE Transactions on Systems, Man, and cybernetics, Part C: Applications and Reviews, 28(1),67-79.

[3] Carpenter, G. A.(1997). Distributed learning, recognition, and prediction by ART and ARTMAP neural networks. Neural Networks, 10(8), 1473-1494

[4] Grossberg, S. (1976). Adaptive pattern recognition and universal encoding II: Feedback, expectation, olfaction, and illusions. Biological Cybernetics,23, 187-202.

[5] Carpenter, G. A., Milenova, B. L.,& Noeske, B, W, (1998). Distributed ARTMAP: A neural network for distributed supervised learning. Neural Networks, 11,793-813.

[6] Donald F. Specht, Probabilistic neural networks for classification, mapping, or associative memory. Proc.IEEE Second Int. Conf. On Neural Networks, San Diego, CA, July 1988, vol I, pp.525-532.

[7] Donald F. Specht, Probabilistic neural networks. Neural Networks, 3:109–118, 1990.

[8] Pietro Burrascano. Learning vector quantization for the probabilistic neural network. IEEE Transactions on Neural Networks, 2(4):458–461, July 1991.

[9] Intel Corporation. Ni1000 Beta Release 2.3 documentation, 1994.

[10] E. Parzen, On estimation of a probability density function and mode. Annals of Mathematical Statistics, 33:1065–1076, 1962.

[11] T. Cacoullos. Estimation of a multivariate density. Annals of the Institute of Statistical Mathematics (Tokyo), 18(2):179–189, 1966.

[12] M.Holler, C.Park, J.Diamond, U.Santoni, S.C.The, M.Glier and C.L.Scofield, A high performance adaptive classifier using radial basis functions. Proceedings of Government Microcircuit Applications Conference, 1992, pp.261-264.

[13] Michael R. Berthold and Jay Diamond. Boosting the performance of RBF networks with Dynamic Decay Adjustment. In G.Tesauro, D.S. Touretzky and T.K. Leen editor, Advances in Neural Information Processing Systems 7, pages 521-528 Cambridge MA. The MIT Press.

[14] Douglas L. Reilly, Leon N. Cooper and Charles Elbaum. A neural model for category learning. In Biological Cybernetics 45, pages 35-41, 1982.

[15] Michael H. Hudak, RCE classiers: Theory and practice. In Cybernetics and Systems, volume 23, pages 484-515. Hemisphere Publishing Corporation 1992.

[16] Constantinos Constantinopoulos, Aristidis Likas. Semi-supervised and activelearning with the probabilistic RBF. Journal Neurocomputing archive, Volume 71 Issue 13-15 , August, 2008. Elsevier Science Publishers

## EANN 2012

By Chrisina Jayne, ab1527@coventry.ac.uk

The 13th EANN 2012 conference was held at the London Campus of Coventry University, UK during September, 2012. The primary sponsor for the conference was the International Neural Network Society (INNS).

The conference attracted delegates from twenty-four countries across the world: Russia, USA, South Africa, Germany, Italy, New Zeeland, UK, Greece, Switzerland, Spain, Brazil, India, Ukraine, France, Poland, Turkey, Chile, Israel, China, Cyprus, Taiwan, Portugal, Belgium, and Finland. The papers presented at the conference included variety of applications of neural networks and other computational intelligence approaches to challenging problems relevant to society and the economy. These included areas such as: intelligent transport, environmental engineering, computer security, civil engineering, financial forecasting, virtual learning environments, language interpretation, bioinformatics and general engineering.

Two very lively workshops took place as part of the conference: Workshop on Applying Computational Intelligence Techniques in Financial Time Series Forecasting and Trading (ACIFF) and Workshop on the Computational Intelligence Applications in Bioinformatics (CIAB).

There were amazing inspirational keynote lectures presented by

- Professor Nikola Kasabov, Director and Founder, Knowledge Engineering and Discovery Research Institute (KEDRI), Chair of Knowledge Engineering, Auckland University of Technology, Institute for Neuroinformatics - ETH and University of Zurich
- Dr. Danil Prokhorov, President-Elect of INNS, Toyota Research Institute NA, Ann Arbor, Michigan
- Professor Kevin Warwick, University of Reading, England and Fellow of The Institution of Engineering and Technology
- Professor Richard J. Duro, Grupo Integrado de Ingeniera Escuela Politecnica Superior, Universidade da Coruña.

The tutorial on "Fuzzy Networks with Modular Rule Bases" was presented by Dr Alexander Gegov from the University of Portsmouth, UK.

The International Neural network Society kindly provided two student awards for the EANN 2012 best student papers. The awards were: one for the best student paper at $250 including two-year student INNS membership and another award at $150 including the same for the runner-up.

The best student paper award was given to: Sakyasingha Dasgupta, PhD Student at Georg-August-Universität Göttingen for his paper Information Theoretic Self-Organised Adaptation in Reservoirs for Temporal Memory Tasks".

The runner up student award went to Po-Chuan Cho, PhD student at National Taipei University of Technology, Taiwan for the paper "A Double Layer Dementia Diagnosis System Using Machine Learning Techniques".

On behalf of the Organizing Committee I would like to thank all that contributed to the success of the EANN 2012 conference.

Dr Chrisina Jayne, Chair of EANN SIG and Organizing Chair for EANN 2012

## Call for Papers

# Modeling and Aiding Intuitions in Organizational Decision Making

Special Issue of the

## Journal of Applied Research in Memory and Cognition

*Guest editors:*

Julian N. Marewski and Ulrich Hoffrage

The *Journal of Applied Research in Memory and Cognition* (JARMAC) will publish a special issue on "Modeling and Aiding Intuitions in Organizational Decision Making", edited by Julian N. Marewski and Ulrich Hoffrage. Interested contributors are referred to a detailed outline of the intended contents below.

How do managers, civil servants, politicians, and other administrators make decisions? An avalanche of studies suggests that not only careful rational analyses, but also intuitions, gut feelings, and heuristics play an extremely important role in professional decision making—for the better or for the worse.

According to **dual-process** theories (e.g., Sloman, 1996), for instance, decision making stems from two cognitive systems; one which is rational, rule-based and one which is intuitive. Similarly, following the **heuristics-and-biases** program (e.g., Kahneman, Slovic, & Tversky, 1982), decisions are prone to a set of biases and irrational fallacies that are often attributed to the intuitive system. The **fast-and-frugal heuristics** framework (e.g., Gigerenzer, Todd, & the ABC Research Group, 1999), in contrast, stresses also what might be conceived of as the positive side of intuitions: According to this framework, successful decision makers smartly choose from an adaptive toolbox of efficient rules of thumb, labeled fast-and-frugal heuristics. Intuitions reflect the workings of these heuristics.

The different, partially segregated, theoretical approaches not only offer contradictory conclusions about the role of intuitions in organizational decision making, but also differ in the methodologies they rely upon. Dual-process theories and the heuristics-and-biases program often invoke verbal, informal accounts of decision making whereas the fast-and-frugal heuristics and other frameworks strive to formulate computational, algorithmic models of the underlying cognitive processes. For example, **cognitive architectures** (e.g., Anderson, 2007) and **connectionist** theories (e.g., Rumelhart, McClelland, & the PDP Research Group, 1986), potentially allow understanding decision processes in terms of very detailed formal models. The approaches also differ in terms of the benchmarks they use to assess the success of heuristic, intuitive decision processes. The heuristics-and-biases program, for instance, typically invokes the **laws of logic** and models that come from the **subjective expected utility maximization** tradition as normative yardsticks for successful decision making and human **rationality**. The fast-and-frugal heuristics framework, in turn, aims at assessing how well decision processes are adapted to the statistical structure of the environment in which they operate—an **ecological view of rationality** that is rooted in Herbert Simon's work (e.g., 1956). Finally, the various approaches differ in terms of how much emphasis they place on actually examining professional decision making in the real-world—as opposed to in the lab— with the **naturalistic decision making** community (e.g., Klein, 2004), making the study of intuitions in the wild one of its methodological priorities.

This rich but partially segregated literature does not offer a consensus as to (i) **how intuitive organizational decision making processes should be modeled and** (ii) **how organizational decision makers can be aided to make better decisions**. Yet, especially the latter question is of great importance to practitioners—such as managers, politicians, or civil servants—who strive to improve decision making processes in institutions.

This special issue intends to contribute to establishing such a consensus, helping practitioners and theorists alike in their endeavor to both understand and aid intuitive organizational decision making. In line with this goal, the special issue will not only present cutting-edge research in this domain, but also offer a synopsis of the various theoretical and methodological approaches in one volume. To further foster exchanges among these approaches, authors of accepted papers will be invited to publish a commentary on the contributions of the other authors (in the same volume).

Submitted articles should make a new theoretical, methodological, or empirical contribution, for example, by presenting theoretical arguments, experimental or observational findings, simulation results, and mathematical analyses. Articles that are explicitly written for practitioners are also solicited.

**Specific topics of full articles include but are by no means limited to:**

(a) How do intuitions guide managers, civil servants, politicians, and other administrators, for instance, when making high-stake and low-stake decisions?

(b) How can managers, civil servants, politicians, and other administrators avoid falling prey to cognitive biases by training their intuitions?

(c) How can heuristics and intuitions be systematically used to aid (rational) decision analysis, for instance, by guiding the construction of complex decision trees and by informing simulations of business scenarios?

(d) How can heuristics be implemented as decision aids in organizations?

(e) How can simple heuristic principles contribute to the robustness of organizations, institutions, or even society (cf. Taleb, 2010)?

(f) Why are there comparatively few detailed computational models of the cognitive processes associated with intuitive organizational decision making?

(g) How can cognitive architectures, connectionist models, and other computational theories of cognition aid the study of intuitive organizational decision making?

(h) How can the rational analysis approach from the cognitive and decision sciences (e.g., Anderson, 1991; Oaksford & Chater, 1998) be useful for studying intuitive decision making in organizations?

(i) When should correspondence criteria and when should coherence criteria (e.g., Hammond, 1996) come into play as normative yardsticks for assessing the success of intuitive decisions in organizations?

(j) How do intuitive decision making processes differ depending on whether they are studied in the wild or in the lab?

(k) How can the Brunswikian methodological imperative of representative experimental design (e.g., Brunswik, 1955) be applied in the study of intuitive organizational decision making?

(l) How can the different theoretical and methodological approaches to intuitive organizational decision making be integrated into an overarching framework?

Interested contributors are requested to contact Julian Marewski and Ulrich Hoffrage (by e mail: julian.marewski@unil.ch, ulrich.hoffrage@unil.ch; for more information about the guest editors, see www.modeling-adaptive-cognition.org) and to submit, as a preliminary step, a summary of the intended contribution (about 200 words). Each summary will be evaluated by the guest editors in terms of the intended contribution's scope and suitability for the special issue. Summaries that are submitted prior to December 31st will be given full consideration for the special issue; summaries that are submitted on a later date will also be considered; however, full consideration of late summaries will only be guaranteed as long as projected number of intended contributions does not exceed the available journal space. The deadline for submitting full papers is October 15th, 2013. Submitted papers will be reviewed within 4 weeks after their reception.

All submissions will be subject to the journal's regular peer review process under the direction of the guest editors and Ronald Fisher, the journal's editor-in-chief. The final version of accepted articles must adhere to the journal's author guidelines.

One goal of the *Journal of Applied Research in Memory and Cognition* is to reach not only scientists but also professionals and practitioners who seek to understand, apply, and benefit from research on memory and cognition. Editorial board members are JR. Belli, R. Bjork, N. Brewer, S. Charman, J. Dunlosky, R. Engle, B. Fischhoff, M. Garry, S. Gathercole, M. Goldsmith, P.A Granhag, A. Healy, P. Hertel, S. Kassin, G. Keren, J. Marewski, M. McDaniel, C. Meissner, J. Metcalfe, K Pezdek, D. Poole, H. Roediger III, B. Schwartz, N. Schwarz, D. Simon, B. Spellman, A. Vrij, G. Wells, C. Wickens, J. Wixted, and D. Wright. The journal is owned by the Society for Applied Research in Memory and Cognition, and published by Elsevier.

**Cited references**


Anderson, J. R. (1991). Is human cognition adaptive? *Behavioral and Brain Sciences*, *14*, 471–517.

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* New York: Oxford University Press.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*, 193–217.

Gigerenzer, G., Todd, P. M., & the ABC Research Group. (1999). *Simple heuristics that make us smart.* New York: Oxford University Press.

Hammond, K. R. (1996). *Human judgment and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice.* Oxford: Oxford University Press.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment Under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press.

Klein, G., (2004). *The Power of Intuition: How to Use Your Gut Feelings to Make Better Decisions at Work.* New York: Currency.

Oaksford, M., & Chater, N. (Eds.). (1998). Rational models of cognition. Oxford, England: Oxford University Press.

Rumelhart, D. E., McClelland, J.L., & the PDP Research Group (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations*, Cambridge, MA: MIT Press.

Simon, H. A. (1956). Rational choice and the structure of environments. *Psychological Review*, 63, 129-138.

Sloman S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3-22.

Taleb, N. N. (2010). *The black swan: The impact of the highly improbable (second edition).* New York: Random House.


## Philosophical Approaches to Social Neuroscience

Special Issue of **Cognitive Systems Research**


Edited by

Leslie Marsh (Medical School, University of British Columbia) and Philip Robbins (Department of Philosophy, University of Missouri)


### Confluence of Interest

It's been twenty-five years or so since Gazzaniga's (1985) empirically motivated work that understood the brain as a kind of hermeneutic device or "interpreter" that evolved in response to social forces. This work could be considered a landmark in the nascent field of social neuroscience (SN). From a philosophical perspective it's also been some twenty-five years since Churchland (1986) broke ranks with the a priorism characteristic of the prevailing philosophy of mind by taking heed of developments within neuroscience.

Social neuroscience, by definition, is an acknowledgement that the nervous system cannot be considered in isolation from the social environments in which humans have evolved. By the same token, the non-Cartesian wing of cognitive science is also a de facto acknowledgement that ubiquitous sociality must be factored into philosophy of mind. This said, there is still a very limited literature dealing with this clear confluence of interest. Of course, social neuroscience is not totally unknown to philosophy – possibly the most famous instance being the work of Gallese et al (1996), given philosophical currency via Gallese and Goldman (1998). But given the diversity of research projects that drive social neuroscience and "situated" philosophy of mind, the possible topics of philosophical investigation go well beyond mirror neurons.

The motivation behind this special issue is to harvest some of the results from SN with a view to:

(a) empirically enriching philosophy of mind, and

(b) philosophically informing social neuroscience.

To this end, we seek philosophical assessments of work being done in and around SN – including (but not limited to) work on mindreading, moral cognition, judgment and decision making, law and testimony, and social epistemology. Contributors are encouraged to scan the contents of two major journals that have social neuroscience as a dedicated interest: Neuroimaging (Elsevier) and Social Neuroscience (Taylor and Francis) as well as journals that have SN as a major interest, namely Neuropsychologica ( Elseveier), Journal of Cognitive Neuroscience (MIT), Journal of Personality and Social Psychology (APA) and Brain Research (Elsevier).

The list of topics includes empathy, altruism, social pain, attribution, the self, stereotyping (race, gender, etc.), and collective intentionality.

Some overlapping questions for consideration:

1) Methodologically speaking, how social is (or can) neuroscience really be if all that is measured is brain activity in non-social contexts, i.e. fMRI scanners? (Keysers & McKay, 2011). Put another way, does social cognition draw upon a distinct set of processes dissociable from non-social processes? (Jenkins & Mitchell, 2011)
2) What count as foundational results in SN? (Ochsner, 2004)
3) What sort of metaphysical and epistemological commitments does research in SN presuppose? To what extent is SN opposed to reductionism in the philosophy of science? (Decety & Cacioppo, 2010)
4) What drives the "techno-ebullience" surrounding neuroimaging in general, and neuroimaging in SN particular, and how might it be problematic for the field? (Vul et al, 2009; Decety & Cacioppo, 2010).

## Timeline

Official start: December 1, 2012
Final drafts due: February 1, 2014
Refereeing: February/March 2014
Final versions due: August 1, 2014

In the first instance we are looking for proposals of not more than 500 words. The aim is to have a broad spread of interest comprising the issue. Final papers should be between 7,500 and 9,000 words. Please send your proposals to both Philip and Leslie:

robbinsp@missouri.edu leslie.marsh@ubc.ca

## References

Churchland, P. (1986). Neurophilosophy: Toward a Unified Science of the Mind-Brain. Cambridge, MA: MIT Press.

Decety, J. & Cacioppo, J. T. (2010). Problems and Prospects in Social Neuroscience. Japanese Journal of Physiological Psychology and Psychophysiology 28(1): 5-16.

Decety, J. & Keenan, J. P. (2006). Social Neuroscience: A new journal. Social Neuroscience, 1.1, 1-4.

Gallese, V., Fadiga, L., Fogassi, L. and Rizzolatti, G. (1996). Action recognition in the premotor cortex. Brain 119: 593-609.

Gallese, V. & Goldman, A. (1998) Mirror neurons and the simulation theory of mind- reading. Trends in Cognitive Sciences, 12:493-501.

Gazzaniga, M. S. (1985). The Social Brain: Discovering the Network of the Mind. New York: Basic Books.

Jenkins, A.C. & Mitchell, J.P. (2001) How has Cognitive Neuroscience Contributed to Social Psychological Theory? In: Todorov et al.

Kahneman, D. (2011). Thinking, Fast and Slow. New York: Doubleday

Keysers, C. & McKay (2011). How to Make Social Neuroscience Social. Psychological Inquiry: An International Journal for the Advancement of Psychological Theory 22:3, 210-216.

McEwen, B. S. & Akil, H. (2011). Introduction to Social Neuroscience: Gene, Environment, Brain, Body. Annals of the New York Academy of Sciences 1231, vii-ix.

Ochsner, K. N. (2004). Current directions in social cognitive neuroscience. Current Opinion in Neurobiology 14: 254-258.

Ogawa, S, Lee, T. M, Nayak, A. S. & Glynn, P. (1990). Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. Magnetic Resonance Medicine 14: 68-78.

Todorov, A., Fiske, S.T., & Prentice, D.A. (2011). Social Neuroscience: Toward Understanding the Underpinnings of the Social Mind. Oxford: Oxford University Press.

Vul, E., et al. (2009). Voodoo correlations in social neuroscience. Perspectives on Psychological Science, 4(3): 299-307.

European Neural Network Society

International Neural Network Society
Institute of Information and Communication Technologies, BAS
Technical University of Sofia
Union of Automation and Informatics, Bulgaria


## Main Topics:

Neural network theory & models
Computational neuroscience
Neuronal automata
Connectionist cognitive science
Pattern recognition
Neuroeconomics
Neurofinance Graphical network models
Brain machine interfaces
Evolutionary neural networks
Neurodynamics
Complex systems
Neuroinformatics
Neuroengineering
Hybrid systems
Computational biology
Neural hardware
Bioinspired embedded systems
Collective intelligence
Self-orgnization
Embodied robotics
Learning algorithms
Neural data analysis
Cognitive models
Cellular Neural Networks

*Conference website: http://icann2013.org*
*Contact: icann2013@tu-sofia.bg ; pkoprinkova@yahoo.com*

## Organizing committee chairs:

**Honorary Chair:** Nikola Kasabov
**General co-chairs:**
Valeri Mladenov
Vassil Sgurev

**Program co-chairs:**
Petia Koprinkova-Hristova
Alessandro Villa
Wlodzislaw Duch
Günther Palm
Mincho Hadjiski

## Deadlines:

**Workshops/special sessions proposals**
March 1, 2013

**Tutorials/competition proposals**
March 1, 2013

**Papers**
March 15, 2013

# ICONIP2013

The 20th International Conference on Neural Information Processing
3 - 7 NOVEMBER 2013 / DAEGU, KOREA

## Paper Submission

Authors are invited to submit extended abstracts on specific topics or full-length papers (8 pages maximum) by the submission deadline through the online submission system. **Extended abstracts** must include the aim or rationale for the following specific topics: **Neuroscience, Neuro-psychology, Industrial applications, Cognitive science, and Demonstration.**
The submission of a paper implies that the paper is original and has not been submitted under review, or copyright-protected elsewhere and will be presented by an author if accepted. All submitted papers will be reviewed by area experts based on the criteria of originality, significance, quality, and clarity. The authors of accepted papers will have an opportunity to revise their papers and take consideration of the reviewers' comments and suggestions. The proceedings of ICONIP 2013 will be published by Springer in its series of Lecture Notes in Computer Science. Selected papers will be included in special issues of **Neural Networks, Neurocomputing, IEEE Trans. on Industrial Informatics (tentative), and Neural Processing Letters (tentative), etc.**

## Sponsored by

**Honorary Chair**
**Shun-ichi Amari**, RIKEN, Japan
**Soo-Young Lee**, KAIST, Korea

**General Chair**
**Minho Lee**, Kyungpook National University, Korea

**Program Chair**
**Akira Hirose**, The University of Tokyo, Japan
**Zeng-Guang Hou**, The Chinese Academy of Sciences, China
**Rhee Man Kil**, Sungkyunkwan University, Korea

## DAEGU, KOREA

Daegu Metropolitan City is the third largest city in Korea and is located in the southeastern part of the Korean peninsula. With an efficient transportation network linked in all directions, Daegu Metropolitan City is a gateway to numerous cultural heritages and tourist attractions, all proudly displaying the unique Korean culture. Daegu is at the heart of Buddhist and Confucian traditions and is therefore a perfect place to appreciate the true character of Asia. The organizing committee of ICONIP 2013 welcomes you to experience and enjoy the following tour programs.

### Historical / Cultural Tour-"UNESCO World Cultural Heritage sites"

Daegu Metropolitan City is surrounded by UNESCO World Cultural Heritage sites-such as Andong Folk Village, Yangdong Folk Village, Seokguram Grotto, Bulguksa Temple and Tripitaka Koreana.

Secretariat
iconip2013@gmail.com

### Industrial Tour

The hub of knowledge-based industry in northeast Asia: Daegu Metropolitan City is a place where industry thrives. The birthplace of Samsung, LG Electronics and POSCO, as well as leading multinationals such as Exxon Mobil, Dassault Systemes and Acciona, all have made sizable investments in Daegu Metropolitan and Gyeongbuk Province.


### Korean Wave Tour-Korean Drama / Movie Theme

Which is your favorite Korean drama or movie? If you are not sure, come and join us in this tour that includes the filming locations of drama 'Love Rain', 'Dae Jang-Geum' and 'Queen SeonDeok'.