Advances in Artificial Intelligence

Artificial Intelligence in Neuroscience and Systems Biology: Lessons Learnt, Open Problems, and the Road Ahead

Guest Editors: Daniel Berrar, Naoyuki Sato, and Alfons Schuster

Artificial Intelligence in Neuroscience and Systems Biology: Lessons Learnt, Open Problems, and the Road Ahead

Artificial Intelligence in Neuroscience and Systems Biology: Lessons Learnt, Open Problems, and the Road Ahead

Guest Editors: Daniel Berrar, Naoyuki Sato, and Alfons Schuster

Editorial Board

Contents

Hindawi Publishing Corporation
Advances in Artificial Intelligence
Volume 2010, Article ID 578309, 2 pages
doi:10.1155/2010/578309

Editorial

Artificial Intelligence in Neuroscience and Systems Biology: Lessons Learnt, Open Problems, and the Road Ahead

Daniel Berrar,[1,2] Naoyuki Sato,[3] and Alfons Schuster[4,5]

[1] *Systems Biology Research Group, Centre for Molecular Biosciences, School of Biomedical Sciences, University of Ulster, Cromore Road, BT52 1SA, Coleraine, Northern Ireland*
[2] *Systems Biology Department, Cancer Institute, Japanese Foundation for Cancer Research, Tokyo, Japan*
[3] *Department of Complex Systems, Future University Hakodate, 116-2 Kamedanakano-cho, Hakodate, Hokkaido 041-8655, Japan*
[4] *School of Computing and Mathematics, Faculty of Computing and Engineering, University of Ulster, Shore Road, New-Townabbey, Co. Antrim, BT37 0QB, Northern Ireland*
[5] *Laboratory for Dynamics of Emergent Intelligence, RIKEN Brain Science Institute, Wako-shi, Saitama 351-0198, Japan*

Correspondence should be addressed to Daniel Berrar, dp.berrar@ulster.ac.uk

Received 31 January 2010; Accepted 31 January 2010

> *"We can only see a short distance ahead, but we can see plenty there that needs to be done."*
> Alan M. Turing

Since its conception in the mid 1950s, artificial intelligence with its great ambition to understand intelligence, its origin, and creation, in natural and artificial environments alike, has been a truly multidisciplinary field that reaches out and is inspired by a great diversity of other fields in perpetual motion. Rapid advances in research and technology in various fields have created environments into which artificial intelligence could embed itself naturally and comfortably. Neuroscience with its desire to understand nervous systems of biological organisms and system biology with its longing to comprehend, holistically, the multitude of complex interactions in biological systems are two such fields. They target ideals artificial intelligence has dreamt about for a long time including the computer simulation of an entire biological brain or the creation of new life forms from manipulations on cellular and genetic information in the laboratory.

The scope for artificial intelligence, neuroscience, and computational systems biology is extremely wide. The motivation of this special issue is to create a bird-eye view on areas and challenges where these fields overlap in their defining ambitions and where these fields may benefit from a synergetic mutual exchange of ideas. The rationale behind this special issue is that a multidisciplinary approach in modern artificial intelligence, neuroscience, and systems biology is essential and that progress in these fields requires a multitude of views and contributions from a wide spectrum of contributors. This special issue, therefore, aims to create a centre of gravity pulling together academic researchers and industry practitioners from a variety of areas and backgrounds to share results of current research and development and to discuss existing and emerging theoretical and practical problems in artificial intelligence, neuroscience, and systems biology transporting them beyond the event horizon of their individual domains.

If the contributions in this special issue are to be classified (crudely) according to their main thrust, then the articles close to the neuroscience camp are devoted to the themes of (i) the characterization of in vitro neuronal networks cultured on multielectrode array (MEA) chips, (ii) the simulation of human episodic memory by using a computational model of the hippocampus, and (iii) information processing in natural and artificial olfactory systems. Novellino and Zaldívar propose a combination of recurrence quantification analysis based on recurrence plots and conventional statistical analysis for neuronal electrophysiology. They investigate their approach by studying the variation of spontaneous electrophysiological activity of in vitro neuronal networks on multielectrode array chips. Sato and Yamaguchi review computational models of the hippocampus and discuss their own computational model of human episodic memory based on neural synchronization. Using computer simulations and human eye movement data, they demonstrate the validity of their model to predict human memory recall. From

an evolutionary perspective, Stafford reviews the biological constraints of the physical properties of neurons and the implication for the construction of artificial neural networks. van der Velde discusses fundamentals of human cognitive processes and proposes models for grounded architectures of cognition. Makino and Yano investigate olfaction as a relatively simple biological information processing system and report their computational works with a focus on the temporal dimension.

The article by Browne et al. reviews computational techniques to infer protein-protein interaction networks, which may help decipher novel drug targets.

Two articles are supported by a strong background in artificial intelligence and focus on learning algorithms. Schuster and Yamaguchi investigate game theoretic concepts and present a novel learning algorithm for a paired neuron system. Sridharan reports on novel bootstrapped learning techniques to process visual inputs that allow a mobile robot to autonomously adapt its behavior to illumination changes.

In summary, this special issue informs the research community at large about an exciting and stimulating relationship between artificial intelligence, neuroscience, and systems biology. The special issues provides access to many state-of-the-art theoretical and applied problems in these hugely exciting fields that are so relevant for modern science. This special issue is also intended as a platform to bridge cultural and technological gaps between these disciplines. Ultimately, the contributions in this special issue should convey to its readership the enthusiasm the editors and authors of this issue share for their respective fields.

Daniel Berrar
Naoyuki Sato
Alfons Schuster

Hindawi Publishing Corporation
Advances in Artificial Intelligence
Volume 2010, Article ID 629869, 12 pages
doi:10.1155/2010/629869

Research Article

Quo Vadis, Artificial Intelligence?

Daniel Berrar,[1,2] Naoyuki Sato,[3] and Alfons Schuster[4,5]

[1] *Systems Biology Research Group, Centre for Molecular Biosciences, School of Biomedical Sciences, University of Ulster, Cromore Road, BT52 1SA Coleraine, UK*
[2] *Systems Biology Department, Cancer Institute, Japanese Foundation for Cancer Research, Tokyo 1358550, Japan*
[3] *Department of Complex Systems, Future University Hakodate, 116-2 Kamedanakano-cho, Hakodate, Hokkaido 041-8655, Japan*
[4] *School of Computing and Mathematics, Faculty of Computing and Engineering, University of Ulster, Shore Road, Newtownabbey, County Antrim BT37 0QB, UK*
[5] *Laboratory for Dynamics of Emergent Intelligence, RIKEN Brain Science Institute, Wako-shi, Saitama 351-0198, Japan*

Correspondence should be addressed to Daniel Berrar, dp.berrar@ulster.ac.uk

Received 9 October 2009; Accepted 4 January 2010

Academic Editor: David Glass

Since its conception in the mid 1950s, artificial intelligence with its great ambition to understand and emulate intelligence in natural and artificial environments alike is now a truly multidisciplinary field that reaches out and is inspired by a great diversity of other fields. Rapid advances in research and technology in various fields have created environments into which artificial intelligence could embed itself naturally and comfortably. Neuroscience with its desire to understand nervous systems of biological organisms and systems biology with its longing to comprehend, holistically, the multitude of complex interactions in biological systems are two such fields. They target ideals artificial intelligence has dreamt about for a long time including the computer simulation of an entire biological brain or the creation of new life forms from manipulations of cellular and genetic information in the laboratory. The scope for artificial intelligence in neuroscience and systems biology is extremely wide. This article investigates the standing of artificial intelligence in relation to neuroscience and systems biology and provides an outlook at new and exciting challenges for artificial intelligence in these fields. These challenges include, but are not necessarily limited to, the ability to learn from other projects and to be inventive, to understand the potential and exploit novel computing paradigms and environments, to specify and adhere to stringent standards and robust statistical frameworks, to be integrative, and to embrace openness principles.

1. Introduction

Since its foundation, which is often associated with a conference held at Dartmouth College in New Hampshire in 1956, artificial intelligence (AI) can look back on a rather exciting and successful (though not unproblematic) career.Per definition, artificial intelligence—which may be described as a field of study embracing concepts, methodologies, and techniques as part of a computer program that exhibits characteristics akin to intelligent behavior [1]—is a sensitive and vulnerable field as the term *intelligence* itself lacks a generally (ideally universally) acknowledged definition. There is, of course, no shortage of characterizations for the term intelligence (e.g., [2, pages 6–21] or [3, pages 1–4]), at large; however, the term remains a complicated, subtle, and malleable topic. In addition, although, arguably, many animals are intelligent to some degree, by intelligence people often assume the same scope of intelligence as can be seen in general purpose human action (e.g., [4] mentions the creation of artifacts that are capable of mimicking and expressing human intelligence, thought, consciousness, and emotion). Although the field has been exposed to some (often well-deserved) criticisms over the years, it may be fair to say that AI is a truly great scientific endeavor that reaches out and embraces a great variety of scientific disciplines ranging from areas that are mathematically and conceptually well defined (e.g., computer science, machine learning, biology, and neuroscience) to areas that are more difficult to quantify and deal with (e.g., philosophy of mind, cognitive science, and emotional intelligence). Of course, it is not possible here to acknowledge adequately the many contributions (including those coming from rapid advances

in computer hardware and software) that shaped AI into the exciting discipline it is today, but it may be possible to categorize, crudely, the development of AI into a few key stages (e.g., see [3, pages 1–24]).

Early AI projects had to come to terms with the many challenges related to the production of knowledge-based (expert) systems, including the acquisition of knowledge, its representation and evolution, among other challenges. The General Problem Solver, the early DENDRAL and MYCIN expert systems, or the ELIZA program may stand representatively for this period.The General Problem Solver is a general purpose program to simulate human decision-making, thinking, and reasoning; DENDRAL (a system for analyzing chemicals) and MYCIN (a system for the diagnosis of infectious blood diseases) are knowledge-based expert systems and focus on the problem solving in specific domains rather than on a general problem-solving strategy that is applicable to many (all) domains; ELIZA is a program related to natural language processing (and the Turing test) where humans interacting with a machine in a question-answer script-based mode were led to believe that they were actually participating in a human-human interaction.

Artificial intelligence received a major boost in the mid 1980s from works on artificial neural networks (ANNs). It is possible to say that since the early days of the perceptron the story of artificial neural networks is one of the most successful chapters in the voluminous book of AI. Since the emergence of several techniques, many of them united under the soft computing paradigm then provided AI with a capable and flexible repository for both theoretical research as well as hands-on problem-solving applications. Possibly, from an application point of view, this was one of the most exciting times for AI and the term *knowledge engineering*, which suggests that systems showing some degree of intelligent problem-solving ability could be assembled from a toolbox of available techniques, may capture this excitement quite well.

In more recent times, AI has made it into the limelight through the DARPA (Defense Advanced Research Projects Agency) Grand Challenges (http://www.darpa.mil/grandchallenge/index.asp).

In these challenges, vehicles had to navigate autonomously in increasingly challenging environments in an intelligent manner, avoiding obstacles and solving problems of increasing levels of complexity, without human intervention (e.g., [5]). Without hesitation one needs to acknowledge the achievements in these tasks, paralleled perhaps in their prestige and popularity by the outstanding successes of IBM's Deep Blue supercomputer (the machine that recorded the first win in a game of chess against a reigning World Chess Champion, Garry Kasparov, in 1996) and the Deep Fritz chess engine (the commercial chess engine that triumphed 4-2 in December 2006 against the reigning champion Vladimir Kramnik in a six-game match) (http://www.research.ibm.com/deepblue/). It is beyond the scope of this paper to pay credit to the great variety of important and popular areas in which AI has made substantial contributions (often without being duly credited, taken for granted, or simply neglected) but it is helpful to emphasize that these areas include mainstream utilities such as the World Wide Web (e.g., the so-called *intelligent web* draws its power from algorithms that process information intelligently—Google's page ranking algorithm or algorithms discovering matches on social-networking sites are two examples [6]), toys and gadgets (e.g., computer games [7] or the Lego Mindstorms Robotics Invention System [8]), extremely helpful human-computer interfaces (e.g., user-friendly interfaces using natural language recognition and visualizations including brain-computer interfaces [9, 10]), or humanoid robots functioning in various roles as partners for people in the immediate environment of human beings (e.g., AIBO, ASIMO) (e.g., http://world.honda.com/ASIMO/). Artificial intelligence has even reached out beyond earthly confines and is heavily utilized for numerous tasks by several space agencies including the European Space Agency (ESA)(e.g., [11]) and NASA (National Aeronautics and Space Administration) (e.g., http://www-aig.jpl.nasa.gov/).

The previous points may suggest that contemporary AI should find itself in a comfortable situation from which it should firmly and enthusiastically continue its cause—the study and creation of artificial entities that are capable of expressing human intelligence, thought, consciousness, and emotion. The paper argues that this may not necessarily be so and that AI may need to be cautious and perhaps alert to some degree about its standing because just around the corner there are extremely exciting and powerful fields that touch upon areas—and begin to outshine AI in areas—that are at the very heart of AI itself. These fields include, but are not necessarily limited to, neuroscience [12], synthetic biology, and systems biology [13]. These disciplines comprise several subfields with boundaries that are often blurred. They also exemplify many modern research endeavors that are characterized by their complexity and a truly interdisciplinary nature that draws upon the expertise of scientists from a wide range of academic backgrounds.

The forthcoming sections investigate how AI is situated in this extended environment. Initially, Section 2 takes a closer look at the interplay between AI, neuroscience, synthetic biology, and systems biology. Section 3 identifies several challenging hurdles in this interplay and includes suggestions for how these hurdles may be overcome in order to create a synergetic and productive environment that is beneficial and supportive to practitioners working in these fields. Section 4 provides concluding remarks and ends the paper with a summary.

2. Quo Vadis, Artificial Intelligence?

A starting point for an answer to this question may be one of the most successful scientific endeavors in the last century—the Human Genome Project. The Human Genome Project achieved its goal, the definition of the sequence of chemical base pairs which make up DNA, on 26 June 2000 with the publication of a first draft of the human genome (acknowledged in special issues in Nature [14] and Science [15]). Simply speaking, the human genome consists of DNA

(deoxyribonucleic acid). This DNA is assembled from a small number of basic nucleotides (adenine, guanine, cytosine, and thymine). These elementary building blocks bond together into DNA sequences. Genes, which are another important component in this context, are particular DNA sequences that play a fundamental role in the evolution and production of organisms (e.g., the development of a human being). At root, DNA is a code, and like any code DNA contains instructions and information (e.g., for the building of complex, 3-dimensional proteins). The genome contains the full set of instructions. Understanding (in its entirety) this code (the genome) is a major goal today. Crucially, the information in the human genome unfolds itself fully over time at different biological levels of structural organization and complexity, each level with its own important functions. From the bottom upwards, these levels may be molecules, genes, more complex cell formations, organs, and, on the highest level, complex higher organisms including human beings. Of course, the abundance of inspiring challenges in this environment invites a growing number of scientists from many fields, including biologists, AI practitioners, and other individuals whose fields have a long tradition in studying complex systems, data, codes, and information processing to pick and investigate hugely exciting problems in this overwhelmingly rich application area of study.

2.1. AI in Synthetic Biology and Systems Biology. A feature that unites systems biology and synthetic biology is the tremendous complexity that is inherent in both fields. Asked for a dividing line, one may argue that systems biology is motivated to create a complete and detailed understanding of existing biological systems whereas synthetic biology is driven by the vision to create novel biological entities from first principles.

2.1.1. Synthetic Biology. Synthetic biology is defined as the design and construction of new biological parts, devices, and systems, and the redesign of existing, natural biological systems for useful purposes (e.g., http://syntheticbiology.org/). Hence, synthetic biology envisages the design and fabrication of biological components and systems that do not exist in the natural world from non-living, raw material and programming them with desired (novel) chemical functionality. The field also envisages the redesign of existing biological systems [16]. In addition, it is helpful to understand that the terms (artificial) synthetic life and artificial life (not to be confused with Alife, the field of study that examines systems related to life mainly through computer simulation applications) are related to synthetic biology and apply when the goal is to recreate life from non-living (abiotic) materials (e.g., [17]). It is important to understand that artificial synthetic life and synthetic biology have the support of a strong lobby and that there are strong beliefs that artificial cells will eventually be created. Anyway, in the context of this paper it is necessary to underline that there are various concepts in these fields that should be of great interest for the AI community. For instance, the *minimal genome* (the smallest set of genes needed to support a simple living cell) is an important aim of artificial synthetic life [18]. From an AI perspective, this concept of a minimal genome is rather interesting and may map to the term *minimal intelligence*, which may be defined as the smallest amount of information needed to support intelligent behavior. The production of a minimal cell could intrinsically involve some intelligent processing that may be realized on the DNA level. It is possible to envisage a simulation of this processing involving some form of AI in an environment that exploits features of the popular DNA computing paradigm as models for various machine learning techniques (e.g., rough set analysis [19]) and other modeling approaches (e.g., Petri nets [20]) exist in the DNA computing world for quite some time. In this case, questions may arise such as: how much artificial intelligence or computing is needed for the construction or support of an artificial cell? Concepts such as *minimum cell computation* or *minimum cell information* may also emerge in this context. Overall, it is relatively easy to see that all these terms are highly relevant to AI and that synthetic biology certainly and quite naturally addresses several fundamental issues AI has had on its agenda since its very beginning.

2.1.2. Systems Biology. The field of systems biology dedicates itself to the study of complex biological systems, their properties, interactions and dynamics [13], and draws on a similarly wide pool of fields including several of the fields mentioned before. Of paramount importance is the holistic paradigm (e.g., http://www.systems-biology.org/), which is in stark contrast to the reductionist approach that was prevailing in molecular biology during the last decades. It should be noted here, however, that systems-level approaches to natural phenomena are not an invention of the 21st century. For example, Alan Turing, widely regarded as the grandfather of AI, adopted systems-level approaches in his studies on the chemical basis of morphogenesis [21], and could therefore rightfully be regarded as a pioneer in systems biology.

The characteristics of complex biological systems often emerge naturally via the interplay of individual system components. A typical example of such an emergent phenomenon is *robustness* [22]. Robust systems are inherently able to maintain their function despite internal or external perturbations, changes in the environment they operate (or live) in, or unreliable components. However, these individual components themselves may not be considered robust. Hence, robustness is an emergent property of the system as a whole and cannot be described or understood by studying the components alone. This preservation of function does not prevent complex systems from evolvability—in contrast, evolutionary principles appear to favor robust systems [22]. Although the principles behind robustness are not fully understood, there are specific cornerstones upon which robustness is believed to rely in various domains: (i) *modularity* the system is composed of elements that work together synergistically; (ii) *redundancy* some of the components share identical function, hence the depletion of one component can be compensated by others; (iii) *feedback control* the system is equipped with a sensor for the detection

of changes in the environment and a controller for reacting to these changes, which allows dynamic system behavior. This feedback control also includes mechanisms for systems repair such as purging, the deliberate destruction of components that fail to operate properly (e.g., outside of their defined range of action); (iv) *spatial compartmentalization* the complex system has a physical or virtual embodiment that is subdivided into areas or compartments that may exchange information with each other; (v) *distributed processing* the modular elements in the compartments collectively give rise to a higher, system-level function, phenotype, or morphology. It is hypothesized that for example cancer is a complex robust system [23], relying on a functional redundancy of genetically heterogeneous cells, and on the ability to maintain homeostasis and functionality despite changes of the tumor microenvironment. Although these changes may affect an individual tumor cell to a larger extent than a healthy cell, the collective of tumor cells is robust due to functional redundancy caused by the genetic variability of its components. Hence, insights about the mechanisms leading to robustness could help us identify potential Achilles' heels in systems whose function we seek to destroy. Unraveling key mechanisms leading to robustness represents one of the grand challenges of systems biology. While methods from cybernetics and control theory are arguably at the forefront of modeling systems dynamics, there is a clear scope for AI approaches. For example, the multi-facetted appeal of robustness for AI has been addressed in a recent edited book [24] that investigates robustness in a diverse range of areas related to AI including computer hardware and software, computer networks and protocols, brain-computer interfaces, biological networks and immune systems, humanoid robotics, image processing, artificial neural networks, genetic algorithms, chaos theory, and other soft computing techniques, as well as space system design and bioregenerative life support systems. This could be complemented by approaches from artificial life, which aims at understanding properties of life by abstracting its fundamental, evolutionary principles, and recreating and emulating them using computer programs. In fact, experiments with digital organisms have already revealed astonishing insights into the genetic basis of evolution [25]. In any case, the aforementioned book agrees that robustness is not a trivial topic, that in truth robustness is a rather elusive and challenging concept, but the book also clearly emphasizes that features of robustness, as they appear in nature, have a great potential for being utilized as general design principles for AI systems [26].

Further, the philosophy of systems biology is rooted in the desire to develop mathematically-founded, testable theories to *understand* whole biological organisms. Such theories, however, are only in their infancy. Mathematical models are still simplistic compared to the daunting complexity of real biology and therefore often met with a justified skepticism [27]. Yet, even the very notion of understanding in this context may be debatable—what do we mean by *understanding* a living system, actually? Does understanding imply that we grasp the *nuts and bolts* of the intricate living machinery? Or is it sufficient to make reliable predictions about the behavior of the system under normal conditions and under specific perturbations? Would it be possible to devise tests for systems biology models that are able to determine whether we have *understood* the system at hand? In AI, the Turing test is the ultimate (though not unchallenged) benchmark test for machine intelligence [28]. What could represent an equivalent test for a computational model in systems biology that claims to *understand* the living system? Harel [29] proposed several modifications to the Turing test in order to validate an artificial model of a living system, but the question of *understanding* is more fundamental in essence. It seems sensible to consider these delicate questions in order to have a yardstick against which we can measure our models (in systems biology, neuroscience and other disciplines); this would also preclude a moving of the goalpost after the models have been developed. Scientists from (philosophical) AI could provide valuable input to this debate.

2.2. AI and Neuroscience. As mentioned before in Section 2, the information in the genome unfolds itself fully over time at different biological levels of organization and complexity (e.g., from molecules to genes, more complex cell formations and organs, up to complex higher organisms such as a human being). Neuroscience with its diverse range of subfields complements this chain in a natural way by investigating (among many other things) the cognitive functions inherent in such organisms. The phrase *from molecules to cognition* therefore is sometimes used to summarize the field.

Neuroscience is a tremendously complex field with many subfields (e.g., [12]) and offers a wide range of opportunities for AI. Neural signalling, for instance, investigates how neural systems acquire, coordinate, and disseminate information. Knowledge about these processes is fundamental to understanding brain pathologies, but also for the development of novel approaches to diagnosing and treating such problems. Artificial intelligence systems can benefit from such an understanding in the field of artificial neural networks in particular and other application areas where networks play a fundamental role in general (e.g., pervasive/ubiquitous computing [30] and autonomic computing [31] aim for a new type of networked communication systems that are autonomously controlled, self-organized, radically distributed, technology independent, scale-free, and can manage themselves to various degrees given high-level objectives from administrators).

Sensations and sensory signal processing, including data storage in memory, are other areas that are key to neuroscience (e.g., [32]). The signal processing in biological vision and olfactory systems, for instance, is heavily investigated and there is a direct link to pattern recognition, knowledge representation, and related AI areas. The perception of information from an environment (internal and external) and the accurate and meaningful digestion and interpretation of such information in real-time is a vitally important task for biological nervous systems in general, and, holistically, the field of robotics is an AI working area that encapsulates many of these issues.

The changing brain and complex brain functions are fundamentally important in neuroscience. The development of a brain throughout the lifespan of a biological organism, where the brain changes both its size (natural development) and its content (real-life experiences) is also highly relevant to AI. For example, it is relatively easy to draw parallels to research on autonomous intelligent agents roaming in an (intelligent) World Wide Web that undergoes changes in size and information content in perpetual motion [6].

It is important to understand that the relationship between AI and neuroscience works vice versa. For example, the insight that many networks in nature are scale-free [33] may be informative for neuroscience, it is, however, a finding that comes from a study researching the large-scale topology of (a portion) of the World Wide Web that had very little to do with neuroscience. It is also important to understand that some projects in neuroscience are similar in character and scope to the Human Genome Project, they are large-scale and have an extremely ambitious, clear-cut goal. The prime example is the well-known Blue Brain Project at École Polytechnique Fédérale de Lausanne (http://bluebrain.epfl.ch/) in its attempt to reverse-engineer the mammalian brain in order to understand brain function and dysfunction through detailed simulations, which is an incredible undertaking. However, this goal (the creation of artificial entities that may eventually be capable of intelligence, thought, consciousness, and emotion) is very much what AI practitioners have been dreaming about for a long time—but, per se, the Blue Brain Project is a neuroscience project and not a project in AI. (Note that a similar argument holds for the previous sections on synthetic biology and systems biology. These fields address several key AI issues, but the term AI, in relation to its core business, the creation of artificial intelligence), appears relatively rarely in these areas. There are exceptions of course, but very often when AI is mentioned in these areas it is usually in a data analysis, data mining, or applied machine learning way, (e.g., [1].) This does not mean at all that the Blue Brain Project, and other brain projects, explicitly avoid the mentioning of AI and its mission [34]. It is just that AI, at the moment, simply does not play the role it could play (and perhaps due to its tradition should play) in such projects. It is as if AI has somehow disappeared from the Blue Brain Project and neuroscience (and synthetic/systems biology) at large.

On the basis of the previous sections—which identified a possible rich interplay between AI, neuroscience, synthetic biology, and systems biology—this paper feels that this is a somewhat surprising situation. The forthcoming sections therefore examine this situation in more detail and make suggestions that may help AI to raise its profile in neuroscience, synthetic and systems biology.

3. Hurdles and the Road Ahead

This section takes its motivation from the findings just mentioned and suggests several challenges AI may face in order to become more mainstream in neuroscience, synthetic and systems biology. These suggestions include, but are not necessarily limited to the following: (i) to be able to learn from other projects and to be inventive, (ii) utilization of novel computing paradigms and environments, (iii) development of standards, (iv) stringent computational and statistical frameworks, (v) adoption of *openness* principles, and (vi) to be integrative and creative.

3.1. Learning from Other Projects and Being Inventive. Although these first items are more like soft factors than hard factors they can be rather important for AI because success is as much the result of hard work as it is the result from being open and flexible enough to learn from others. It is not without reason that the capacity to learn stands at the heart of many AI systems, and it is clear that learning is an extremely powerful problem-solving strategy not only for AI systems but humans (and other organisms) in general. So what can AI learn from projects such as the Human Genome Project or the Blue Brain Project? Initially, it is necessary perhaps to highlight some of the outstanding features in such projects simply from a project management perspective. In general, these projects are: large-scale, team-oriented, inherently cross-disciplinary, collaborative, distributed, heterogeneous, and global. Crucially, on top of all this these projects are garnished by a truly *grand vision*! (The history of mankind is full of examples for this. In more recent times, the early Apollo missions with their wonderfully inspiring goal of bringing men to the Moon—and the currently orbiting International Space Station research facility alike—perhaps hit the nail on the head.) This does not necessarily imply, of course, that projects lacking these features are doomed to fail, but several positive developments in the former framework suggest that AI may benefit from adopting such an approach for future projects. The DARPA Grand Challenges. (The DARPA Grand Challenges are hugely attractive and successful but carry an ethical dimension with them that should not be underestimated. Some people may be reluctant to participate in these challenges or oppose them in principle because they are organized by the United States Department of Defense, which is responsible for the development of new technology for use by the military; the ethical dimension plays a similarly important role in neuroscience and synthetic biology, too.) mentioned earlier in Section 1 or the concepts behind the organization of various X-Prizes (http://www.xprize.org/) may be trendsetting in this regard in many ways. Essentially, these events generate an extremely rich pool of expertise by drawing together elements of public interest, entrepreneurial spirit, and cross-disciplinary innovation. The dream is that this pool eventually creates breakthroughs that are beneficial to mankind at large.

This does not mean that there is a general lack of projects in AI embracing this philosophy in principle. In many cases, however, such projects operate on a somewhat smaller scale for several reasons (financial support, popularity, etc.). The Loebner Prize in Artificial Intelligence (http://www.loebner.net/Prizef/loebner-prize.html) (a competition in the format of a standard Turing test where a machine has to demonstrate human-like ability/intelligence),

or the RoboCup (http://www.robocup.org/) (a competition where robots are tested in their ability to play soccer) may stand representatively for such projects. It is important now to understand that the scale of a project is a crucial factor, but also that scale alone may not be enough. There is another major obstacle that can deny even extremely exciting projects the support, attention, and success they may deserve—*individualism*. Take the case of humanoid robotics as an example. In the RoboCup challenges just mentioned, teams develop their applications individually and compete against each other again individually. The same accounts for industry driven projects in humanoid robotics. AIBO, for instance, is a Sony project, and ASIMO a humanoid robot created by Honda. If humanoid robotics is important enough for so many enthusiasts and such industry powerhouses then why is there no Humanoid Robot Grand Challenge with all resources open to all? Since robotics is only one of many examples, perhaps, what is required is a new way of thinking and collaboration in science at large. Actually, there are several rather interesting initiatives that point in the right direction. For example, the Santa Fe Institute (http://www.santafe.edu/) devotes itself to creating a new kind of scientific research community, and the Edge Foundation, Inc. (http://www.edge.org/) may be seen as a supersophisticated sphere where some of the greatest minds of our time engage in an intelligible and enthralling discourse on a diverse range of fundamental issues. Since Section 3.5 is going to comment further on this attitude of *putting all cards on the table* this section wants to add that being able to learn from other projects and to be open for change may get you some way but to make a truly large step forward may require another feature—the ability to be *inventive*.

3.1.1. Being Inventive. To be able to be creative and inventive is a great thing to be and nature reminds us kindly and patiently day by day how important it is to be clever and original, equally, in the small and in the large. It is difficult, if not impossible, to specify a process for inventiveness, but one way for AI may be to first of all identify novel and exciting fields that are relatively unexplored by AI or to get involved more deeply into current working areas with great potential for the future. From a potentially larger pool of fields, computer games, virtual reality, nanotechnology, quantum computing, and space exploration may be such fields.

Artificial intelligence has been in touch with the fields of computer games [7] and virtual reality [35] almost from the point when these fields took off. Many projects, including those of commercial organizations, provide game engines and other artifacts for free but what is missing, and what may be desirable according to this paper, is the transition from small-scale investigations conducted by individuals or individual institutions to a coordinated effort in the spirit emphasized in the previous sections. The virtual world *Second Life* (http://secondlife.com/) developed by Linden Lab and launched in 2003 demonstrates rather well the great potential such an approach may have for computer games, virtual reality, and AI. Second Life exploits free software utilization and the internet and provides users with an open environment in which they are free to express their creativity by designing artifacts with which they are free to interact and participate individually or in groups in a continuously changing and evolving virtual world.

In comparison to computer games and virtual reality, nanotechnology and quantum computing are two areas that are rather unexplored by AI. Nanotechnology is a rather diverse field again, and among other things the field researches new approaches to developing and processing new materials with dimensions on the nanoscale. Most people possibly are familiar with ideas such as micro-robots cleaning blood vessels, but then, where there is a robot there should be room for AI and indeed there is [36]. Quantum computing is another field with great potential for AI and there are attempts for combining the two fields (e.g., [37, 38]). In many ways the weird world of quantum computing with its sometimes unsettling conundrums (parallel universes, entanglement, etc.) has much of the charm of the early AI (but arguably a higher degree of robustness and credibility when it comes to interpretations that could easily come from the science fiction camp).

The exciting domain of space exploration with its large-scale dimension has a similar appeal for AI, and in fact there are many projects in which AI already faces the great challenge of space. NASA and ESA, for example, pursue large space research programs in which AI plays a significant role; be it in the design of life-support systems [39] or the development of autonomous systems for satellite path planning [40]. Actually, a very attractive thing to do for AI in this spirit could be a grand challenge with the goal of sending a humanoid robot to the Moon. Perhaps, one of the tasks (not to be taken too seriously here) for this robot could be to play a round of golf on the Moon, mimicking the attempts demonstrated by NASA astronaut Alan Shepard who was the first human being hitting golf balls (with a six iron) on the Moon on the Apollo 14 mission in 1971. Overall, these examples indicate that there is great potential for AI in several cutting-edge areas. Whether AI can exploit these challenges to its advantage (e.g., by clearly defining a stimulating set of (truly) grand challenges (in the areas mentioned here or any other area for that matter) is a question open for speculation that only future AI enthusiasts may be able to answer.

3.2. Exploitation of Novel Computing Areas and Computing Paradigms. The mentioning of the terms nanotechnology or quantum computing in the previous Section 3.1.1 provides an introduction to an issue that is rather important for AI—the exploitation of novel computing areas and computing paradigms. In the past, a standard computer was largely the platform for most AI systems (e.g., a desktop PC or an application embedded in a mobile robot). Our time, however, has seen the coming of new computing paradigms such as quantum computing or DNA computing. This development creates fundamental changes in terms of computer hardware and software (e.g., the difference between a standard PC and a DNA computer is sometimes expressed by saying that for a standard PC computing is computing with *bits*, whereas

for a DNA computer it is computing with *molecules*). Similar arguments hold for work in synthetic biology (Section 2.1.1) where the main working entity (processor) is the cell (e.g., the relatively young field of membrane computing is a new computational model inspired by the processing found in biological cells [41]).

There are also dramatic developments happening in neuroscience at the moment. Historically, computational neuroscience has been a branch of AI [42], with a focus on modeling and theorizing functional aspects behind brain signals. Artificial, yet biologically plausible, neural networks like in the Blue Brain Project have long been playing a central role in the interpretation of brain signals and mechanisms of brain functions. The evaluation of artificial neural networks in this field often requires realistic, real-world environments, which are facilitated by robotics and computer vision. Methods that originated from AI play now an equally important role in the analysis of brain signals from both animals and humans. Sophisticated brain-machine interfaces [10], for instance, exploit the computational power of a human brain, aiming at the translation of brain signals into motor commands for robotic device control [43]. Pilot studies in this field have shown remarkable results with great promises for severely disabled people (e.g., for neuroprostheses [44]).

Along this line, *brain-like* computers are another exciting area in neuroscience. For instance, IBM has engaged into an activity researching brain-like computers and the newly coined term *cognitive computing* encapsulates the key idea to engineer *mind-like* intelligent machines by reverse engineering the structure, dynamics, function and behavior of the brain [45]. Crucially, this project targets an understanding of the *mind* (decade of the mind [46]) based on the findings from the last decade of brain research.

The development of new neuroscience technologies for learning and memory in vitro is another relatively new and highly promising area in neuroscience. Researchers in this field grow mammalian brain cells in culture on multielectrode arrays in order to form a long-term, two-way interface between cultured networks and a computer. The generated cultured nets then can serve as brains of simulated so-called animats or robotic creatures, which opens the door widely for all sorts of AI research related to neuroscience driven robotics (e.g., [34, 47]). Other interesting contributions, related to these studies, come from work investigating predictive behavior within microbial genetic networks where bacteria anticipate changing environments [48]. The fact that bacteria have no brains or nervous system makes this work particularly interesting, and there are suggestions that these microbes experience and learn through evolutionary changes in their complex networks of interacting genes and proteins (i.e., the problem-solving potential is part of the configuration of the system) [49]. A key element of this system is *quorum sensing*, the regulation of gene expression in response to fluctuations in cell-population density [50]. Bacteria can exploit this naturally occurring, self-organizing principle as a means of communication with each other in order to collectively solve problems, or to give rise to collective phenomena such as bioluminescence. Interestingly, there is a parallel between quorum sensing and *swarm behavior*, the emergent property of a collective of individuals acting in concert. Key concepts of the collective behavior are bottom-up (instead of top-down) approaches, and the lack of a central command-and-control structure (i.e., decentralization), which have also been described as *swarm intelligence*—the collective is able to solve problems that are beyond the information processing abilities of the individual. Swarm behavior has been extensively studied in artificial life research and led to numerous algorithms, from ant colony algorithms to find optimal paths to crowd simulations in movie animation technology. There are certainly other examples worth mentioning in this section but for the sake of brevity this text moves on to other issues that may be equally important for AI in the wider context of this paper.

3.3. Development of Standards. Standards facilitate many aspects of product quality such as correctness, reliability, reproducibility, robustness, understandability, interoperability, and maintainability. In a global setting, and in a heavy-demanding computing-permeated environment in particular, standards are an unavoidable necessity. There are several activities directed at the implementation of such standards for systems biology [51], and the *Systems Biology Markup Language* (SBML) (http://sbml.org/) is just one example [52]. SBML provides a machine-readable, XML-based format for describing models of biological processes. Projects in systems biology often involve genomic data from high-throughput experiments such as DNA microarrays [53]. The *Minimum Information About a Microarray Experiment* (MIAME) [54] is a standard that was developed for the exhaustive and unambiguous description of such experiments. Over 50 journals, including the *Nature* series, now require that any published study involving microarray data makes this data publicly available in a format compliant to the MIAME standard.

The development of standards could also address the terminological aspect. In multidisciplinary research environments, the same technical term can refer to very different things [55], adding to the already existing language barrier. For example, the term *sample* may bear a completely different meaning to a statistician and a biologist. While such a misunderstanding can be easily detected and resolved, problems arising from other terms may be more difficult to spot: test set versus validation set, supervised versus unsupervised, and so forth, [56].

Standards for describing computational models do exist, for example, the Predictive Model Markup Language (PMML), an XML-based schema that allows the detailed description of statistical models. In metabolomics research, efforts for establishing minimum standards for reporting data analysis have been undertaken [57]. However, in general, standards for computational and statistical models do not seem to be widely embraced yet, at least life science publications do not widely adhere to these standards, arguably because they are not enforced by journals and other publication sources—yet. There are many critical voices, though, calling for standards implementing *good statistical guidelines* [56]. It can be speculated that in the near

future journals will begin enforcing standardized reports of computational models, similarly to the MIAME standard for microarray data.

3.4. Adherence to Stringent Computational and Statistical Frameworks. Despite the pivotal role that computational and statistical models play in today's interdisciplinary research, their deployment (and the analysis workflow) are often only imprecisely and rudimentarily described. Problems can arise from uncertainties such as: How were the data preprocessed and normalized, precisely? Were outliers removed? If a data re-sampling strategy such as cross-validation was adopted, then how was this performed in detail? Which loss function was used and why? How were the model's parameters calibrated? Published reports of computational and statistical models generally leave many open questions, even in the context of data mining competitions such as the KDD cup (http://www.sigkdd.org/kddcup/). Of crucial importance in this context is an audit trail that logs the detailed analysis steps, allowing a reproducible study. This *bookkeeping*, albeit apparently simple, has been described as the *number one* problem in today's bioinformatics research [58].

Practitioners developing AI-based solutions for interdisciplinary life science research could perhaps play an active role in the specification of the aforementioned standards (even though it may be a bit of a stumbling block for some *old workhorses* in AI). For example, artificial neural networks have shown remarkable performance for classification and prediction tasks, and they are also an integral part of the toolbox for analyzing genomic data [59]. However, their intrinsic black-box character may represent a major impediment to their widespread use in the life sciences. Practitioners in AI must be aware that applications in these fields should, ideally, also have explanatory power, and must be embedded in statistical frameworks that account for experimental artifacts and biases, issues due to multiple hypotheses testing, and the notorious curse of dimensionality (i.e., many more features than samples, which presents a huge challenge for clustering, classification and rule extraction), to name but a few.

3.5. Adoption of Openness Principles. The definition and adherence of rigorous standards and frameworks is closely linked to the principle of *openness*. For software development, for instance, the primary goal of open source is the production of reliable software, its main technique is code sharing, and the main tool for achieving this is the internet. There is a mountain of evidence from many projects (Linux, Ruby on Rails, Python, etc.) about the quality demonstrated by such systems and there is a great belief in many places that the open source model may be superior in many ways to traditional (commercial, individual) development approaches that are happening in closed environments. Openness should not be confused with decentralization or disorganization. Quite the opposite is true and an organization such as the World Wide Web Consortium (W3C), the organization which oversees the standards for the World Wide Web, may hold as the best example for this mode of operation. The W3C is pioneering in its organization and mode of operation and provides a blueprint for how large-scale projects may benefit from the adoption of open and streamlined processes and procedures.

From an AI perspective, openness relates very well to many issues mentioned earlier in this section, and although it may be speculative to say so, there is a feeling that AI may benefit greatly if it decides to embrace some of the key openness concepts that are behind the successes demonstrated by organizations such as the W3C. The same is true for modern biology. (Note that although this section focuses somewhat on biology, the same is true for the field of neuroscience.) Recent breakthroughs in modern biology would arguably not have been possible without the support of highly efficient and freely available, non-patented bioinformatics software like the BLAST algorithm [60]. There is a mountain of evidence from many systems for the impact of open source projects (e.g., the R language and environment for statistical computing, the CellML markup language for the description of biological models, or MathML, which describes mathematical notation, capturing both its structure and content).

Further testimony to the road towards openness comes from the rising number of highly cited articles published in, for instance, the Hindawi, the BioMed Central and the Public Library of Science series embracing the open access policy [61]. These journals adopt a licensing scheme of the Creative Commons (http://creativecommons.org/), which the Neurocommons project (http://neurocommons.org/) is based on, too. This project develops an open source knowledge management platform for biological research, with a focus on neuroscience. These are just some examples of efforts towards increased openness in science, which will enhance the reproducibility, transparency, dissemination and ultimately the quality of research results. But the trend towards openness is not limited to science. Google, for example, is spearheading an openness initiative by allowing users to extract their data from its proprietary products (e.g., Gmail), thereby opening the door for users who wish to switch to another provider (http://www.dataliberation .org/).

In addition, an increasing number of biomedical journals now require that the experimental data are deposited in open repositories (such as Gene Expression Omnibus) prior to publication. While this is a laudable practice, it is not sufficient to warrant truly reproducible results. Ideally, open source principles would be adopted not only for the data, but also for the source code and the implemented analysis workflow(s) in order to allow truly reproducible experiments, and to resolve arising scientific debates [62]. While the need for publishing open source code has been pointed out for years, and many authors do provide it voluntarily in supplementary materials, most life science journals do not require this explicitly. It is tempting to speculate, however, that, in the foreseeable future, numerous life science journals will require the publication of open source code (which nonetheless can be copyrighted, of course), analogously to the publication of raw experimental data.

3.6. Integration of Data, Information, and Knowledge. Modern life science disciplines generate experimental data at an enormous scale and complexity. These disciplines are also characterized by an enormous body of background knowledge in the form of domain expert knowledge, scientific articles, data warehouses, or graphical representations of, for instance, biological pathways. No comparable knowledge infrastructure exists in *classic* domains for AI applications such as data mining for marketing and retail or stock market prediction. The intelligent integration of this wealth of data, information and knowledge presents unprecedented analytical and technological challenges. Intelligent grid computing for a massive parallel processing of this deluge of geographically distributed data could represent a further interesting challenge for AI-inspired methods [63]. Natural language processing and text mining of scientific articles may be the tool for digging out hidden gold nuggets of knowledge and for enriching life science data analysis [64]. Still, truly intelligent approaches for leveraging this domain knowledge are only in their infancy. Intelligent human-computer interfaces, aided by natural language processing and pattern recognition software offer further scope for approaches from AI in this context [9].

Advances of modern life science also go hand in glove with the development of increasingly sophisticated technological instruments. Gene expression profiling based on microarray technology, for example, generates thousands of measurements for one single biological specimen, entailing what is commonly known as the *curse of dimensionality* or *small-n-large-p problem*. This problem is exacerbated by an inherently high level of technical and biological noise, systematic errors (such as experimenter bias), and missing data. Conventional statistical methods that originate from *classic* data mining domains (marketing, retail, etc.) are not tailored for the idiosyncrasies of life science data. Here, innovations from machine learning have made significant contributions. But paradoxically, current high-throughput technologies generate both too many and too little data. Ideally, in order to *understand* system dynamics, we wish to measure multiple biological parameters for the same sample, at the same time and under the same experimental conditions. However, this is generally not feasible with the currently available, highly specialized instruments (e.g., common microarray platforms quantify only RNA abundance of a sample at a moment of time). Hence, time series experiments are necessary to assemble the *snapshots of time* into a full picture that allows the analysis of system dynamics. Still, biochemical experiments generally involve a *breaking-up* of the biological entity in order to measure the molecular components, whereas what we are really interested in are the dynamics of the complete, living entity in its natural environment. For example, primary tumors are composed of a multiple, genetically heterogeneous subpopulations of cells with different proclivity to metastasize, and the tumor microenvironment is a crucial determinant in the pathogenesis [65].

3.7. Computational Creativity. Creativity is a frequent element in the mythology, philosophy, or religion of many cultures and it is fair to say that it is one of these malleable concepts again that has fascinated mankind for centuries. (The term creativity (e.g., [66]) is similarly problematic in nature to the term intelligence. Among a manifold of definitions, [67] defines creativity as a cognitive process to generate novel or unconventional solutions. This cognitive process relies on two essential mechanisms: (i) divergent thinking, which generates original, new ideas and (ii) convergent thinking, which logically evaluates a variety of possible solutions to find the optimal one.) For instance, philosophers such as Pythagoras and the Pythagoreans contemplated beauty as an objective principle in beings which maintain harmony, order, and balance. (In this view, beauty could be the harmony witnessed in the cosmos, but also an expression of normal human behavior.) From beauty, however, it is only a small step to creativity—people admire the beauty of artifacts of various kinds but very often these artifacts are the product of a creative process undertaken by an artist. Creativity and beauty are not restricted to the liberal arts only. Many theories in science are considered to be the outcome of an equally creative process and people often mention the elegance or beauty of a theory. In the more recent mid-80s, for instance, science encountered a discourse with beauty and the creative forces in nature through chaos theory, the inspirational field of science that captured, among many other things, the dynamic of natural systems in images (called *fractals*) of astonishing beauty [68].

Creativity has strong ties to computing for some time. The goal of web design, for instance, is not to add to the functionality of an application but to make an application aesthetically pleasing and accessible to its users. (This does not mean that the production of code is a mundane task. On the contrary, many regard good coding as a highly creative activity.) As another example, take the field of humanoid robotics where the physical appearance of a robot, its gestures or its tone of voice may have an impact on user acceptance (e.g., in health care). Computational creativity, which is a relatively young field, relates to many of these issues. The field, which by its very nature is a multidisciplinary scientific endeavor again, carries the vision to better understand human creativity and to construct (via computers and intelligent algorithms) artifacts demonstrating human-level creativity or tools that are able to support the creative processes of humans. In relation to AI it is necessary to mention that creativity has been on the AI agenda for some time (e.g., [69]) and that it is a challenging question to ask whether AI could assist this creative process, or perhaps even emulate the human process of generating creative solutions? The BISON project is a recent initiative aiming at the design and implementation of a comprehensive computational realization of a bisociative information discovery framework with applications in systems biology (http://www.bisonet.eu/. *Bisociation* is a term [70] coined to denote a creative process involving "... *the sudden interlocking of two previously unrelated skills*, *or matrices of thought*."). One aspect of BISON is that in large, complex, and heterogeneous domains association techniques fail to discover relevant information that is not related

in obvious associative ways, in particular information that is related across different contexts. In reality, however, it is often the case that context/domain-crossing associations are needed in order to generate innovative domains. From this perspective it is an interesting idea to think about the threesome of AI, neuroscience and systems biology where novel information related to these areas is generated by some form of creative AI.

4. Summary

The motivation in this paper was to investigate and position the standing of AI in a modern science context, in particular a modern life science context with a focus on systems biology and neuroscience. An important finding in the paper is the fact that systems biology and neuroscience offer a fertile ground for approaches from AI and that it is fair to say that these fields are united by a great overlap in their defining dreams and visions. Research in these areas (and in life science in general) is characterized by an ever-increasing, complex data and knowledge proliferation, which presents unprecedented challenges on multiple levels, ranging from the acquisition of data up to its high-level interpretation and utilization. With the desire to model and understand complex dynamic systems, these disciplines therefore share a common goal, and they could undoubtedly benefit from a synergetic mutual exchange of ideas and discussion of problems. The paper mentioned that several methods from machine learning and data mining that have their roots in AI are now the backbone of data analysis in systems biology and neuroscience, but also that there may be an atmosphere of individualism that may stand against the creation of a stimulating synergistic environment that can be beneficial for practitioners working in these fields. The paper highlighted that some of the main challenges in such hugely interdisciplinary research environments are not of a technical nature but rather of a cultural nature. In order to overcome these challenges the paper suggested several measures (soft and hard) including, but not necessarily limited to: the ability to learn from other projects and to be inventive, to exploit novel computing paradigms and environments, to develop and adhere to stringent standards and frameworks, to be integrative, and to embrace openness principles. Taking a bird-eye view on AI, the paper believes that at this point in time AI is still a very lively and fascinating field, but also that it may find itself at a crossroads where it has to set its course intelligently and wisely in order to sustain its privileged standing as an inspiring and visionary modern discipline.

Acknowledgments

The first author gratefully acknowledges the support of the Japan Society for the Promotion of Science (JSPS fellowship no. P08625) and of Dr. Hiroaki Kitano, The Systems Biology Institute, Tokyo, Japan. The third author gratefully acknowledges the support of the Japan Society for the Promotion of Science (JSPS fellowship no. S09168) and wishes to express his thankfulness to Dr. Yoko Yamaguchi, laboratory head of Laboratory for Dynamics of Emergent Intelligence, RIKEN Brain Science Institute, Wako-shi, Saitama 351-0198, Japan.

References

[1] W. Dubitzky and F. Azuaje, "Preface," in *Artificial Intelligence Methods and Tools for Systems Biology*, W. Dubitzky and F. Azuaje, Eds., p. 221, Springer, Berlin, Germany, 2004.

[2] R. Pfeifer and C. Scheier, *Understanding Intelligence*, MIT Press, Cambridge, Mass, USA, 2000.

[3] M. Negnevitsky, *Artificial Intelligence: A Guide to Intelligent Systems*, Addison Wesley, Harlow, UK, 2nd edition, 2004.

[4] H. Brighton and H. Selina, *Introducing Artificial Intelligence*, Icon Books, 2003.

[5] K. Iagnemma and M. Buehler, "Editorial for Journal of Field Robotics—special issue on the DARPA grand challenge," *Journal of Field Robotics*, vol. 23, no. 9, pp. 655–656, 2006.

[6] H. Marmanis and D. Babenko, *Algorithms of the Intelligent Web*, Manning Publications, 2009.

[7] I. Millington, *Artificial Intelligence for Games*, Morgan Kaufmann, San Francisco, Calif, USA, 2006.

[8] S. Kelly and A. Schuster, "Application of a fuzzy controller on a lego mindstorms robot," in *Proceedings of the 1st International Conference on Automation, Control and Instrumentation (IADAT '05)*, pp. 200–203, Bilbao, Spain, February 2005.

[9] W. Duch and J. Mandziuk, "Quo vadis, computational intelligence?" in *Machine Intelligence: Quo Vadis? Advances in Fuzzy Systems Applications and Theory*, pp. 3–28, 2004.

[10] M. A. Lebedev and M. A. L. Nicolelis, "Brain-machine interfaces: past, present and future," *Trends in Neurosciences*, vol. 29, no. 9, pp. 536–546, 2006.

[11] D. Girimonte and D. Izzo, "Artificial intelligence for space applications," in *Intelligent Computing Everywhere*, A. Schuster, Ed., pp. 235–243, Springer, London, UK, 2007.

[12] D. Purves, G. J. Augustine, D. Fitzpatrick, et al., *Neuroscience*, Sinauer Associates, Sunderland, Mass, USA, 3rd edition, 2004.

[13] H. Kitano, "Systems biology: a brief overview," *Science*, vol. 295, no. 5560, pp. 1662–1664, 2002.

[14] E. S. Lander, L. M. Linton, B. Birren, et al., "Initial sequencing and analysis of the human genome," *Nature*, vol. 409, no. 6822, pp. 860–921, 2001.

[15] J. C. Venter, M. D. Adams, E. W. Myers, et al., "The sequence of the human genome," *Science*, vol. 291, no. 5507, pp. 1304–1351, 2001.

[16] S. Rasmussen, L. Chen, D. Deamer, et al., "Transitions from nonliving to living matter," *Science*, vol. 303, no. 5660, pp. 963–965, 2004.

[17] B. Holmes, "Alive!," *New Scientist*, vol. 185, no. 2486, pp. 28–33, 2005.

[18] C. Ainsworth, "The facts of life," *New Scientist*, vol. 178, no. 2397, pp. 28–31, 2003.

[19] A. Schuster, "DNA algorithms for rough set analysis," in *Proceedings of the 4th International Conference on Intelligent Data Engineering and Automated Learning (IDEAL '03)*, J. Liu, Y. M. Cheung, and H. Yin, Eds., vol. 2690 of *Lecture Notes in Computer Science*, pp. 498–513, Springer, Hong Kong, 2004.

[20] A. Schuster, "DNA algorithms for Petri net modeling," *Informatica*, vol. 32, no. 4, pp. 421–427, 2008.

[21] A. M. Turing, "The chemical basis of morphogenesis," *Philosophical Transactions of the Royal Society of London B*, vol. 237, no. 641, pp. 37–72, 1952.

[22] H. Kitano, "Towards a theory of biological robustness," *Molecular Systems Biology*, vol. 3, article 137, 2007.

[23] H. Kitano, "Cancer as a robust system: implications for anticancer therapy," *Nature Reviews Cancer*, vol. 4, no. 3, pp. 227–235, 2004.

[24] A. Schuster, Ed., *Robust Intelligent Systems*, Springer, London, UK, 2008.

[25] C. Adami, "Digital genetics: unravelling the genetic basis of evolution," *Nature Reviews Genetics*, vol. 7, no. 2, pp. 109–118, 2006.

[26] A. Schuster, "Robustness in nature as a design principle for artificial intelligence," in *Robust Intelligent Systems*, A. Schuster, Ed., pp. 173–197, Springer, London, UK, 2008.

[27] E. Voit and J. Schwacke, "Understanding through modeling: a historical perspective and review of biochemical systems theory as a powerful tool for systems biology," in *Handbook of Systems Biology: Principles, Methods, and Concepts*, A. Konopka, Ed., pp. 27–82, CRC Press, Boca Raton, Fla, USA, 2007.

[28] A. M. Turing, "Computing machinery and intelligence," *Mind*, vol. 59, pp. 433–460, 1950.

[29] D. Harel, "A Turing-like test for biological modeling," *Nature Biotechnology*, vol. 23, no. 4, pp. 495–496, 2005.

[30] D. Saha and A. Mukherjee, "Pervasive computing: a paradigm for the 21st century," *Computer*, vol. 36, no. 3, pp. 25–31, 2003.

[31] J. O. Kephart and D. M. Chess, "The vision of autonomic computing," *Computer*, vol. 36, no. 1, pp. 41–50, 2003.

[32] N. Sato and Y. Yamaguchi, "Computational model-based prediction of human episodic memory performance based on eye movements," *IEICE Transactions on Communications*, vol. 91B, no. 7, pp. 2142–2143, 2008.

[33] A.-L. Barabási and E. Bonabeau, "Scale-free networks," *Scientific American*, vol. 288, no. 5, pp. 60–69, 2003.

[34] S. M. Potter, "What can artificial intelligence get from neuroscience?" in *Artificial Intelligence Festschrift: The Next 50 Years*, M. Lungarella, J. Bongard, and R. Pfeifer, Eds., pp. 174–185, Springer, Berlin, Germany, 2007.

[35] M. Luck and R. Aylett, "Applying artificial intelligence to virtual reality: intelligent virtual environments," *Applied Artificial Intelligence*, vol. 14, no. 1, pp. 3–32, 2000.

[36] A. Huw Arnall, "Future technologies, today's choices—nanotechnology, artificial intelligence and robotics: a technical, political and institutional map of emerging technologies," Tech. Rep., Department of Environmental Science and Technology, Environmental Policy and Management Group, Faculty of Life Sciences, Imperial College London, University of London, London, UK, 2003.

[37] K. N. Sgarbas, "The road to quantum artificial intelligence," in *Current Trends in Informatics*, T. Papatheodorou, D. Christodoulakis, and N. Karanikolas, Eds., vol. A, pp. 469–477, New Technologies Publications, Athens, Greece, 2007.

[38] S. Kak, "Quantum mechanics and articial intelligence," in *Intelligent Computing Everywhere*, A. Schuster, Ed., pp. 81–101, Springer, London, UK, 2007.

[39] L. Rodriguez, H. Jiang, K. Stark, S. Bell, and D. Kortenkamp, "Validation of heuristic techniques for design of life support systems," in *Proceedings of the Workshop on Artificial Intelligence for Space Applications (IJCAI '07)*, Hyderabad, India, 2007.

[40] D. Izzo and L. Pettazzi, "Autonomous and distributed motion planning for satellite swarm," *Journal of Guidance, Control, and Dynamics*, vol. 30, no. 2, pp. 449–459, 2007.

[41] G. Paun, "Introduction to membrane computing," in *Applications of Membrane Computing*, G. Ciobanu, M. J. Perez-Jimenez, and G. Paun, Eds., Natural Computing Series, pp. 1–42, Springer, Berlin, Germany, 2006.

[42] M. A. Arbib, *The Handbook of Brain Theory and Neural Networks*, MIT Press, Cambridge, Mass, USA, 2nd edition, 2003.

[43] G. Santhanam, S. I. Ryu, B. M. Yu, A. Afshar, and K. V. Shenoy, "A high-performance brain-computer interface," *Nature*, vol. 442, no. 7099, pp. 195–198, 2006.

[44] L. R. Hochberg, M. D. Serruya, G. M. Friehs, et al., "Neuronal ensemble control of prosthetic devices by a human with tetraplegia," *Nature*, vol. 442, no. 7099, pp. 164–171, 2006.

[45] F. Roberta, "Reverse engineering the brain," *Biomedical Computation Review*, vol. 5, no. 2, pp. 10–17, 2009.

[46] J. S. Albus, G. A. Bekey, J. H. Holland, et al., "A proposal for a decade of the mind initiative," *Science*, vol. 317, no. 5843, p. 1321, 2007.

[47] D. J. Bakkum, P. M. Gamblen, G. Ben-Ary, Z. C. Chao, and S. M. Potter, "Meart: the semi-living artist," *Frontiers in NeuroRobotics*, vol. 1, no. 5, 2007.

[48] T. Saigusa, A. Tero, T. Nakagaki, and Y. Kuramoto, "Amoebae anticipate periodic events," *Physical Review Letters*, vol. 100, no. 1, Article ID 018101, 4 pages, 2008.

[49] I. Tagkopoulos, Y.-C. Liu, and S. Tavazoie, "Predictive behavior within microbial genetic networks," *Science*, vol. 320, no. 5881, pp. 1313–1317, 2008.

[50] M. B. Miller and B. L. Bassler, "Quorum sensing in bacteria," *Annual Review of Microbiology*, vol. 55, pp. 165–199, 2001.

[51] A. Brazma, M. Krestyaninova, and U. Sarkans, "Standards for systems biology," *Nature Reviews Genetics*, vol. 7, no. 8, pp. 593–605, 2006.

[52] M. Hucka, A. Finney, H. M. Sauro, et al., "The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models," *Bioinformatics*, vol. 19, no. 4, pp. 524–531, 2003.

[53] M. Schena, D. Shalon, R. W. Davis, and P. O. Brown, "Quantitative monitoring of gene expression patterns with a complementary DNA microarray," *Science*, vol. 270, no. 5235, pp. 467–470, 1995.

[54] A. Brazma, P. Hingamp, J. Quackenbush, et al., "Minimum information about a microarray experiment (MIAME)—toward standards for microarray data," *Nature Genetics*, vol. 29, no. 4, pp. 365–371, 2001.

[55] W. Dubitzky, M. Granzow, and D. Berrar, *Fundamentals of Data Mining in Genomics and Proteomics*, Springer, New York, NY, USA, 2007.

[56] M. A. Troester, R. C. Millikan, and C. M. Perou, "Microarrays and epidemiology: ensuring the impact and accessibility of research findings," *Cancer Epidemiology Biomarkers and Prevention*, vol. 18, no. 1, pp. 1–4, 2009.

[57] R. Goodacre, D. Broadhurst, A. K. Smilde, et al., "Proposed minimum reporting standards for data analysis in metabolomics," *Metabolomics*, vol. 3, no. 3, pp. 231–241, 2007.

[58] L. Savage, "Forensic bioinformatician aims to solve mysteries of biomarker studies," *Journal of the National Cancer Institute*, vol. 100, no. 14, pp. 983–987, 2008.

[59] J. Khan, J. S. Wei, M. Ringnér, et al., "Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks," *Nature Medicine*, vol. 7, no. 6, pp. 673–679, 2001.

[60] J. Quackenbush and S. L. Salzberg, "It is time to end the patenting of software," *Bioinformatics*, vol. 22, no. 12, pp. 1416–1417, 2006.

[61] G. Eysenbach, "Citation advantage of open access articles," *PLoS Biology*, vol. 4, no. 5, article e157, 2006.
[62] K. A. Baggerly, K. R. Coombes, and E. S. Neeley, "Run batch effects potentially compromise the usefulness of genomic signatures for ovarian cancer," *Journal of Clinical Oncology*, vol. 26, no. 7, pp. 1186–1187, 2008.
[63] Y. Gil, E. Deelman, J. Blythe, C. Kesselman, and H. Tangmunarunkit, "Artificial intelligence and grids: workflow planning and beyond," *IEEE Intelligent Systems*, vol. 19, no. 1, pp. 26–33, 2004.
[64] J. Natarajan, D. Berrar, W. Dubitzky, et al., "Text mining of full-text journal articles combined with gene expression analysis reveals a relationship between sphingosine-1-phosphate and invasiveness of a glioblastoma cell line," *BMC Bioinformatics*, vol. 7, article 373, 2006.
[65] I. J. Fidler, "The organ microenvironment and cancer metastasis," *Differentiation*, vol. 70, no. 9-10, pp. 498–505, 2002.
[66] I. Taylor, "The nature of the creative process," in *Creativity: An Examination of the Creative Process*, P. Smith, Ed., pp. 51–82, Hastlings House, New York, NY, USA, 1959.
[67] V. V. Kryssanov, H. Tamaki, and S. Kitamura, "Understanding design fundamentals: how synthesis and analysis drive creativity, resulting in emergence," *Artificial Intelligence in Engineering*, vol. 15, no. 4, pp. 329–342, 2001.
[68] B. B. Mandelbrot, *The Fractal Geometry of Nature*, W.H. Freeman, New York, NY, USA, 1982.
[69] M. A. Boden, "Creativity and artificial intelligence," *Artificial Intelligence*, vol. 103, no. 1-2, pp. 347–356, 1998.
[70] A. Koestler, *The Act of Creation*, Hutchinson, London, UK, 1964.

Hindawi Publishing Corporation
Advances in Artificial Intelligence
Volume 2010, Article ID 918062, 18 pages
doi:10.1155/2010/918062

Review Article

Where Artificial Intelligence and Neuroscience Meet: The Search for Grounded Architectures of Cognition

Frank van der Velde

Cognitive Psychology, Leiden University, Wassenaarseweg 52, 2333 AK Leiden, The Netherlands

Correspondence should be addressed to Frank van der Velde, vdvelde@fsw.leidenuniv.nl

Received 31 August 2009; Revised 11 November 2009; Accepted 12 December 2009

Academic Editor: Daniel Berrar

The collaboration between artificial intelligence and neuroscience can produce an understanding of the mechanisms in the brain that generate human cognition. This article reviews multidisciplinary research lines that could achieve this understanding. Artificial intelligence has an important role to play in research, because artificial intelligence focuses on the mechanisms that generate intelligence and cognition. Artificial intelligence can also benefit from studying the neural mechanisms of cognition, because this research can reveal important information about the nature of intelligence and cognition itself. I will illustrate this aspect by discussing the grounded nature of human cognition. Human cognition is perhaps unique because it combines grounded representations with computational productivity. I will illustrate that this combination requires specific neural architectures. Investigating and simulating these architectures can reveal how they are instantiated in the brain. The way these architectures implement cognitive processes could also provide answers to fundamental problems facing the study of cognition.

1. Introduction

Intelligence has been a topic of investigation for many centuries, dating back to the ancient Greek philosophers. But it is fair to say that it is a topic of a more scientific approach for just about 60 years. Crucial in this respect is the emergence of artificial intelligence (AI) in the mid 20th century. As the word "artificial" suggests, AI aimed and aims not only to understand intelligence but also to build intelligent devices. The latter aim adds something to the study of intelligence that was missing until then: a focus on the mechanisms that generate intelligence and cognition (here, I will make no distinction between these two concepts).

The focus on mechanisms touches upon the core of what intelligence and cognition are all about. Intelligence and cognition are about mechanisms. Only a true mechanistic process can transform a sensory impression into a motor action. Without it, cognition and intelligence would not have any survival value. This is quite clear for processes like pattern recognition or motor planning, but it also holds for "higher" forms of intelligence (cognition), like communication or planning. Consequently, a theory of a cognitive process that does not describe a true mechanism (one that, at least in principle, can be executed) is not a full theory of that process, but at best an introduction to a theory or a philosophical account.

In this respect, AI is not different from other sciences like physics, chemistry, astronomy, and genetics. Each of these sciences became successful because (and often when) they focussed on an understanding of the mechanisms underlying the phenomena and processes they study. Yet, the focus on mechanisms was not always shared by other sciences that study intelligence or cognition, like psychology or neuroscience. For the most part, psychology concerned (and still concerns) itself with a description of the behavior related to a particular cognitive process. Neuroscience, of course, studied and studies the physiology of neurons, which aims for a mechanistic understanding. Yet, for a long time it stopped short at a translation from physiology to cognition.

However, the emergence of cognitive neuroscience in the 1990s introduced a focus on a mechanistic account of natural intelligence within neuroscience and related sciences. Gazzaniga, one of the founders of cognitive neuroscience, makes this point explicitly: "At some point in the future,

cognitive neuroscience will be able to describe the algorithms that drive structural neural elements into the physiological activity that results in perception, cognition, and perhaps even consciousness. To reach this goal, the field has departed from the more limited aims of neuropsychology and basic neuroscience. Simple descriptions of clinical disorders are a beginning, as is understanding basic mechanisms of neural action. The future of the field, however, is in working toward a science that truly relates brain and cognition in a mechanistic way." [1, page xiii].

It is not difficult to see the relation with the aims of AI in this quote. Gazzaniga even explicitly refers to the description of "algorithms" as the basis for understanding how the brain produces cognition. Based on its close ties with computer science, AI has always described the mechanisms of intelligence in terms of algorithms. Here, I will discuss what the algorithms as intended by Gazzaniga and the algorithms aimed for by AI could have in common. I will argue that much can be gained by a close collaboration in developing these algorithms. In fact, a collaboration between cognitive neuroscience and AI may be necessary to understand human intelligence and cognition in full.

Before discussing this in more detail, I will first discuss why AI would be needed at all to study human cognition. After all, (cognitive) neuroscience studies the (human) brain, and so it could very well achieve this aim on its own. Clearly, (cognitive) neuroscience is crucial in this respect, but the difference between human and animal cognition does suggest that AI has a role to play as well (in combination with (cognitive) neuroscience. The next section discusses this point in more detail.

2. Animal versus Human Cognition

Many of the features of human cognition can be found in animals as well. These include perception, motor behavior and memory. But there are also substantial differences between human and animal cognition. Animals, primates included, do not engage in science (such as neuroscience or AI) or philosophy. These are unique human inventions. So are space travel, telescopes, universities, computers, the internet, football, fine cooking, piano playing, money, stock markets and the credit crisis, to name but a few.

And yet, we do these things with a brain that has many features in common with animal brains, in particular that of mammals. These similarities are even more striking in case of the neocortex, which is in particular involved in cognitive processing. In an extensive study of the cortex of the mouse, Braitenberg [2] and Braitenberg and Schüz [3] observed striking similarities between the cortex of the mouse and that of humans. In the words of Braitenberg [2, page 82]: "All the essential features of the cerebral cortex which impress us in the human neuroanatomy can be found in the mouse too, except of course for a difference in size by a factor 1000. It is a task requiring some experience to tell a histological section of the mouse cortex from a human one. . . . With electronmicrographs the task would actually be almost impossible."

It is hazardous to directly relate brain size to cognitive abilities. But the size of the neocortex is a different matter. There seems to be a direct relation between the size of the neocortex and cognitive abilities [4]. For example, the size of the human cortex is about four times that of chimpanzees, our closest relatives. This difference is not comparable to the difference in body size or weight between humans and chimpanzees.

So, somehow the unique features of human cognition are related to the features of the human cortex. How do we study this relation? Invasive animal studies have been extremely useful for understanding features of cognition shared by animals and humans. An example is visual perception. Animal research has provided a detailed account of the visual cortex as found in primates (e.g., macaques [5]). Based on that research, AI models of perception have emerged that excel in comparison to previous models [6]. Furthermore, neuroimaging research begins to relate the structure of the visual cortex as found in animals to that of humans [7].

So, in the case of visual perception we have the ideal combination of neuroscience and AI, producing a mechanistic account of perception. But what about the unique features of human cognition?

In invasive animal studies, electrodes can penetrate the cortex at arbitrary locations, the cortex can be lesioned at arbitrary locations, and the animal can be sacrificed to see the effects of these invasions. On occasion, electrodes can be used to study the human cortex, when it is invaded for medical reasons [8]. But the rigorous methods as used with animals are not available with humans. We can use neuroimaging, but the methods of neuroimaging are crude compared to the methods of animal research. EEG (electroencephalogram) provides good temporal resolution but its spatial resolution is poor. For fMRI (functional magnetic resonance imaging), the reverse holds. So, these methods on their own will not provide us with the detailed information provided by animal research.

This is in particular a problem for studying the parts of the human brain that produce space travel, telescopes, universities, computers, the internet, football, fine cooking, piano playing, money, stock markets and the credit crisis, if indeed there are such distinguishable parts. It is certainly a problem for studying the parts of the human brain that produce language and reasoning, which are at the basis of these unique human inventions. For these aspects of cognition, there is no animal model that we can use as a basis, as in the case of visual perception. (Indeed, if there were such animal models, that is, if animal cognition was on a par with human cognition, we would have to question the ethical foundations of doing this kind of research.)

So, not surprisingly, our knowledge of the neural mechanisms of language or reasoning is not comparable to that of visual perception. In fact, we do not have neural models that can account for even the basic aspects of language processing or reasoning.

In his book on the foundation of language, Jackendoff [9] summarized the most important problems, the "four challenges for cognitive neuroscience", that arise with a neural implementation of combinatorial structures, as found in

human cognition. These challenges illustrate the difficulties that occur when combinatorial hierarchical structures are implemented with neural structures. Consider the first two challenges analyzed by Jackendoff.

The first challenge concerns the massiveness of the binding problem as it occurs in language, for example in hierarchical sentence structures. For example, in the sentence *The little star is besides the big star*, there are bindings between adjectives and nouns (e.g, *little star* versus *big star*), but also bindings between the noun phrase *the little star* and the verb phrase *is besides the big star* or between the prepositional phrase *besides the big star* and verb *is*.

The second challenge concerns the problem of multiple instantiations, or the "problem of 2", that arises when the same neural structure occurs more than once in a combinatorial structure. For example, in the sentence *The little star is besides the big star*, the word *star* occurs twice, first as subject of the sentence and later as the noun of the prepositional phrase.

These challenges (and the other two) were not met by any neural model at the time of Jackendoff's book. For example, consider synfire chains [10]. A synfire chain can arise in a feedforward network when activity in one layer cascades to another layer in a synchronous manner. In a way, it is a neural assembly, as proposed by Hebb [11] with a temporal dimension added to it [3]. Synfire chains have sometimes been related to compositional processing [12], which is needed in the case of language.

But is clear that synfire chains do not meet the challenges discussed by Jackendoff. For example, in *The little star is besides the big star* a binding (or compositional representation) is needed for *little star* and *big star*, but not for *little big star* (this noun phrase is not a part of the sentence). With synfire chains (and Hebbian assemblies in general [13]), we would have synfire chains for *star*, *little* and *big*. The phrase *little star* would then consist of a binding (link) between the synfire chains for *little* and *star*. At the same time, the phrase *big star* would consist of a binding between the synfire chains for *big* and *star*. However, the combination of the bindings between the synfire chains for *little*, *big* and *star* would represent the phrase *little big star*, contrary to the structure of the sentence.

This example shows that synfire chains fail to account for the "problem of two". Because the word *star* occurs twice in the sentence, somehow these occurrences have to be distinguished. Yet, a neural representation of a concept or word, like *star*, is always the same representation (in this case the same synfire). Indeed, this is one of the important features of neural cognition, as I will argue below. But this form of conceptual representation precludes the use of direct links between synfire chains (or assemblies) as the basis for the compositional structures found in language (see [13] for a more extensive analysis).

3. Investigating the Neural Basis of Human Cognition

Given the additional difficulties involved in studying the neural basis of the specific human forms of cognition, as outlined above, the question arises how we can study the neural basis of human cognition.

Perhaps we should first study the basic aspects of neural processing, before we could even address this question. That is, the study of human forms of cognition would have to wait until we acquire more insight into the behavior of neurons and synapses, and smaller neural circuits and networks.

However, this bottom-up approach may not be the most fruitful one. First, because it confuses the nature of understanding with the way to achieve understanding. In the end, a story about the neural basis of human cognition would begin with neurons and synapses (or even genes) and would show how these components form neural circuits and networks, and how these structures produce complex forms of cognition. This is indeed the aim of understanding the neural basis of human cognition. But is not necessarily the description of the sequence in which this understanding should or even could be obtained.

A good example of this difference is found in the study of the material world. In the end, this story would begin with an understanding of elementary particles, how these particles combine to make atoms, how atoms combine to make molecules, how molecules combine to make fluids, gases and minerals, how these combine to make planets, how planets and stars combine to make solar systems, how these combine to make galaxies, and how galaxies combine to form the structure of the universe.

This may be the final aim of understanding the material world, but it is not the way in which this understanding is achieved. Physics and astronomy did not begin with elementary particles, or even atoms. In fact, they began with the study of the solar system. This study provided the first laws of physics (e.g., dynamics) which could then be used to study other aspects of the material world as well, such as the behavior of atoms and molecules. The lesson here is that new levels or organization produce new regularities of behavior, and these regularities can also provide information about the lower levels of organization. Understanding does not necessarily proceed from bottom to top, it can also proceed from top to bottom.

Perhaps the best way to achieve understanding is to combine bottom-up and top-down information. The discussion above about the foundations of language provides an example. We can study language (as we can study planets) and obtain valuable information about the structure of language. This information then sets the boundary conditions, such as the two challenges discussed above, that need to be fulfilled in a neural account of language structure. In fact, these boundary conditions provide information that may be difficult to come by in a pure bottom-up approach.

The study of the material world also provides information of how the interaction between the bottom-up and to-down approach might proceed. Astronomy studies objects (stars and galaxies) that are in a way inaccessible. That is we cannot visit them or study them in a laboratory setting. In a way, this resembles the study of the human brain, which is inaccessible in the sense that we cannot do the rigorous experiments as we do with animals.

Yet, astronomy has acquired a profound understanding of stars and galaxies. It can, for example, describe the evolution of stars even though that proceeds over millions of years. In the 19th century, however, astronomy was still restricted to describing the position of stars and their relative magnitude. But physics can study the properties of matter in a laboratory. Combined with theoretical understanding (e.g., quantum physics), it can show how light provides information about the structure of matter. This information can be used to study the properties of stars as well. Furthermore, theoretical understanding of matter (e.g., statistical physics) can also provide information about how stars could evolve, which in turn can be investigated with astronomical observations.

In short, the success of astronomy depends on a combination of studying the basics of matter (physics), observing the properties of stars (astronomy) and combining these levels with theoretical analysis. In this three-fold combination, each component depends on the other. As a result, seemingly inaccessible phenomena can be studied and understood on a substantial level of complexity.

A similar approach could be successful in studying the seemingly inaccessible neural basis of human cognition (as exemplified in language and reasoning). That is, detailed investigation of basic neural structures, observations of brain processes based on neuroimaging, and theoretical or computational research which investigates how cognitive processes as found in humans can be produced with neural structures and how the behavior of these structures can be related to observations based on neuroimaging. As in the case of astronomy, each of these components is necessary. But the role of AI will be restricted to the computational part. So, I will focus on that in the remainder of this paper.

4. Large-Scale Simulations

An important development in the collaboration between AI and neuroscience is the possibility of large-scale simulations of neural processes that generate intelligence. For example, the mouse cortex has approximately 8×10^6 neurons and 8000 synapses per neuron. Recently, an IBM research group represented 8×10^6 neurons and 6400 synapses per neuron on the IBM Blue Gene processor, and ran 1 s of model time in 10 s of real time [14]. With this kind of computing power, and its expected increase over the coming years, it can be expected that large sections of the human cortex (which is about 1000 times larger than the mouse cortex [3]) can be modelled in comparable detail in the near future.

These large-scale simulations will provide a virtual research tool by which characteristics of the human brain, and their relation to cognitive function, can be investigated on a scale and level of detail that is not hampered by the practical and ethical limitations of (invasive) brain research. For example, large-scale simulations can be used to study the interaction between thousands of neurons in realistic detail, or to investigate the effect of specific lesions on these interactions, or to investigate the role of specific neurotransmitters on neuronal interactions. In this way, the limitations of experimental methods can be augmented. No experimental method gives detailed information about the interaction of thousands of neurons, and no experimental method can vary parameters in the interaction at will to study their effect. The Blue Brain Project [15] is an attempt to study how the brain functions in this way, and to serve as a tool for neuroscientists and medical researchers.

But the Blue Brain Project is focused on creating a physiological simulation for biomedical applications. By its own admission, it is not (yet) an artificial intelligence project. However, from an AI perspective, large-scale simulations of neural processes can be used as a virtual laboratory to study the neural architectures that generate natural intelligence and cognition. These architectures depend on the structure of the brain, and the neocortex in particular, as outlined below.

4.1. Structure of the Neocortex. In the last decades, a wealth of knowledge has been acquired about the structure of the cortex (e.g., [16]). A comparison of the structure of the cortex in different mammals shows that the basic structure of the cortex in all mammals is remarkably uniform. The one factor that distinguishes the cortex of different mammals is their size. For example, the cortex of humans is about 1000 times that of a mouse, but at a detailed (microscopically) level it is very hard to distinguish the two [3]. This finding suggests that the unique features of human cognition might derive from the fact that more information can be processed, stored and interrelated in the extended networks and systems of networks as found in the human neocortex.

Furthermore, the basic structure of the cortex itself is highly regular. Everywhere within the cortex, neurons are organized in horizontal layers (i.e., parallel to the cortical surface) and in small vertical columns. The basic layered structure consists of six layers, which are organized in three groups: a middle layer (layer 4), the superficial layers (layers above layer 4) and the deep layers (layers below layer 4). The distribution of different kinds of neurons within the layers and columns is similar in all parts of the cortex. More than 70% of all neurons in the cortex are pyramidal neurons. Pyramidal neurons are excitatory, and they are the only neurons that form long-range connections in the cortex (i.e., outside their local environment). The probability that any two pyramidal neurons have more than two synaptic contacts with each other is small. Yet, substantially more than two synaptic inputs are needed to fire a pyramidal neuron. This indicates that neurons in the cortex operate in groups or populations. Furthermore, neurons within a given column in the cortex often have similar response characteristics, which also indicates that they operate as a group or population. In all parts of the cortex, similar basic cortical circuits are found. These circuits consist of interacting populations of neurons, which can be located in different layers.

At the highest level of organization, the cortex consists of different areas and connection structures ("pathways") in which these areas interact. Many pathways in the cortex are organized as a chain or hierarchy of cortical areas. Processing in these pathways initially proceeds in a feedforward manner, in which the lower areas in the hierarchy process input information first, and then transmit it to higher areas in

the hierarchy. However, almost all feedforward connections in the pathways of the cortex are matched by feedback connections, which initiate feedback processing in these pathways. The connection patterns in the pathways, consisting of feedforward, feedback and lateral connections, begin and terminate in specific layers. For example, feedforward connections terminate in layer 4, whereas feedback connections do not terminate in this layer.

An example of the relation between cortical structures and cognitive processing is given by visual perception. Processing visual information is a dominant form of processing in the brain. About 40% of the human cortex is devoted to it (in primates even more than 50%). The seemingly effortless ability to recognize shapes and colors, and to navigate in a complex environment is the result of a substantial effort on the part of the brain (cortex). The basic features of the visual system are known (e.g., [5]). The visual cortex consists of some 30 cortical areas, that are organized in different pathways. The different pathways process different forms of visual information, or "visual features", like shape, color, motion, or position in visual space.

All pathways originate from the primary visual cortex, which is the first area of the cortex to receive retinal information. Information is transmitted from the retina in a retinotopic (topographic) manner to the primary visual cortex. Each pathway consists of a chain or hierarchy of cortical areas, in which information is initially processed in a feedforward direction. The lower areas in each pathway represent visual information in a retinotopic manner. From the lower areas onwards, the pathways begin to diverge.

Object recognition (shape, color) in the visual cortex begins in the primary visual cortex, located in the occipital lobe. Processing then proceeds in a pathway that consists of a sequence of visual areas, going from the primary visual cortex to the temporal cortex. The pathway operates initially as a feedforward network (familiar objects are recognized fast, to the extent that there is little time for extensive feedforward-feedback interaction). Objects (shapes) can be recognized irrespective of their location in the visual field (i.e., relative to the point of fixation), and irrespective of their size.

Processing information about the spatial position of an object occurs in a number of pathways, depending on the output information produced in each pathway. For example, a specific pathway processes position information in eye-centered coordinates, to steer eye movements. Other pathways exist for processing position information in body-, head-, arm- or finger-centered coordinates. Each of these pathways consist of a sequence of visual areas, going from the primary visual cortex to the parietal cortex (and to the prefrontal cortex in the case of eye movements).

5. From Neural Mechanisms to Cognitive Architectures

Although several levels of organization can be distinguished in the brain, ranging from the cell level to systems of interacting neural networks, the neural mechanisms that fully account for the generation of cognition emerge at the level of neural networks and systems (or architectures) of these networks. A number of important issues can be distinguished here.

The structure of the cortex seems to suggest that the implementation of cognitive processes in the brain occurs with networks and systems of networks based on the uniform local structures (layers, columns, basic local circuits) as building blocks. The organization at the level of networks and systems of networks can be described as "architectures" that determine how specific cognitive processes are implemented, or indeed what these cognitive processes are.

Large-scale simulations of these architectures provide a unique way to investigate how specific architectures produce specific cognitive processes. In the simulation, the specific features of an architecture can be manipulated, to understand how they affect the cognitive process at hand. Furthermore, human cognition is characterized by certain unique features that are not found in animal cognition, or in a reduced form only (e.g., as in language, reasoning, planning). These features have to be accounted for in the analysis of the neural architectures that implement human cognitive processes. An interesting characteristic of these architectures is that they would consist of the same kind of building blocks and cortical structures as found in all mammalian brains. Investigating the computational features of these building blocks provides important information for understanding these architectures.

Because the cortex consists of arrays of columns, containing microcircuits, the understanding of local cortical circuits is a prerequisite for understanding the global stability of a highly recurrent and excitatory network as the cortex. An important issue here is whether the computational characteristics of these microcircuits can be characterized by a relatively small number of parameters [17]. A small number of parameters which are essential for the function of local circuits, as opposed to the large number of neural and network parameters, would significantly reduce the burden of simulating large numbers of these circuits, as required for the large-scale simulation of cognitive processes. It would also emphasize the uniform nature of columns as building blocks of the cortex.

Another important issue concerns the computational characteristics of the interaction between feedforward and feedback networks in the cortex. Connections in the feedforward direction originate for the most part in the superficial layers and sometimes in the deep layers, and they terminate in the middle layer (layer 4) of the next area. Within that area, the transformation from input activity (layer 4) to output activity (superficial or deep layers) occurs in the local cortical circuits (as found in the columns) that connect the neural populations in the different layers. Feedback processing starts in the higher areas in a hierarchy and proceeds to the lower areas. Feedback connections originate and terminate in the superficial and deep layers of the cortex.

So, it seems that feedforward activity carries information derived from the outside world (bottom up information), whereas feedback activity is more related to expectations generated at higher areas within an architecture (top-down

expectations). The difference between the role of feedforward activation and that of feedback activation is emphasized by the fact that they initially activate different layers in the cortex. In particular, feedback activation terminates in the layers that also produce the input for feedforward activity in the next area. This suggest that feedback activity (top-down expectation) modulates the bottom-up information as carried by feedforward activity. It is clear that this modulation occurs in the microcircuits (columns) that interconnect the different layers of the cortex, which again emphasizes the role of these circuits and illustrates the interrelation between the different computational features of the cortex.

The large-scale simulation of cortical mechanisms works very well when there is a match between the knowledge of a cortical architecture and the cognitive processes it generates, as in the case of the visual cortex. For example, the object recognition model of Serre et al. is based on cortex-like mechanisms [6]. It shows good performance, which illustrates the usefulness of cortical mechanisms for AI purposes. Also, the model is based on neural networks which could be implemented in parallel hardware, which would increase their processing speed. Moreover, the weight and energy consumption of devices based on direct parallel implementation of networks would be less than that of standard computers, which enhances the usefulness of these models in mobile systems.

So, when a cortical architecture of a cognitive process is (relatively) well known, as in the visual cortex, one could say that AI follows the lead of (cognitive) neuroscience. But not all cortical architectures of cognition are as well known as the visual cortex. Knowledge of the visual cortex derives to a large extent from detailed animal experiments. Because these experiments are not available for cognitive processes that are more typically human, such as language and reasoning, detailed information about their cortical mechanisms is missing.

Given the uniform structure of the cortex, we can make the assumption that the cortical architectures for these cognitive processes are based on the cortical building blocks as described above. But additional information is needed to unravel these cortical architectures. It can be found in the nature of the cognitive processes they implement. Because specific neural architectures in the cortex implement specific cognitive processes, the characteristics of these processes provide information about their underlying neural mechanisms. In particular, the specific features of human cognition have to be accounted for in the analysis and modelling of the neural architectures involved. Therefore, the analysis of these features provides important information about the neural architectures instantiated in the brain.

6. From Cognitive Architectures to Neural Mechanisms

AI might take the lead in the analysis of mechanisms that can generate features of human cognition. So, AI could provide important information about the neural architectures instantiated in the brain when the mechanisms it provides are combined with knowledge of cortical mechanisms. A number of features of (human) cognition can be distinguished where insight in cognitive mechanisms is important to understand the cortical architectures involved.

6.1. Parallel versus Sequential Processing. A cognitive neural architecture can be characterized by the way it processes information. A main division is that between parallel processing of spatially ordered information and processing of sequentially ordered information.

Parallel processing of spatially ordered information is found in visual perception. An important topic in this respect is the location and size invariant identification of objects in parallel distributed networks. How this can be achieved in a feedforward network is not yet fully understood, even though important progress has been made for object recognition (e.g., [6]). An understanding of this ability is important, because visual processing is a part of many cognitive tasks. However, understanding the computational mechanisms of location and size invariant processing in the brain is also important in its own right, given the applications that could follow from this understanding.

Sequentially ordered information is found in almost all forms of cognitive processing. In visual perception, for example, a fixation of the eyes lasts for about 200 ms. Then a new fixation occurs, which brings another part of the environment in the focal field of vision. In this way, the environment is explored in a sequence of fixations. Other forms of sequential processing occur in auditory perception and language processing. Motor behavior also has clear sequential features. The way in which sequentially ordered information can be represented, processed and produced in neural architectures is just beginning to be understood [13]. Given its importance for understanding neurocognition, this is an important topic for further research.

6.2. Representation. Many forms of representation in the brain are determined by a frame of reference. On the input side, the frame of reference is based on the sensory modality involved. For example, the initial frame of reference in visual perception is retinotopic. That is, in the early (or lower) areas of the visual cortex, information is represented topographically, in relation with the stimulation on the retina. On the output side, the frame of reference is determined by the body parts that are involved in the execution of a movement. For example, eye positions and eye movements are represented in eye-centered coordinates. Thus, to move the eyes to a visual target, the location of the target in space has to be represented in eye-centered coordinates. Other examples of (different) "motor representations" are head-, body-, arm-, or finger-centered coordinates. The nature of these representations and the transformations between their frames of reference have to be understood. Three important issues can be distinguished in particular.

The first one concerns the nature of feedforward transformations. When sensory information is used to guide an action, sensory representations are transformed into motor representations. For example, to grasp an object with visual

guidance, the visual information about its location has to be transformed into the motor representations needed to grasp the object. In this case, the transformations to the motor representations start from a retinotopic representation. The question is what the different forms of motor representation are, how the neural transformations between retinotopic representation and these different motor representations proceed, and how they are learned.

The second one concerns the integration of motor systems. An action often involves the movement of different body parts. The question is how these different motion systems are integrated. That is, how are the transformations between different motor representations performed, and how are they learned. In particular, the question is whether the transformations between motor systems are direct (e.g., from head to body representation and vice versa), or whether they proceed through a common intermediary representation. Suggestions have been made that eye-centered coordinates function as such an intermediary representation (lingua franca). In this way, one motor representation is first transformed into eye-centered coordinates before it is transformed into another motor transformation. An answer to this question is also of relevance for visual motor guidance (e.g., the effect of visual attention on action preparation, [18]).

The third one concerns the effect of feedback transformations. These transformations concern the effect of motor planning on sensory (e.g., visual) processing. For example, due to an eye shift a new part of the visual space is projected on a given location of the retina, replacing the previous projection. In physical terms, there is no difference between a new projection on the retina produced by the onset of a new stimulus (i.e., a stimulus not yet present in the visual field), or a new projection on the same retinal location produced by a stimulus (already present in the visual field) due to an eye shift. In both cases, there is an onset of a stimulus on the given retinal location. However, at least some neurons in the visual cortex respond differently to these two situations. The difference is most likely due to the effect of motor planning and motor execution on the visual representation. In case of an eye shift, information is available that a new stimulus will be projected on a given retinal location. This information is absent in the case of a direct stimulus onset (i.e., the onset of a stimulus not yet present in the visual field). Through a feedback transformation, the motor representation related to the eye shift can be transformed into a retinotopic representation, which can influence the representation of the new visual information. The stability of visual space is related to these feedback transformations. Because the body, head and eyes are moving continuously, the retinal projections also fluctuate continuously due to these movements. Yet, the visual space is perceived as stable. Visual stability thus results from an integration of visual and motor information.

6.3. Productivity. A fundamental feature of human cognition is the practically unlimited productivity of human cognition. Cognitive productivity concerns the ability to process or produce a virtually unlimited number of cognitive structures in a given cognitive domain. For example, a virtually unlimited number of novel sentences can (potentially) be understood or produced by a normal language user. Likewise, visual perception provides the ability to navigate in a virtually unlimited number of novel visual scenes (e.g., novel environments like unknown cities).

In the case of visual perception, productivity is found in animals as well. But with language and reasoning, productivity is uniquely human. A conservative estimate shows that humans can understand a set of 10^{20} (meaningful) sentences or more [19, 20]. This kind of productivity is unlimited in any practical sense of the word. For example, the estimated lifetime of the universe is in the order of 10^{17} to 10^{18} seconds. This number excludes that we could learn each sentence in the set of 10^{20}. Instead, we can understand and produce sentences from this set only in a productive manner.

In computational terms, productivity results from the ability to process information in a combinatorial manner. In combinatorial processing, a cognitive structure (e.g., sentence, visual scene) is processed in terms of its components (or constituents) and the relations between the components that determine the overall structure. Sentences are processed in terms of words and grammatical relations. Visual scenes are processed in terms of visual features like shapes, colors, (relative) locations, and the binding relations between these features.

To understand the neural basis of human cognition, it is essential to understand how combinatorial processing is implemented in neural systems as found in the cortex. A recently proposed hypothesis is that all forms of combinatorial processing in neural systems depend on a specific kind of neural architectures [13]. These architectures can be referred to as neural "blackboard" architectures. They consist of specialized networks that interact through a common neural blackboard.

An example is found in the visual cortex. Visual features like shape, color, motion, position in visual space, are processed and identified in specialized (feedforward) networks. Through feedback processing and interaction in the lower retinotopic areas of the visual cortex, these specialized networks can interact. In this way, the (binding) relations between the visual features of an object can be established [18]. The structure of the neural blackboard architecture for vision is determined by the kind of information it processes, in particular the fact that visual information is (initially) spatially ordered. The characteristics of visual (spatial) information thus provide information about the structure of the neural architecture for vision.

In a similar way, the characteristics of sequentially ordered information, for example, as found in language, or reasoning, or motor planning, and so forth, can be used to determine the structure of the neural architectures involved in these forms of processing. Because combinatorial processing imposes fundamental constraints on neural architectures, these constraints can be used to generate hypotheses about the underlying brain structures and dynamics. In particular, when they are combined with the nature of conceptual representation, as discussed in the next section.

7. Grounded Architectures of Cognition

A potential lead of AI in analyzing the mechanisms of cognition is perhaps most prominent with cognitive processes for which no realistic animal model exists. Examples are language, detailed planning and reasoning. A fascinating characteristic of these processes is that they are most likely produced with the same cortical building blocks as described earlier, that is, the cortical building blocks that also produce cognitive processes shared by humans and animals, such as visual perception and motor behavior.

Apparently, the size of the neocortex plays a crucial role here. The human cortex is about four times the size of that of a chimpanzee, 16 times that of a macaque monkey and a 1000 times that of a mouse [3, 4]. Given the similarity of the structure of the cortex, both within the cortex and between cortices of different mammals this relation between size and ability makes sense. Having more of the same basic cortical mechanisms available will make it easier to store more information, but apparently it also provides the ability to recombine information in new ways.

Recombining information is what productivity is about. So, we can expect these more exclusively human forms of cognition to be productive. But the way information is stored should be comparable with the way information is stored in the brain in all forms of cognition. Examples are the forms of representation found in the visual cortex or the motor cortex, as discussed above. This is a challenge for AI and cognitive science: how to combine productivity as found in human cognition with the forms of representation found in the brain. Solving this challenge can provide important information about how these forms of cognition are implemented in the brain. It can also provide information about the unique abilities of human cognition which can be used to enhance the abilities of AI.

To understand the challenge faced by combining cognitive productivity with representation as found in the brain, consider the way productivity is achieved in the classical theory of cognition, or classical cognitivism for short, that arose in the 1960s. Classical cognitive architectures (e.g., [21, 22]) achieve productivity because they use symbol manipulation to process or create compositional (or combinatorial) structures.

Symbol manipulation depends on the ability to make copies of symbols and to transport them to other locations. As described by Newell [22, page 74]: "The symbol token is the device in the medium that determines where to go outside the local region to obtain more structure. The process has two phases: first, the opening of access to the distal structure that is needed; and second, the retrieval (transport) of that structure from its distal location to the local site, so it can actually affect the processing. (...) Thus, when processing "The cat is on the mat" (which is itself a physical structure of some sort) the local computation at some point encounters "cat"; it must go from "cat" to a body of (encoded) knowledge associated with "cat" and bring back something that represents that a cat is being referred to, that the word "cat" is a noun (and perhaps other possibilities), and so on."

Symbols can be used to access and retrieve information because they can be copied and transported. In the same way, symbols can be used to create combinatorial structures. In fact, making combinatorial structures with symbols is easy. This is why symbolic architectures excel in storing, processing and transporting huge amounts of information, ranging from tax returns to computer games. The capacity of symbolic architectures to store (represent) and process these forms of information far exceeds that of humans.

But interpreting information in a way that could produce meaningful answers or purposive actions is far more difficult with symbolic architectures. In part, this is due to the ungrounded nature of symbols. The ungrounded nature of symbols is a direct consequence of using symbols to access and retrieve information, as described by Newell. When a symbol token is copied and transported from one location to another, all its relations and associations at the first location are lost. For example, the perceptual information related to the concept *cat* is lost when the symbol token for *cat* is copied and transported to a new location outside the location where perceptual information is processed. At the new location, the perceptual information related to cats is not directly available. Indeed, as Newell noted, symbols are used to escape the limited information that can be stored at one site. So, when a symbol is used to transport information to other locations, at least some of the information at the original site is not transported.

The ungrounded nature of symbol tokens has consequences for processing. Because different kinds of information related to a concept are stored and processed at different locations, they can be related to each other only by an active decision to gain access to other locations, to retrieve the information needed. This raises the question of who (or what) in the architecture makes these decisions, and on the basis of what information. Furthermore, given that it takes time to search and retrieve information, there are limits on the amount of information that can be retrieved and the frequency with which information can be renewed.

So, when a symbol needs to be interpreted, not all of its semantic information is directly available, and the process to obtain that information is very time consuming. And this process needs to be initiated by some cognitive agent. Furthermore, implicit information related to concepts (e.g., patterns of motor behavior) cannot be transported to other sites in the architecture.

7.1. Grounded Representations. In contrast to symbolic representations, conceptual representations in human cognition are grounded in experiences (perception, action, emotion) and (conceptual) relations (e.g., [23, 24]). The forms of representation discussed in Section 6.2 are all grounded in this way. For example, grounding of visual representations begins with the retinotopic (topographic) representations in the early visual cortex. Likewise, motor representations are grounded because they are based on the frame of reference determined by the body parts that are involved in the execution of a movement. An arbitrary symbol is not grounded in this way.

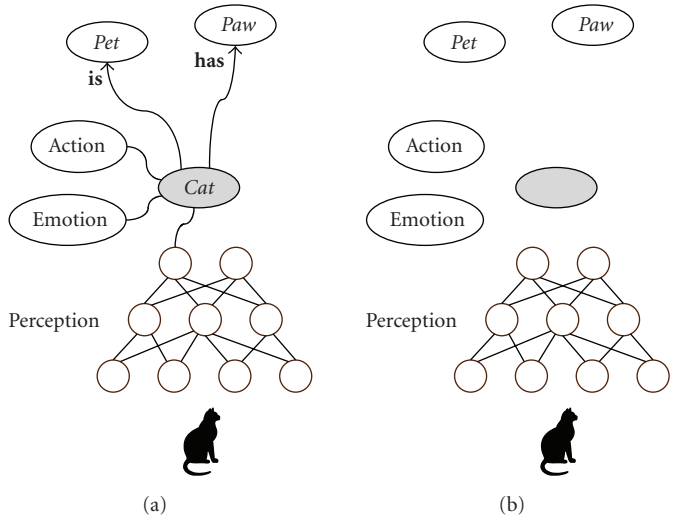


Figure 1: (a) illustration of the grounded structure of the concept *cat*. The circles and ovals represent populations of neurons. The central population labeled *cat* can be used to bind the grounded representation to combinatorial structures. (b) without the overall connection structure, the central population no longer forms a representation of the concept *cat*.

The consequence of grounding, however, is that representations cannot be copied and transported elsewhere. Instead, they consists of a network structure distributed over the cortex (and other brain areas). An illustration is given in Figure 1, which illustrates the grounded structure of the concept *cat*.

The grounded representation of *cat* interconnects all features related to cats. It interconnects all perceptual information about cats with action processes related to cats (e.g., the embodied experience of stroking a cat, or the ability to pronounce the word *cat*), and emotional content associated with cats. Other information associated or related to cats is also included in the grounded representation, such as the (negative) association between cats and dogs and the semantic information that a cat is a pet or has paws.

It is clear that a representation of this kind develops over time. It is in fact the grounded nature of the representation that allows this to happen. For example, the network labeled "perception" indicates that networks located in the visual cortex learn to identify cats or learn to categorize them as animals. In the process of learning to identify or categorize cats they will modify their connection structure, by growing new connections or synapses or by changing the synaptic efficacies. Other networks will be located in the auditory cortex, or in the motor cortex or in parts of the brain related to emotions. For these networks as well, learning about cats results in a modified network structure. Precisely because these networks remain located in their respective parts of the cortex, learning can be a gradual and continuous process. Moreover, even though these networks are located in different brain areas, connections can develop over time between them because their positions relative to each other remain stable as well.

The grounded network structure for *cat* illustrates why grounded concepts are different from symbols. There is no well designated neural structure like a symbol that can be copied or transported. When the conceptual representation of *cat* is embodied in a network structure as illustrated in Figure 1, it is difficult to see what should be copied to represent *cat* in sentences like these.

For example, the grey oval in Figure 1, labeled *cat*, plays an important role in the grounded representation of the concept *cat*. It represents a central neural population that interconnects the neural structures that represent and process information related to cats. However, it would be wrong to see this central neural population itself as a neural representation of *cat* that could be copied and transported like a symbol. As Figure 1 (b) illustrates, the representational value of the central neural population labeled *cat* derives entirely from the network structure of which it is a part. When the connections between this central neural population and the other networks and neural populations in the structure of *cat* are disrupted, the central neural population no longer constitutes a representation of the concept *cat*. For example, because it is no longer activated by the perceptual networks that identify cats. So, when the internal network structure of the central neural population (or its pattern of activation) is copied and transported, the copy of the central neural population is separated from the network structure that represents *cat*. In this way, it has lost its grounding in perception, emotion, action, associations and relations.

7.2. Grounded Representations and Productivity. Making combinatorial structures with symbols is easy. All that is required is to make copies of the symbols (e.g., words) needed and to paste them into the combinatorial structure as required. This, of course, is the way how computers operate and how they are very successful in storing and processing large amounts of data. But as noted above, semantic interpretation is much more difficult in this way, as is the binding with more implicit forms of information

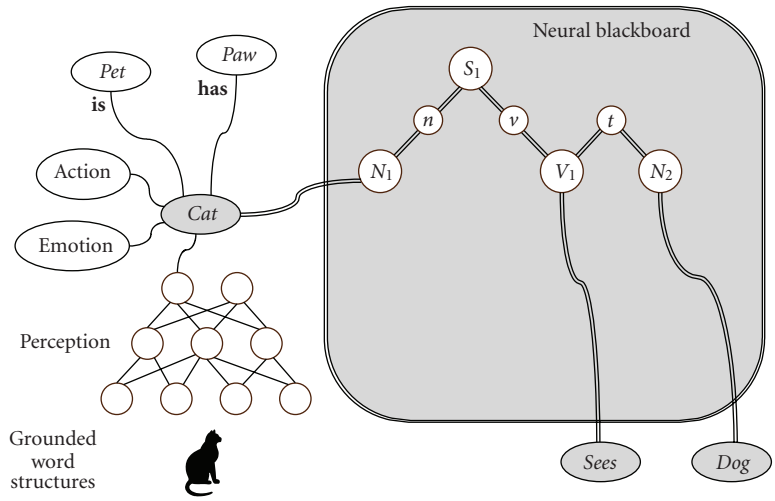


Figure 2: Illustration of the combinatorial structure *The cat sees the dog* (ignoring *the*), with grounded representations for the words. The circles in the neural blackboard represent populations and circuits of neurons. The double line connections represent conditional connections. (N, n = noun; S = sentence; t = theme; V, v = verb.)

storing found in embodied cognition. Yet, grounding representations and at the same time providing the ability to create novel combinatorial structures with these representations is a challenge, which the human brain seems to have solved.

At face value, there seems to be a tension between the grounded nature of human cognition and its productivity. The grounded nature of cognition depends on structures as illustrated in Figure 1. At a given moment, they consist of a fixed network structure distributed over one or more brain areas (depending on the nature of the concept). Over time, they can be modified by learning or development, but during any specific instance of information processing they remain stable and fixed.

But productivity requires that new combinatorial structures can be created and processed on the fly. For, as noted above, humans can understand and (potentially) produce in the order of 10^{20} (meaningful) sentences or more. Because this numbers exceeds the lifetime of the universe in seconds, it precludes that these sentences are somehow encoded in the brain by learning or genetic coding. Thus, most of the sentences humans can understand are novel combinatorial structures (based on familiar words), never heard or seen before. The ability to create or process these novel combinatorial structures was a main motivation for the claim that human cognition depends on symbolic architectures (e.g., [25]).

Figure 2 illustrates that grounded representations of the words *cat*, *sees* and *dog* can be used to create a combinatorial (compositional) structure of the sentence *The cat sees the dog* (ignoring *the*). The structure is created by forming temporal interconnections between the grounded representations of *cat*, *sees*, and *dog* in a "neural blackboard architecture" for sentence structure [13]. The neural blackboard consists of neural structures that represent syntactical type information (or "structure assemblies") such as structure assemblies for sentence (S_1), noun phrase (here, N_1 and N_2) and verb phrase (V_1). In the process of creating a sentence structure, the structure assemblies are temporarily connected (bound) to word structures of the same syntactical type. For example, *cat* and *dog* are bound to the noun phrase structure assemblies N_1 and N_2, respectively. In turn, the structure assemblies are temporarily bound to each other, in accordance with the sentence structure. So, *cat* is bound to N_1, which is bound to S_1 as the subject of the sentence, and *sees* is bound to V_1, which is bound to S_1 as the main verb of the sentence. Furthermore, *dog* is bound to N_2, which is bound to V_1 as its theme (object).

Figure 3 illustrates the neural structures involved in the representation of the sentence *cat sees dog* in more detail. To simplify matters, I have used the basic sentence structure in which the noun *cat* is connected directly to the verb *sees* as its agent. This kind of sentence structure is characteristic of a protolanguage [26], which later on develops into the more elaborate structure illustrated in Figure 2 (here, *cat* is the subject of the sentence, instead of just the agent of *sees*).

Figure 3(a) illustrates the structure of *cat sees dog*. The ovals are the grounded word structures, as in Figure 2. They are connected to their structure assemblies with memory circuits. The structure assemblies have an internal structure. For example, a noun phrase structure consists of a main part (e.g., N_1) and subparts, such as a part for agent (a) and one for theme (t). Subparts are connected to their main parts by gating circuits. In turn, similar subparts (or "subassemblies") of different structure assemblies are connected to each other by memory circuits. In this way, N_1 and V_1 are connected with their agent subassemblies and V_1 and N_2 are connected with their theme subassemblies. This represents that *cat* is the agent of *sees* and *dog* is its theme.

The structure assemblies (main parts and subparts alike) consists of pools or "populations" of neurons. So, each circle in Figure 3 represents a population. The neurons in a population are strongly interconnected, which entails that a population behaves as a unity, and its behavior can be modeled with population dynamics [13]. Furthermore,

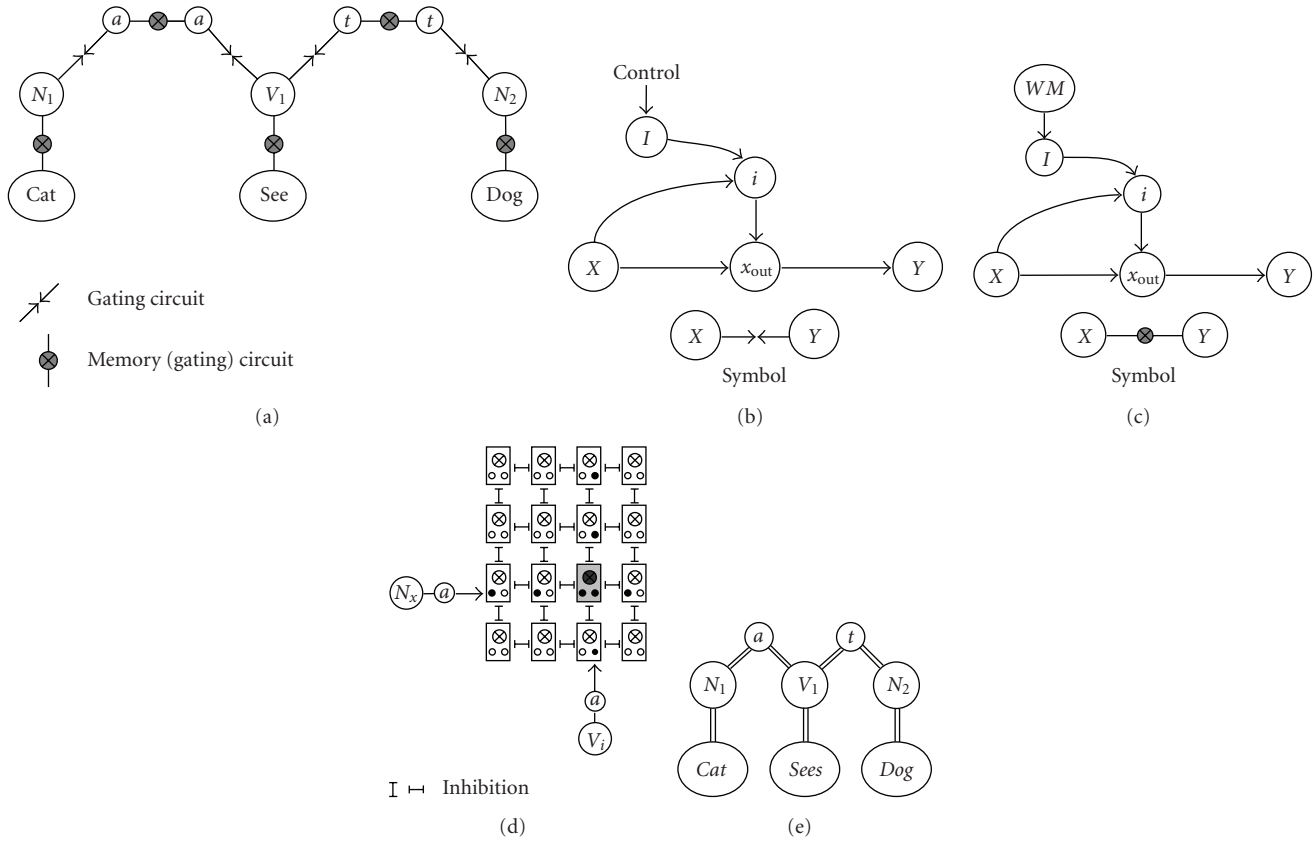


Figure 3: Illustration of the detailed neural structures involved in a sentence representation as illustrated in Figure 2. Ovals represent grounded word structures. The oval *WM* represents a working memory population, that remains active for a while after being activated. Circles represent populations of neurons. *I* and *i* are inhibitory neuron populations. The other ones are excitatory populations. (*a* = agent; *N* = noun; *t* = theme; *V* = verb.)

a population can retain activation for a while, due to the reverberation of activity within the population [27].

Figure 3(b) illustrates a gating circuit between two populations (*X* and *Y*). It consists of a disinhibition circuit. Activation can flow from *X* to *Y* when a control circuit activates population *I*, which in turn inhibits population *i*. The combination of gating circuits from *X* to *Y* and from *Y* to *X* is represented by the symbol illustrated in Figure 3(b). Gating circuits provide control of activation. They prevent that interconnected word structures form an associative structure, in which all word structures become automatically activated when one of them is active. Instead, activation from one word structure to another depends on specific control signals that activate specific gating circuits. In this way, information can be stored and retrieved in a precise manner. For example, the architecture can answer the question "What does the cat see?" or "Who sees the dog?" in this way [13].

Figure 3(c) illustrates a memory circuit between two populations (*X* and *Y*). It consists of a gating circuit that is activated by a working memory (*WM*) population. The *WM* population is activated when *X* and *Y* have been activated simultaneously (using another circuit not shown here [13]). So, the *WM* population stores the "memory" that *X* and *Y* have been activated simultaneously. Activation in the *WM* population consists of reverberating (or "delay") activity, which remains active for a while [27]. The combination of memory circuits from *X* to *Y* and from *Y* to *X* is represented by the symbol illustrated in Figure 3(c). When the *WM* population is active, activation can flow between *X* and *Y*. In this way, *X* and *Y* are "bound" into one population. Binding lasts as long as the *WM* population is active.

Bindings in the architecture are between subassemblies of the same kind (this is, in fact, also the case for the bindings between word assemblies and structures assemblies, although these subassemblies are ignored here). Figure 3(d) shows the connection matrix for binding between the agent subassemblies of noun phrase and verb phrase structure assemblies. All other subassembly bindings depend on a similar connection matrix. Arbitrary noun phrase and verb phrase structure assemblies can bind in this way. Binding occurs in a "neural column" that interconnects their respective subassemblies (agent subassemblies in this case). The neural column consists of the memory circuits needed for binding (and the circuit that activate the *WM* population). Neural columns for the same noun phrase or verb phrase structure assembly inhibit each other, which ensures that a noun phrase can bind to only one verb phrase structure assembly (and vice versa) with the same subassembly.

Figure 3(e) illustrates a "shorthand" representation of the entire connection structure of the sentence *cat sees dog* illustrated in Figure 3. When subassemblies are bound by memory circuits, they effectively merge into one population, so they are represented as one. The gating circuits, and the memory circuits between word and structure assemblies, are represented by double lines. The structure as represented in Figure 3(e) in fact consists of more than 100 populations, consisting of the populations that represent the structure assemblies and the populations found in the gating and memory circuits. To "see" these populations, one would have to "unwrap" the shorthand representation, inserting the connection matrices, gating and memory circuits and structure assemblies involved.

In the remainder of the paper, I will use the shorthand notion, as I have done in Figure 2. But the full structure is always implied, consisting of over 100 populations (substantially more for more complex sentences). So, for example, the circle labeled "*n*" in Figure 2 represents the "noun" subassemblies of the N_1 and S_1 structure assemblies, and the memory circuit that connects them. In this way, N_1 is bound to S_1 as its subject. Likewise, S_1 and V_1 are connected with their "verb" (*v*) subassemblies.

All bindings in this architecture are of a temporal nature. Binding is a dynamic process that activates specific connections in the architecture. The syntax populations (structure assemblies) play a crucial role in this process, because they allow these connections to be formed. For example, each word structure corresponding to a noun has connections to each noun phrase population in the architecture. However, as noted, these connections are not just associative connections, due to the neural (gating) circuits that control the flow of activation through the connection.

To make a connection active, its control circuit has to be activated. This is an essential feature of the architecture, because it provides control of activation, which is not possible in a purely associative connection structure. In this way, relations instead of just associations can be represented. Figure 1 also illustrates an example of relations. They consist of the conditional connections between the word structure of *cat* and the word structures of *pet* and *paw*. For example, the connection between *cat* and *pet* is conditional because it consists of a circuit that can be activated by a query of the form *cat is*. The **is** part of this query activates the circuit connection between *cat* and *pet*, so that *pet* is activated as the answer to the query. Thus, in conditional connections the control of activation can be controlled. For example, the **is** and **has** labels in Figure 1 indicate that information of the kind *cat is* or *cat has* controls the flow of activation between the word structures.

In Figures 2 and 3, the connections in the neural blackboard and between the word structures and the blackboard are also conditional connections, in which flow of activation and binding are controlled by circuits that parse the syntactic structure of the sentence. These circuits, for example, detect (simply stated) that *cat* is a noun and that it is the subject of the sentence *cat sees dog*. However, the specific details of the control and parsing processes that allow these temporal connections to be formed are not the main focus of this article. Details can be found in [9]. Here, I will focus on the general characteristics that are required by any architecture that combines grounded representations in a productive way. Understanding these general features is important for the interaction between AI and neuroscience.

7.3. Characteristics of Grounded Architectures. The first characteristic is the grounded nature of representations in combinatorial structures. In Figures 2 and 3, the representations of *cat*, *sees*, and *dog* remain grounded in the whole binding process. But the structure of the sentence is compositional. The syntax populations (structure assemblies) play a crucial role in this process, because they allow temporal connections to be formed between grounded word representations. For example, the productivity of language requires that we can form a relation between an arbitrary verb and an arbitrary noun as its subject. But we can hardly assume that all word structures for nouns are connected with all word structures for verbs, certainly not for noun verb combinations that are novel. Yet, we can assume that there are connections between words structures for nouns and a limited set of noun phrase populations, and that there are connections between words structures for verbs and a limited set of verb phrase populations. And we can assume that there are connections between noun phrase and verb phrase populations. So, using the indirect link provided by syntax populations we can create new (temporal) connections between arbitrary noun and verbs, and temporal connections between words of other syntactic types as well.

The second characteristic is the use of conditional and temporal connections in the architecture. Conditional connections provide a control of the flow of activation in connections. This control of activation is necessary to encode relational information. By controlling the flow of activation the architecture can answer specific queries such as *what does the cat see?* or *who sees the dog?*. Without such control of activation, only associations between word (concept) structures could be formed. But when connections are conditional and temporal (i.e., their activation is temporal), arbitrary and novel combinations can be formed in the same architecture (see [13]).

The third characteristic is the ability to create combinatorial structures in which the same grounded representation is used more than once. Because grounded representations cannot be copied, another solution is needed to solve this problem of multiple instantiations, that is, the "problem of two" [9]. Figure 4 illustrates this solution with the sentences *The cat sees the dog* and *The dog sees the cat* (ignoring *the*). The combinatorial structures of these two sentences can be stored simultaneously in the blackboard architecture, without making copies of the representations for *cat*, *sees* and *dog*. Furthermore, *cat* and *dog* have different syntactic roles in the two sentences.

Figure 4 illustrates that the syntax populations eliminate the need for copying representations to form sentences. Instead of making a copy, the grounded representation of *cat* is connected to N_1 in the sentence *cat sees dog* and to N_4 in the sentence *dog sees cat*. Because N_1 is connected to S_1,

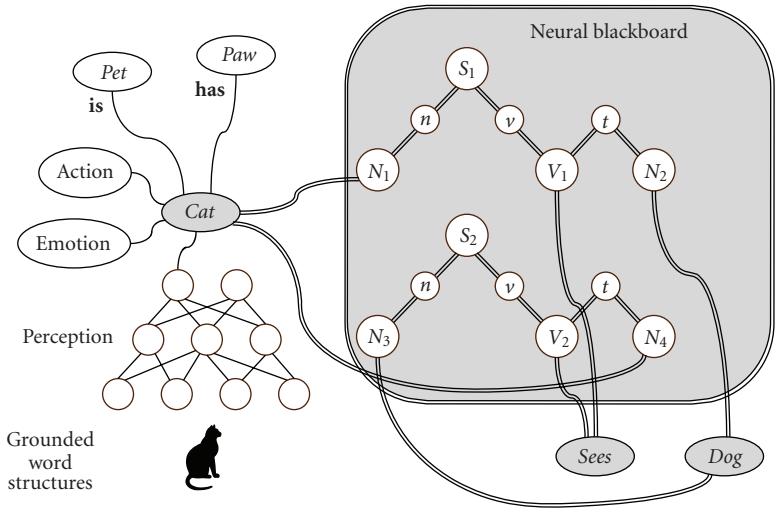


Figure 4: Illustration of the combinatorial structures of *The cat sees the dog* and *The dog sees the cat* (ignoring *the*), with grounded representations for the words. The circles in the neural blackboard represent populations and circuits of neurons. The double line connections represent conditional connections. (N, n = noun; S = sentence; t = theme; V, v = verb.)

cat is the subject in the sentence *cat sees dog*. It is the theme (object) in the sentence *dog sees cat*, because N_4 is connected to V_2 as its theme. The multiple binding of the grounded representations *dog and sees* proceeds in a similar way.

The fourth characteristic concerns the (often sequential) control of activation in the architecture. As I noted above, the conditional connections provide the ability to control the flow of activation within the architecture. Without this control, the architecture cannot represent and process combinatorial structures and relations. Control of activation results from neural circuits that interact with the combinatorial structures. Examples of control circuits can be found in [13, 28].

Figure 5 illustrates how these control circuits can affect and regulate the dynamics in the architecture, and with it the ability to process and produce information. With control of activation, the architecture can answer specific queries like *what does the cat see?* (or *cat sees?,* for short). The query *cat sees?* activates the grounded representations *cat* and *sees*. When the sentences *cat sees d*og and *dog sees cat* are stored in the blackboard, *cat* activates N_1 and N_4, because it is temporarily bound with these syntax populations. Likewise, *sees* activates V_1 and V_2.

But the query *cat sees?* also provides the information that *cat* is the subject of a verb. Using this information, control circuits can activate the conditional connections between subject syntax populations. In Figure 5 these are the connections between N_1 and S_1 and between N_3 and S_2. Because *cat* has activated N_1, but not N_3, N_1 activates S_1. Notice that the activation of N_4 by *cat* has no effect here, because N_4 is bound to V_2 as its theme (t), and these conditional connections are not activated by the query (yet). Because *cat* is the subject of a verb (*sees*), this information can be used to activate the conditional connections between the S_i and V_j populations in the architecture. Because S_1 is the only active S_i population, this results in the activation of V_1 by S_1.

At this point, a fifth characteristic of grounded cognition emerges: the importance of dynamics. Figure 5 shows why dynamics is important. Because *sees* is grounded, the query *cat sees?* has activated all V_j populations bound to *sees*, here V_1 and V_2. This would block the answer to the query, because that consists of activating the theme of V_1 but not the theme of V_2. However, due to the process described above, S_1 also activates V_1. Because populations of the same nature compete in the architecture (by inhibition), V_1 wins the competition with V_2.

When V_1 has won the competition with the other V_j populations, the query can be answered. The query *cat sees?* asks for the theme of the verb for which *cat* is the subject. That is, its asks for the theme of a syntax population bound to *sees*. After the competition, V_1 has emerged as the winning syntax population bound to that verb, so the query asks for the theme of V_1. It can do so by activating the conditional connections between V_1 and N_2 (see [9]). This will result in the activation of N_2 and with that of *dog* as the answer to the query.

The sequential nature of control illustrated in Figure 5 resembles that of control of movement. Executing a particular movement usually consists of sequential activation of a set of muscles. For example, when we swing an arm back and forth, its muscles have to be activated and deactivated in the correct sequence. More complex movement patterns like dancing or piano playing require elaborate sequential control of muscles being activated and deactivated. The motor programs for these movement patterns could in fact be a basis for the development of grounded representations. After all, muscles are "grounded" by nature. That is, we have just one set of muscles that we use to make specific movement sequence.

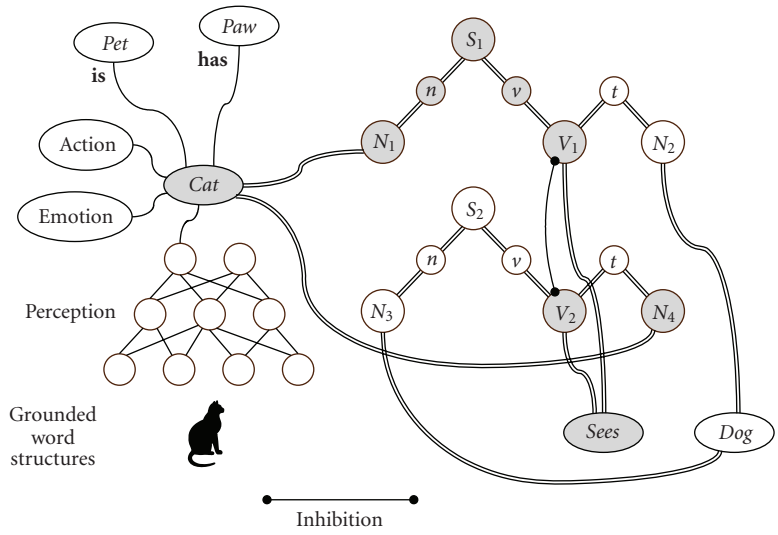


Figure 5: Illustration of the combinatorial structures of *The cat sees the dog* and *The dog sees the cat* (ignoring *the*), with grounded representations for the words. The circles in the neural blackboard represent populations and circuits of neurons. The grey nodes represent activate populations initiated by the query *cat sees?*. The double line connections represent conditional connections. (N, n = noun; S = sentence; t = theme; V, v = verb.)

7.4. Blackboard Architectures for Cognitive Processing. The combination of productivity and grounding requires certain architectures in which the grounded representations can be combined temporarily into combinatorial structures. The neural blackboard for sentence structure illustrated in Figures 2 and 3 is an illustration of such an architecture.

The neural blackboard illustrated in Figures 2, 3 and 4 provides the ability to form sentence structures. But words, for example, also have a phonological structure, and these structures are productive (combinatorial) as well. So, words would also be a part of a phonological neural blackboard. Words (concepts) can be used in reasoning processes based on sentence structures, which would require a specific blackboard architecture as well [13]. But words could also be a part of nonsentence like sequences, which could be used for other specific forms of reasoning [29]. Because the sentence blackboard is not suited for these sequences, a specific sequence blackboard is required as well.

Thus, grounded conceptual representations will be embedded in neural blackboards for sentence structure, phonological structure, sequences and reasoning processes, and potentially other blackboards as well. One might argue that this is overly complex. But complexity is needed to account for human cognition. Complexity is hidden in symbol manipulation as well. For example, when a specific symbol manipulation process is executed on a computer, a lot of its complexity is hidden in the underlying machinery provided by the computer. As a model of cognition, this machinery has to be assumed as a part of the model.

Furthermore, the embedding of representations in different blackboards is a direct consequence of the grounded nature of representations. Because these representations always remain "in situ", they have to be connected to architectures like blackboards to form combinatorial structures and to execute processes on the basis of these structures. In fact, the grounded representations form the link between the different blackboard architectures. When processes occur in one blackboard, the grounded representation can also induce processes in the other blackboards, which could in turn influence the process in the first blackboard. In this way, an interaction occurs between local information embodied in specific blackboards and global information embodied in grounded representations.

Viewed in this way, architectures of grounded cognition reverse the relation between control and representation as found in symbolic architectures of cognition. In the latter, resembling the digital computer, control is provided by a central "fixed" entity (e.g., the CPU) and representations move around in the architecture, when they are copied and transported. In grounded cognition, however, the representations are "fixed", whereas control moves around within and between blackboards.

8. Research Directions: Searching for Grounded Architectures of Cognition

The analysis given above suggests that cognition on the level of human cognition arises from the interaction between grounded representations and productive (blackboard) architectures. If so, these grounded architectures (for short) would have to be instantiated in the brain. This raises the question of how one could demonstrate that these architectures exist, and how their properties could be studied.

Empirical techniques such as electrodes, EEG (electroencephalogram) and fMRI (functional magnetic resonance imaging) are used to study "cognition in the brain". Each of these techniques provides valuable information about how the brain instantiates cognition. But each of them

is also limited. EEG provides information about groups of neurons (typically in the millions), for the most part located at the surface of the cortex. It's temporal resolution is very high, whereas its spatial resolution is relatively low. Functional MRI provides better spatial resolution (although not on the level of the neuronal circuits as found in cortical columns), but it's temporal resolution is too low to capture the dynamics of cognition.

Electrodes, inserted in the cortex, have the best spatial and temporal resolution. But the number of electrodes that can be inserted is limited relative to the number of neurons involved in a cognitive process. Moreover, it's use in humans is restricted to specific cases that arise when humans need brain surgery for medical reasons (e.g., [8]). A rigorous use as in animal experiments is excluded with humans for obvious ethical reasons. But the consequence of that is that detailed theories and models of human cognitions could never be tested empirically in detail.

It is important to emphasize this point, because it entails an additional difficulty that the study of human cognition faces. A scientific requirement of theories and models is that they can be tested empirically. Sometimes, theories and models cannot be tested (in full) because they are (partly) too vague or ambiguous. Such theories and models do not meet scientific standards in full. But in the case of human cognition, theories and models could be exact, detailed and unambiguous, but fail empirical testing due to ethical reasons. This is particularly true for the features of cognition that are specifically human. Detailed information is available for visual processing, for example, because we have animal models to test and investigate vision. But animal models are missing for language, planning, reasoning and other more exclusively human forms of cognition.

Perhaps the only way to test theories and models for these features of human cognition is large-scale brain modelling. Animal models could be of value because they provide the initial information and testing for simulating cortical columns, areas and pathways. Given the uniform nature of the cortex, between and within animals, these simulations could form the basis for cortical models of cognitive processes that are more specifically human. As suggested here, these models would consist of grounded architectures. These architectures require more than the simulation of cortical structures suited for animal cognition. For example, specific connection structures are needed to create blackboard architectures [9]. So, simulations need to investigate how these connection structures can be formed with cortical columns, or how other connection structures can be formed with cortical columns that have the same functional abilities.

In this process, AI would take a leading role, because it can develop detailed models of cognitive processes based on neural architectures. These models could then be used as a target for cortical simulations. That is, with cortical simulations it could be investigated whether and how the neural models developed by AI can be instantiated with the cortical building blocks found in the brain. In turn, these cortical simulations could be investigated by deriving virtual measurements from them, resembling electrode, EEG and even fMRI measurements. The latter could then be compared with measurements derived from actual brains.

The role of AI in this process is to analyze the mechanisms that can produce high-level processes of human cognition, and to develop neural instantiations for these mechanisms, such as the neural blackboard architectures discussed in the previous section. Neuroscience would provide the detailed information about the cortical building blocks, as discussed earlier. Large-scale simulations would integrate and further develop these two lines of investigation. So, AI has an important role to play in this research. But AI may also benefit from it, because this research could also solve important issues concerning the nature and mechanisms of intelligence and cognition. I will briefly discuss some of them in the final section.

9. Investigating Deep Problems

A number of issues in the study of (human) cognition can be characterized as "deep" problems. They concern the very nature of human-level cognition, and they have been the topic of speculation from the very beginning of thinking about cognition. But they largely remain as problems to be solved. The lack of progress with these problems also has a clear negative effect on the development of artificial forms of intelligence. The solution of these problems is most likely to be found in the unique way in which the human brain produces cognition, and thus in the unique computational and cognitive features of the neural architectures in the brain. Motivation for this assumption is found in the fact that the human brain is the only known example of a system that produces (human-level) cognition. Investigating the neural architectures of cognition thus provides the possibility to study at least some of these problems in a way that has not been available before. A few of these problems can be singled out.

9.1. Conceptual Structure (Meaning). Arbitrary symbols, gestures or sounds can be used to convey meaning, such as words and sentences in language. The question is how arbitrary symbols and sounds acquire meaning, what the nature (structure) of their meaning is, and how they succeed in conveying their meaning. An indication of the profound nature of these questions is the fact that meaning is one of the major problems in automatic language translation. Neuroimaging research has already demonstrated that there are relations between the neural representation of certain words and sensory-motor representations in the brain (e.g., action verbs activate parts of the motor cortex that are involved in the actions these verbs denote). Given these relations, it can be assumed that the nature and development of certain conceptual representations in the brain are related to the nature and development of sensory representations (e.g., sensory categorizations), motor representations, or transformations between representations. Thus, the study of sensory-motor representations and their transformations in neural architectures (as outlined above) could also be used to study the nature and development of those conceptual

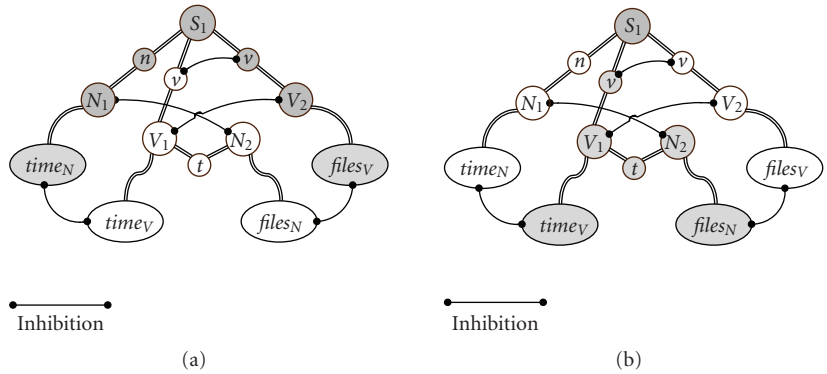


Figure 6: Competing neural blackboard structures for *time flies*. In (a), the competition results in $time_N flies_V$. In (b) the competition results in $time_V flies_N$. The ovals and circles represent populations as in Figures 2 and 3. Grey circles and ovals are active. (N, n = noun; S = sentence; t = theme; V, v = verb.)

structures that are related to sensory representations (e.g., nouns or adjectives), motor representations (e.g., verbs), or transformations (e.g., prepositions, [9]).

9.2. Selection of Information (Resolution of Ambiguity). Information is often ambiguous. A good illustration is given by language. Almost all words in language have multiple meanings. Consequently, sentences are very often ambiguous. For example, in the sentence *Time flies like an arrow*, the word *time* can be a noun, a verb, or even an adjective (i.e., *time flies* as in *fire flies*). Furthermore, the word *flies* can be a verb or a (plural) noun and the word *like* can be a verb or an adverb. Each of these choices provide a different interpretation for this sentence, for which at least five different interpretations are possible [19]. Artificial (computer) programs for sentence interpretation and translation have substantial difficulties in handling these forms of ambiguity.

Ambiguities are common in language and cognition in general, but humans often do not notice them [30, 31]. This is also the case for the sentence *Time flies like an arrow*. The usual interpretation of this sentence is in terms of a metaphor, that states that time changes very fast. Humans usually end up with this (one) interpretation, but a computer program of sentence analysis (based on symbol manipulation) gave all five interpretations [19]. The fact that humans can operate remarkably well with ambiguous sentences indicates that they have the ability to select the relevant or intended meaning from the ambiguous information they receive.

The difficulty of artificial intelligence systems to select relevant information has been another major problem in their development (sometimes referred to as the frame problem). Selecting relevant information is in particular a problem for generative (rule-based) processing. It is in fact the downside of the productivity of this form of processing. With generative processing, too many possibilities to be explored are often produced in a given situation. In contrast, associative structures such as neural assemblies are very suited for selecting relevant information. For example, when information in a neural assembly is partly activated, the assembly will reactivate all related information as well. The ability to select relevant information in human cognition could thus result from a combination of generative and associative processing. The development of grounded neural architectures of cognition, in which neural assemblies are combined with generative processing in neural blackboard architectures, as illustrated above, provides a way to investigate this possibility.

Figure 6 illustrates how ambiguity resolution could occur in a neural architecture of grounded cognition. In the architecture, dynamical interactions can occur between sentence structures [9]. Similar interactions can also influence the binding process, that is, the process by which a sentence structure is formed [19]. Figure 6 shows the competing sentence structures of *time flies*. The word *time* activates two grounded (word) structures, one for *time* as a noun ($time_N$) and one for *time* as a verb ($time_V$). In the same way, *flies* activates $flies_N$ and $flies_V$.

Initially each of the word structures binds to corresponding syntax populations, such as N_1 and V_1. These syntax populations then form competing sentence structures. One is the sentence structure for $time_N flies_V$ (the grey nodes in Figure 6(a)). Here, $time_N$ is the subject of the sentence and $flies_V$ is the main verb. The other is the sentence structure for $time_V flies_N$ (the grey nodes in Figure 6(b)). Here, $flies_N$ is the theme (t) of the verb $time_N$.

In the architecture, there is a dynamic competition between the sentence structures and between word structures. In particular, the word structures for $time_N$ and for $time_V$, and those for $flies_N$ and $flies_V$ inhibit each other. This competition implements the constraint that a word can have only one interpretation at the same time in a sentence structure. Between the sentence structures there is a competition (inhibition) between the circuits that activate conditional connections of the same kind (in Figure 6 those for the verb connections), and inhibition between similar syntax populations (e.g., between the noun phrases N_1 and N_2 and between the verb phrases V_1 and V_2).

The outcome of the competition is either the structure illustrated with the grey nodes in Figure 6(a), or the structure with the grey nodes in Figure 6(b). The competition is

resolved when there is a clear advantage for one of the competing structures [13, 28]. In Figure 6, an advantage for one of the sentence structures can arise from the fact that the interpretation of *time* as a noun is more frequent than the interpretation of *time* as a verb. In that case, the activation of $time_N$ will be stronger than that of $time_V$, so that the first inhibits the second. Then, $flies_V$ inhibits $flies_N$, because $flies_V$ is activated by $time_N$ through the sentence structure, whereas N_2 is inhibited by N_1 (this inhibition becomes stronger with increasing activation of $time_N$). In this way, the grey structure in Figure 6(a) remains as the active structure to which the rest of the sentence, *like an arrow*, binds.

The competition in Figure 6 illustrates why grounded representations are important, and why they have to remain grounded in combinatorial structures. The competition that solves the ambiguity of *time flies*, for example, results from the interaction between the structures of $time_N$ and $time_V$. The assumption is that $time_N$ wins this competition because it is used more frequently than $time_V$ in natural language. Due to the grounded nature of representations, the more frequent use of $time_N$ will affect the grounded representation of $time_N$ directly, because this representation is always used to represent $time_N$. Furthermore, the difference between $time_N$ and $time_V$ is found only in sentence contexts, thus in combinatorial structures. So, when grounded representations remain grounded in combinatorial structures, the more frequently used type of combinatorial structure can influence the grounded structures involved directly.

9.3. Learning and Development. This is a topic of extensive research, which is very important for understanding cognition. One problem concerning learning and development perhaps stands out. It concerns the difference between associative versus generative (rule-based) processing, which in turn relates to the age-old debate between nature and nurture.

Associative processing plays an important role in human cognition. Examples are the neural assemblies proposed by Hebb [11]. Furthermore, the learning mechanisms discovered in the brain (e.g., long-term potentiation) concern the forming of new associations. Thus associative processing gives an account of the development of cognition (nurture). Examples are the associations that can develop within and between grounded conceptual representations. However, generative processing is needed for the productivity of cognition. But the development (learning) of generative processing is difficult to account for, which has led to the assumption that the basic principles of generative processing are innate (nature). Yet, these innate abilities develop only with proper stimulation (experience).

The problem thus concerns the question of what features of generative processing are innate, and how this innate ability develops on the basis of experience. If neural architectures of generative processing are adequately captured in a model, this problem could for the first time be addressed in a more experimental way, by using a backtracking procedure (reverse engineering). With this procedure one can simplify the known (fully developed) neural architecture and then investigate how the fully developed architecture can evolve from the more simplified version of it. This approach could be repeated in several steps, leading to a more and more elementary architecture as the basis of the fully developed architecture.

10. Conclusion

For the first time in history, it is possible to investigate the neural mechanisms that produce human cognition. It can be done because the experimental methods and techniques are now available to investigate the structure of the brain, because the theoretical knowledge is available that provides the possibility of a theoretical analysis of neural mechanisms of cognition, and because the computer power is now available that provides the possibility of large-scale simulations and numerical analyses of these mechanisms.

However, the complexity of the brain, and the cognitive processes it produces, entails that integrated multidisciplinary expertise is needed to combine these lines of research. The computational perspective on neurocognition, aimed at understanding how the neural dynamics and neural mechanisms of the brain produce cognition, can play a fundamental role in this respect, because it focuses on the ultimate aim of neurocognition [1]. So, AI has an important role to play in this process.

But AI can also benefit from it, because a detailed analysis of how the brain produces cognition could provide important information about the nature of cognition itself. Here, I have argued that understanding the neural basis of cognition could reveal important characteristics of its grounded nature. For example, combinatorial structures can be created with grounded representations, but not all structures are equally feasible [13]. And, as illustrated in Figure 6, the combinatorial structures formed are influenced by dynamics, which provides additional constraints on the ability to create combinatorial structures. The example given in Figure 6 show that these constraints prevent the excessive production of sentence interpretations, as found in systems with unlimited productivity based on symbol manipulation. But, on occasion, it can also result in misrepresentations, which is indeed found in human cognition as well.

The combination of grounding and productivity could solve a problem about cognition addressed by Fodor. Although he supported the computational view of cognition from its beginning, more recently Fodor has argued that a computational (symbol manipulation) account of cognition is incomplete [32]. In particular, because the computational processes provided by symbol manipulation are always local (as illustrated in the quote from Newell [22]). Local processing, in the view of Fodor, does not capture the global flexibility of cognition, which may be the most important feature of human cognition [32].

Grounded cognition, as presented here, is both local and global, and it is productive. Processes that occur within specific blackboards are local, but the grounded representations involved are global. The interaction between blackboards and grounded representations thus provides a basis for the productivity and global flexibility of cognition.

References

[1] M. S. Gazzaniga, "Preface," in *The Cognitive Neurosciences*, M. S. Gazzaniga, Ed., MIT Press, Cambridge, Mass, USA, 1995.

[2] V. Braitenberg, "Two views of the cerebral cortex," in *Brain Theory*, G. Palm and A. Aertsen, Eds., Springer, Berlin, Germany, 1986.

[3] V. Braitenberg and A. Schüz, *Anatomy of the Cortex: Statistics and Geometry*, Springer, Berlin, Germany, 1991.

[4] W. H. Calvin, "Cortical columns, modules, and Hebbian cell assemblies," in *The Handbook of Brain Theory and Neural Networks*, M. A. Arbib, G. Adelman, and P. H. Arbib, Eds., pp. 269–272, MIT Press, Cambridge, Mass, USA, 1995.

[5] D. J. Felleman and D. C. Van Essen, "Distributed hierarchical processing in the primate cerebral cortex," *Cerebral Cortex*, vol. 1, no. 1, pp. 1–47, 1991.

[6] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, "Robust object recognition with cortex-like mechanisms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 3, pp. 411–426, 2007.

[7] K. Grill-Spector and R. Malach, "The human visual cortex," *Annual Review of Neuroscience*, vol. 27, pp. 649–677, 2004.

[8] R. Q. Quiroga, L. Reddy, G. Kreiman, C. Koch, and I. Fried, "Invariant visual representation by single neurons in the human brain," *Nature*, vol. 435, no. 7045, pp. 1102–1107, 2005.

[9] R. Jackendoff, *Foundations of Language*, Oxford University Press, Oxford, UK, 2002.

[10] M. Abeles, *Corticonics: Neural Circuits of the Cerebral Cortex*, Cambridge University Press, New York, NY, USA, 1991.

[11] D. O. Hebb, *The Organization of Behavior*, John Wiley & Sons, New York, NY, USA, 1949.

[12] E. Bienenstock, "Composition," in *Brain Theory: Biological Basis and Computational Theory of Vision*, A. Aertsen and V. Braitenberg, Eds., pp. 269–300, Elsevier, New York, NY, USA, 1996.

[13] F. van der Velde and M. de Kamps, "Neural blackboard architectures of combinatorial structures in cognition," *Behavioral and Brain Sciences*, vol. 29, no. 1, pp. 37–70, 2006.

[14] J. Frye, R. Ananthanarayanan, and D. S. Modha, "Towards real-time, mouse-scale cortical simulations," IBM Research Report RJ10404, 2007.

[15] H. Markram, "The blue brain project," *Nature Reviews Neuroscience*, vol. 7, no. 2, pp. 153–160, 2006.

[16] G. M. Shepherd, *Neurobiology*, Oxford University Press, Oxford, UK, 1983.

[17] R. J. Douglas and K. A. C. Martin, "Neocortex," in *The Synaptic Organization of the Brain*, G. M. Shepherd, Ed., pp. 389–438, Oxford University Press, Oxford, UK, 3rd edition, 1990.

[18] F. van der Velde and M. de Kamps, "From knowing what to knowing where: modeling object-based attention with feedback disinhibition of activation," *Journal of Cognitive Neuroscience*, vol. 13, no. 4, pp. 479–491, 2001.

[19] S. Pinker, *The Language Instinct*, Penguin, London, UK, 1994.

[20] G. A. Miller, *The Psychology of Communication*, Penguin, London, UK, 1967.

[21] J. R. Anderson, *The Architecture of Cognition*, Harvard University Press, Cambridge, Mass, USA, 1983.

[22] A. Newell, *Unified Theories of Cognition*, Harvard University Press, Cambridge, Mass, USA, 1990.

[23] S. Harnad, "The symbol grounding problem," in *Emergent Computation: Self-Organizing, Collective, and Cooperative Phenomena in Natural and Artificial Computing Networks*, S. Forrest, Ed., MIT Press, Cambridge, Mass, USA, 1991.

[24] L. W. Barsalou, "Perceptual symbol systems," *Behavioral and Brain Sciences*, vol. 22, no. 4, pp. 577–660, 1999.

[25] J. A. Fodor and Z. W. Pylyshyn, "Connectionism and cognitive architecture: a critical analysis," in *Connections and Symbols*, S. Pinker and J. Mehler, Eds., pp. 3–71, MIT Press, Cambridge, Mass, USA, 1988.

[26] W. H. Calvin and D. Bickerton, *Lingua ex Machina: Reconciling Darwin and Chomsky with the Human Brain*, MIT Press, Cambridge, Mass, USA, 2000.

[27] D. J. Amit, "The Hebbian paradigm reintegrated: local reverberations as internal representations," *Behavioral and Brain Sciences*, vol. 18, no. 4, pp. 617–657, 1995.

[28] F. van der Velde and M. de Kamps, "Learning of control in a neural architecture of grounded language processing," *Cognitive Systems Research*, vol. 11, no. 1, pp. 93–107, 2010.

[29] R. F. Hadley, "The problem of rapid variable creation," *Neural Computation*, vol. 21, no. 2, pp. 510–532, 2009.

[30] B. J. Baars and S. Franklin, "How conscious experience and working memory interact," *Trends in Cognitive Sciences*, vol. 7, no. 4, pp. 166–172, 2003.

[31] B. J. Baars, "Conscious cognition and blackboard architectures," *Behavioral and Brain Sciences*, vol. 29, no. 1, pp. 70–71, 2006.

[32] J. A. Fodor, *The Mind Doesn't Work That Way*, MIT Press, Cambridge, Mass, USA, 2000.

Hindawi Publishing Corporation
Advances in Artificial Intelligence
Volume 2010, Article ID 209254, 10 pages
doi:10.1155/2010/209254

Research Article

Recurrence Quantification Analysis of Spontaneous Electrophysiological Activity during Development: Characterization of In Vitro Neuronal Networks Cultured on Multi Electrode Array Chips

Antonio Novellino and José-Manuel Zaldívar

Institute for Health and Consumer Protection—(IHCP), Joint Research Centre—(JRC), European Commission—(EC), Via E. Fermi 2749, 21207 Ispra, Italy

Correspondence should be addressed to Antonio Novellino, antonio.novellino@jrc.ec.europa.eu

Received 24 August 2009; Revised 14 October 2009; Accepted 28 October 2009

Academic Editor: Naoyuki Sato

The combination of a nonlinear time series analysis technique, Recurrence Quantification Analysis (RQA) based on Recurrence Plots (RPs), and traditional statistical analysis for neuronal electrophysiology is proposed in this paper as an innovative paradigm for studying the variation of spontaneous electrophysiological activity of in vitro Neuronal Networks (NNs) coupled to Multielectrode Array (MEA) chips. Recurrence, determinism, entropy, distance of activity patterns, and correlation in correspondence to spike and burst parameters (e.g., mean spiking rate, mean bursting rate, burst duration, spike in burst, etc.) have been computed to characterize and assess the daily changes of the neuronal electrophysiology during neuronal network development and maturation. The results show the similarities/differences between several channels and time periods as well as the evolution of the spontaneous activity in the MEA chip. RPs could be used for graphically exploring possible neuronal dynamic breaking/changing points, whereas RQA parameters are suited for locating them. The combination of RQA with traditional approaches improves the identification, description, and prediction of electrophysiological changes and it will be used to allow intercomparison between results obtained from different MEA chips. Results suggest the proposed processing paradigm as a valuable tool to analyze neuronal activity for screening purposes (e.g., toxicology, neurodevelopmental toxicology).

1. Introduction

In vitro neuronal networks are a simplified and accessible model of the central nervous system (CNS). Moreover, they exhibit morphological and physiological properties [1] and activity-dependent path-specific synaptic modification similar to the in vivo tissue [2, 3]. In vitro Neuronal Networks (NNs) of cortical cells grown on multielectrode array (MEA) chips have been shown to be a valuable tool to study fundamental properties of network activity patterns [4–8], plasticity [2, 9], learning in vitro [10–12], and pharmacological screening [13–18]. While growing, the networks of cortical cells manifest different spontaneous electrophysiological dynamics [7, 19, 20] and therefore analysis and characterisation of the network dynamics can provide important information for understanding network behaviour and optimising experimental design.

The spike trains can be accurately extracted from MEA recordings, for example, [21, 22], and thus neuronal activity is translated into time series of discrete events. The ensembles of spike trains simultaneously recorded from many channels represent multidimensional nonstationary point-process time series which makes the data analysis highly challenging [23]. Identification, description, and prediction of the changes of such dynamic during the neuronal network development are even more complex and challenging.

A promising tool for the qualitative and quantitative analysis of nonstationary signals is Recurrence Quantification Analysis (RQA) [24] which was developed to quantify

the parameters of Recurrence Plots (RPs) [25]. An RP is a two-dimensional graph which reveals all the moments at which the state space of the dynamical system recurs, that is, visits the same region in state space (recurrence of states). The recurrence of states, meaning that states are arbitrary close after some time, is a fundamental property of deterministic dynamical systems and is typical for nonlinear or chaotic systems. The RQA is a method for assessing the dynamics of noisy, short, and nonstationary signals typical of a dynamical system (e.g., neuronal electrophysiology). RPs of neuronal spike trains were analyzed by Kalużny and Tarnecki [26] where they showed changes in the structure of the interspike interval sequences recorded from cerebellum and red nucleus of anesthetized cats. The application of RP was also proposed by Novák and Schmidt [27] for the analysis of the time series generated by their model of neuronal stochastic activity. More recently, Marwan and Meinke [28] used RQA to detect transitions in event-related brain potentials and Bergner et al. [29] to analyze synchronization in neuronal networks.

In this work, we used a combination of well-known electrophysiological parameters (e.g., mean firing rate, mean bursting rate, etc.) and RQA parameters (e.g., percentage of recurrence, of determinism, laminarity, etc.) for identifying and characterizing the key events that underlie the electrophysiological dynamics as it changes daily during the neuronal network growth.

The analysis showed the similarities and differences of the neuronal dynamic in the space (recording channels) and time (different days). Our results suggest that the combination of RQA with traditional approaches improves the identification, description, and prediction of electrophysiological changes in neuronal networks coupled to MEA chips. The application of the presented methodology could be used to perform intercomparison between results obtained from different MEA chips, when the characterization and quantification of electrophysiological endpoints are needed for further assessment (e.g., toxicology).

Finally, our result suggests that the spontaneous activity of in vitro network of neurons is less noisy than it was expected, and the identified deterministic and laminar behaviours represent a main advantage for developing more robust and accurate *in silico* model of neuronal network.

2. Materials and Methods

2.1. Neuronal Network. The experiments were preformed with Cryo preserved neurons from mouse cortex (Clonetics C57 black, E14-E16) (Lonza, M-CX-300). Cryo preserved cells are very delicate: we used cryovials containing 4 million cells and we recorded a viability ranging from 15% to 25% (Trypan Blue assessment). The culture medium is PNBM, L-Glutamine, Gentamycine, AmphotericinB, and NSF-1 (cryo cell bullet kit, Lonza CC-4461). The final cell density was about 35.000 cells per chip. The chips were then incubated at 37°C in 5% CO_2, 20% O_2. Starting at 3 DIV (Day In Vitro), half of the culture medium was exchanged twice a week under the laminar hood. The study of the electrophysiology started after the second day of in vitro incubation.

2.2. Chip Preparation. Standard 60-electrode MEA chips (with 30 μm diameter electrodes, 200 μm interelectrode spacing with an integrated reference electrode) were employed. Prior to plating cells, the MEA chip was sterilized (2 hours in oven at 122°C). The sterilized chip was coated with 10% New Born Calf Serum (NBCS, Invitrogen/Gibco 16010-159). Laminin (Sigma L2020) and Poli-D-Lysin (Sigma P6407) were then applied to the surface of the MEA to render the surface cellphilic and to promote neurites outgrowth. Every coater deposition was carried out at room temperature into laminar hood. The chips with Laminin were incubated three hours into incubator and then it was gently removed. The Poly-D-lysin was left overnight. Once the second coater was removed, the chip was ready to host cells.

2.3. Recording System. The activity was recorded by the MEA120-2-System from Multichannel Systems (MCS GmbH, Ruetlingen, Germany, http://www.multichannel-systems.com). In particular, the MEA was fed into the MEA Amplifier (Gain 1000x) and data were recorded by the MC_Rack software at a sampling rate of 10 kHz. A bandpass digital filter (60 Hz–4000 Hz) was also applied. Figure 1 shows an example of the neuronal network coupled to the MEA chip and the typical recorded electrophysiological signal. The system also included a temperature controller (TC02, MCS GmbH) that allowed heating the MEA chips and thus the medium from the bottom.

2.4. Signal Processing

2.4.1. Characterization of the Spontaneous Electrophysiology. Spikes were extracted when the raw signals overcame a threshold set at −6.5 times the standard deviation of the mean square root noise. The raw signal was then translated into time series of discrete events, that is, the Spike Train:

$$\mathrm{ST}(t) = \sum_{s=1}^{N} \delta(t - t_s). \tag{1}$$

The recorded spike trains were processed to extract descriptors of the spontaneous electrophysiology at both spike and burst levels. In particular we extracted the Mean Firing Rate (MFR) and Mean Burst Rate (MBR), and we also studied the number of spikes per burst (i.e., Burst Amplitude (BA)), Burst Duration (BD) and Interburst Interval (IBI). Bursts were extracted from spike trains according to already presented methods [30]. Briefly, a burst is a dense sequence of spikes, and we applied the following burst detection parameters: a burst is composed of 5 spikes at least, the interspike interval between two successive spikes belonging to the same burst is 100 milliseconds at most, and two serial bursts are separated by 100 milliseconds of non bursting activity.

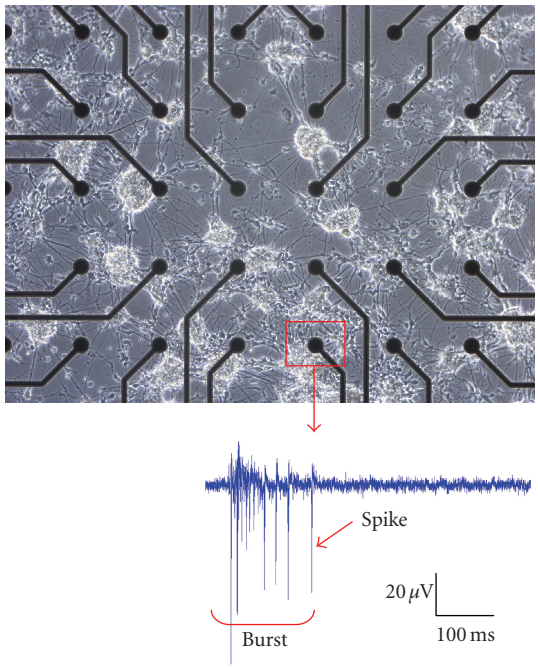


Figure 1: An in vitro neuronal network coupled to a multielectrode array chip and the typical recordable electrophysiological activity. The NN is randomly self-reassembled from cryo preserved cortical neurons of mouse. The microelectrode is 30 μm in diameter and the interelectrode distance is 200 μm. After few days of in vitro culture it is possible to record both spikes and packages of spike (bursts).

It is then possible to describe the Burst Train BT(t) such as

$$\mathrm{BT}(t) = \sum_{b=1}^{M} \left(\mathrm{BA}_b \Pi \left(\frac{t - t_b - \tau_b/2}{\tau_b} \right) \right), \tag{2}$$

where M is the number of bursts within the train, t_b denotes the starting time of the bth burst, $\Pi(t/\tau)$ is the rectangular function denoting the occurrence of a burst at time $t = t_b$ and lasting τ, and BA_b is the number of spike per burst (i.e., Burst Amplitude):

$$\mathrm{BA}_b = \frac{1}{\tau_b} \int_{\tau_b} \sum_{s=1}^{N} \delta(t - t_s) dt = \frac{N_{\tau_b}}{\tau_b}, \tag{3}$$

where τ_b is the burst duration, and N_{τ_b} is the number of the spikes belonging to that burst.

2.4.2. Linear Correlation Assessment. To check if the spikes time series or the originally raw signals were linearly correlated, we used the correlation matrix and the Cross Correlation Function (CCF). The correlation coefficient matrix represents the normalized measure of the strength of linear relationship between variables. The correlation coefficient R of two variables x and y is given by

$$R(x, y) = \frac{\mathrm{COV}(x, y)}{\sqrt{\mathrm{VAR}(x)\mathrm{VAR}(y)}}, \tag{4}$$

where $\mathrm{COV}(x, y)$ is the covariance matrix.

The correlation coefficients range from -1 to 1, where values close to 1 suggest that there is a positive linear relationship between the data columns, and values close to -1 suggest that one column of data has a negative linear relationship to another column of data (anticorrelation). Values close or equal to 0 suggest that there is no linear relationship between the data columns.

The significance of each correlation was evaluated by the t-test.

Cross correlation is a generalization of the correlation coefficient and it is a standard method for estimating the degree to which two series are correlated when we shift them one in respect to the others [31].

Consider two series x_i and y_i where $i = 1, 2, \ldots, n$. The cross correlation R at delay d is defined as

$$R(x, y, d) = \frac{\sum_i [(x_i - \overline{x})(y_{i-d} - \overline{y})]}{\sqrt{\sum_i (x_i - \overline{x})^2} \sqrt{\sum_i (y_i - \overline{y})^2}}. \tag{5}$$

If the above is computed for all delays $d = 0, 1, 2, \ldots, n - 1$, then it results in a cross correlation series of twice the length as the original series. For $d = 0$ it becomes the linear correlation coefficient $R(x, y)$.

2.4.3. Recurrence Plot Analysis. A recurrence plot (RP) is a two-dimensional graph which reveals all the moments at which the state space of the dynamical system recurs, that is, visits the same region in state space. Recurrence Plots (RPs) were introduced by Eckmann et al. [25] to provide insight into periodic structures and clustering properties that were not apparent in the original time series. In particular the RP is based on the computation of the distance matrix between the reconstructed points in the phase space, that is, $\mathbf{s}_i = \{s(t), s(t - \tau), s(t - 2\tau), \ldots, s(t + (d_E - 1)\tau\}$,

$$d_{ij} = \left\| \mathbf{s}_i - \mathbf{s}_j \right\|, \tag{6}$$

where τ and d_E are the reconstruction parameters, that is, the time delay (the lag between data when reconstructing the phase space) and the embedding dimension (the dimension of the space required to unfold the dynamics).

The result is the array of distances that forms an $N \times N$ square matrix, $\mathbf{D}$, where N is the number of points under study. Any distance between any couple of points i, j can be represented by a pixel where the pixel coordinates are (i, j). Eckmann et al. [25] showed that if we darken the pixels which correspond to distances lower than a predetermined cutoff threshold, that is, a ball of radius ε centred at $\mathbf{s}_i$, and if we require $\varepsilon_i = \varepsilon_j$, then the plot is symmetric and presents a darkened main diagonal. The darkened diagonal corresponds to the identity line and the darkened points individuate the recurrences of the dynamical systems.

The RP method was then made more quantitative by Zbilut and Webber [24] who defined several measures of complexity to quantify the small scale structures in RP. These measures are based on the recurrence point density and the diagonal and vertical line structures of the RP (i.e., recurrence quantification analysis RQA).

A computation of these measures in small windows (submatrices) of the RP moving along the main diagonal yields the time dependent behaviour of these variables [32]. Studies based on RQA measures show that they are able to identify bifurcation points, especially chaos-order [33] and chaos-to-chaos transitions [34].

In order to quantify complexity of RPs, Zbilut and Webber [24] and Marwan et al. [35] proposed the following parameters.

(i) Measures Based on Recurrence Density. The percentage of recurrence, RR, is the percentage of darkened pixels in recurrence plot:

$$\mathrm{RR}(\varepsilon) = \frac{1}{N^2} \sum_{i,j=1}^{N} \mathbf{R}_{i,j}(\varepsilon), \tag{7}$$

where $\mathbf{R}_{i,j}(\varepsilon)$ is one if the state of the system at time i and the one at time j have a distance lower than ε and zero otherwise. RR is a measure of the density of recurrence points in an RP.

(ii) Measures Based on Diagonal Lines. Let $P(l)$ be the histogram of diagonal lines of length l. The ratio of recurrence points that form diagonal structures, longer than $l_{\min}$, is called the percentage of determinism (DET):

$$\mathrm{DET} = \frac{\sum_{l=l_{\min}}^{N} lP(l)}{\sum_{l=1}^{N} lP(l)}. \tag{8}$$

The percentage of determinism (DET) depends on the value of $l_{\min}$. If $l_{\min} = 1$, then DET = 100. If $l_{\min}$ is too big, the histogram $P(l)$ is likely to be sparse and, thus, the reliability of DET decreases.

A further RQA measure considers the length $L_{\max}$ of the longest diagonal line found in the RP, or its reciprocal, that is, the divergence ($Divergence = 1/L_{\max}$).

$L_{\max}$ and *Divergence* are related to the exponential divergence of the phase space trajectory. The faster the trajectory segments diverge, the shorter the diagonal lines are, and the higher the divergence is.

The entropy (ENTR) calculates the Shannon information entropy of the probability, $p(l)$, to find a diagonal line of length l in the RP. This probability is given by $p(l) = P(l)/N_l$:

$$\mathrm{ENTR} = -\sum_{l=l_{\min}}^{N} p(l) \ln p(l) \tag{9}$$

ENTR tries to capture the complexity of the diagonal lines in the RP. Uncorrelated noise will produce small ENTR values, indicating its low complexity.

TREND measures the paling recurrence points away from the central diagonal. It is a linear regression coefficient over recurrence point density of the diagonals parallel to the main diagonal as a function of the time distance between these diagonals and the main diagonal. It provides information about nonstationarity in the process. TREND is highly correlated to the size of the window [35].

(iii) Measures Based on Vertical Lines. We can find vertical lines in presence of laminar states in intermittence regimes. Let the total number of vertical lines of length v in RP be given by the histogram $P(v)$, then the ratio between the recurrence points forming the vertical structures and the entire set of recurrence points can be computed:

$$\mathrm{LAM} = \frac{\sum_{v=v_{\min}}^{N} vP(v)}{\sum_{v=1}^{N} vP(v)}. \tag{10}$$

This is the percentage of laminarity. The LAM is computed for those vertical lines of length v that exceed a minimal length $v_{\min}$. LAM represents the occurrence of laminar states in the system without describing the length of these laminar phases. LAM decreases if RP consists of more single recurrence points than vertical structures.

The average length of vertical structures is given by

$$\mathrm{TT} = \frac{\sum_{v=v_{\min}}^{N} vP(v)}{\sum_{v=v_{\min}}^{N} P(v)}, \tag{11}$$

which is called trapping time (TT). TT estimates the mean time that the system will be trapped at a specific state.

According to Marwan et al. [34], measures based on vertical lines are able to find chaos-chaos transitions, because these measures are close to zero in periodic dynamics; that is, the RP has only diagonal lines.

3. Results

The experimental results reported here were obtained from a culture continually monitored for 6 weeks. Spontaneous activity monitoring started at 2 DIV (day in vitro) and continued up to DIV 42. The chip was monitored for at least 30 minutes at least three times per week. Results refer to ten minutes of spontaneous activity starting from five minutes after the start of the electrophysiological recording. We waited five minutes for a more stable electrophysiological behavior.

Random spiking activity was evident even at a very early stage of network development; however only a low amount of channels were active and the spike frequency was very low (Figure 2) and unsynchronized.

After one week in culture (DIV 10) the number of active channels was higher and it was possible to record both local spiking and bursting. Until the third week, bursts were generated at low frequency (Figure 3) and low amplitude (i.e., the number of spikes belonging to the same burst) (see Table 1). Burst Duration is the parameter that showed less variation during time (see Table 1). If we consider the bursts at network level and consider a weighted burst duration taking into account the number of active bursting channels (mean burst duration $*$ number of busting channels), then the network was seen to be most active during the third week (Figure 3).

The mean interburst interval was variable if considered at single channel level, while the weighted interburst interval (interburst interval/number of active bursting channels) showed the lowest values during the third week in vitro

Table 1

MBR [burst/min]	DIV 10	13	14	15	16	17	20	21	22	23	28	29	30	42
mean	0.0100	0.0145	0.0103	0.0066	0.0122	0.0112	0.0115	0.0369	0.1184	0.1269	0.1083	0.0574	0.0358	0.0126
std	0.0097	0.0117	0.0042	0.0017	0.0040	0.0078	0.0061	0.0197	0.0946	0.0876	0.0568	0.0366	0.0213	0.0055
errstd	0.0069	0.0083	0.0021	0.0010	0.0023	0.0045	0.0022	0.0070	0.0315	0.0292	0.0201	0.0129	0.0075	0.0019
cv	0.9751	0.8094	0.4047	0.2554	0.3266	0.6953	0.5315	0.5335	0.7988	0.6904	0.5247	0.6375	0.5959	0.4373
Mean Burst Number [burst]	DIV 10	13	14	15	16	17	20	21	22	23	28	29	30	42
mean	6.50	9.50	6.75	4.33	8.00	7.33	7.63	24.38	78.11	83.78	71.50	37.88	23.63	8.25
std	6.36	7.78	2.75	1.15	2.65	5.13	4.07	13.05	62.43	57.83	37.54	24.11	14.07	3.65
errstd	4.50	5.50	1.38	0.67	1.53	2.96	1.44	4.61	20.81	19.28	13.27	8.52	4.97	1.29
cv	0.98	0.82	0.41	0.27	0.33	0.70	0.53	0.54	0.80	0.69	0.53	0.64	0.60	0.44
Burst Amplitude [spike]	DIV 10	13	14	15	16	17	20	21	22	23	28	29	30	42
mean	12.341	14.508	15.444	17.178	16.520	15.624	19.990	24.541	20.501	22.719	25.051	29.409	28.492	22.730
std	2.603	2.487	4.258	4.903	2.379	2.530	7.710	8.143	5.906	7.670	7.536	10.497	9.343	6.042
errstd	1.841	1.758	2.129	2.831	1.374	1.460	2.726	2.879	1.969	2.557	2.664	3.711	3.303	2.136
cv	0.211	0.171	0.276	0.285	0.144	0.162	0.386	0.332	0.288	0.338	0.301	0.357	0.328	0.266
Burst Duration [ms]	DIV 10	13	14	15	16	17	20	21	22	23	28	29	30	42
mean	172.123	222.367	172.589	149.433	141.071	133.695	104.138	144.781	109.939	117.365	133.208	188.610	168.383	155.034
std	55.611	6.741	58.799	29.039	38.274	30.332	40.088	31.857	25.810	25.205	22.030	42.591	40.396	29.909
errstd	39.323	4.767	29.400	16.766	22.098	17.512	14.173	11.263	8.603	8.402	7.789	15.058	14.282	10.574
cv	0.323	0.030	0.341	0.194	0.271	0.227	0.385	0.220	0.235	0.215	0.165	0.226	0.240	0.193
Weighted Burst Duration [ms]	DIV 10	13	14	15	16	17	20	21	22	23	28	29	30	42
mean	344.2455	3335.5	862.9438	747.1667	987.4975	802.1718	833.1023	2895.61	4947.242	5868.258	8525.282	4149.423	3536.049	930.2023
std	111.2215	101.1163	293.9954	145.1942	267.9214	181.9895	320.7046	637.1416	1161.467	1260.242	1409.888	936.9954	848.3061	179.4547
errstd	78.64545	71.5	146.9977	83.82789	154.6845	105.0717	113.3862	225.2636	387.1557	420.0806	498.4709	331.2779	299.9215	63.44683
cv	344.2455	3335.5	862.9438	747.1667	987.4975	802.1718	833.1023	2895.61	4947.242	5868.258	8525.282	4149.423	3536.049	930.2023
Mean Inter Burst Interval [ms]	DIV 10	13	14	15	16	17	20	21	22	23	28	29	30	42
mean	7737.76	12622.38	7295.26	8866.79	22472.71	13806.17	30529.00	19838.58	6854.69	11606.48	7557.16	22677.82	21249.86	8451.73
std	10942.84	17850.74	6365.85	7678.88	16692.84	14362.34	18160.49	9445.50	6587.63	9813.10	4118.10	8995.54	10881.04	7274.24
errstd	7737.76	12622.38	3182.92	4433.40	9637.62	8292.10	6420.70	3339.49	2195.88	3271.03	1455.97	3180.41	3847.03	2571.83
cv	1.41	1.41	0.87	0.87	0.74	1.04	0.59	0.48	0.96	0.85	0.54	0.40	0.51	0.86
Weighted IBI [ms]	DIV 10	13	14	15	16	17	20	21	22	23	28	29	30	42
mean	3868.88	841.4921	1459.052	1773.359	3210.387	2301.029	3816.1248	991.9288	152.3263	232.1296	118.0806	1030.81	1011.898	1408.621
std	5471.422	1190.05	1273.17	1535.776	2384.691	2393.723	2270.0611	472.2748	146.3917	196.2619	64.34534	408.8884	518.1445	1212.373
errstd	3868.88	841.4921	636.5849	886.6809	1376.802	1382.017	802.58782	166.9744	48.79723	65.42064	22.74951	144.5639	183.1918	428.6386
cv	3868.88	841.4921	1459.052	1773.359	3210.387	2301.029	3816.1248	991.9288	152.3263	232.1296	118.0806	1030.81	1011.898	1408.621

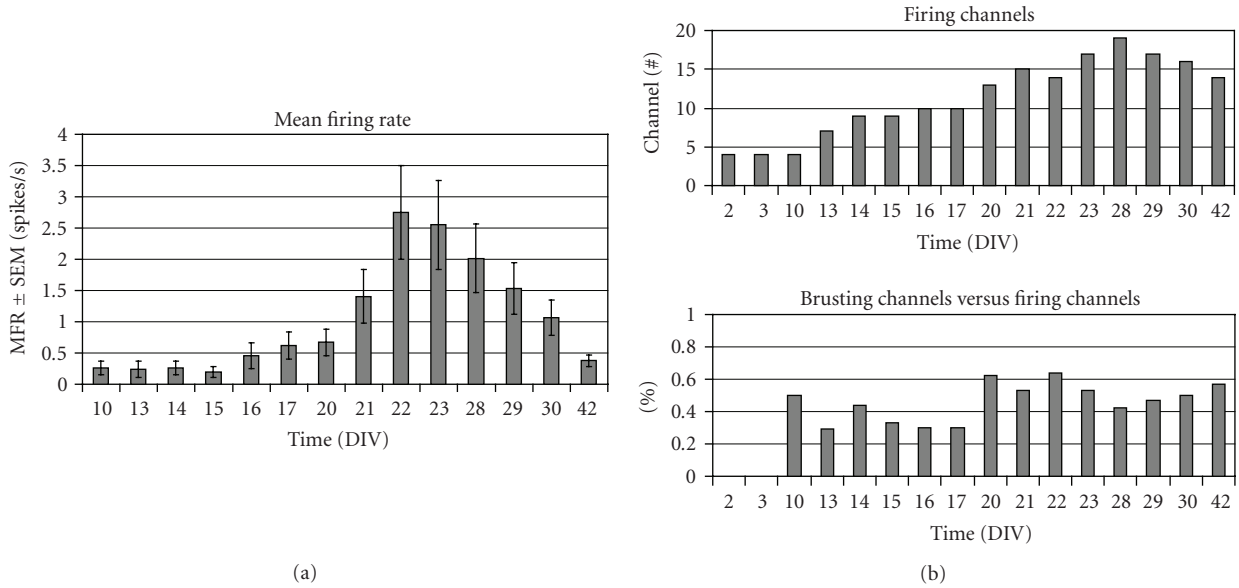


Figure 2: The mean firing rate (MFR) reports the average spike number during the investigation period (i.e., 10 minutes) and it is an indicator of the spontaneous neuronal activity. Random spiking activity can be recorded at a very early stage of the network development, that is, DIV 2 or 3. The network is still immature and spiking is slow, sporadic, and sparse (i.e., active channels are likely to be far and unsynchronized). After one week of in vitro culture (e.g., DIV10) it is possible to record more frequent activity. The network reaches its maturity at the third week in vitro when activity reaches its peak, and the maximum number of active channels is recorded (top right), and then it starts decreasing. Burst behavior appears during the second week of culture (bottom right). We monitored the network behavior up to DIV 42.

(Figure 3) and confirmed that the rate of burst events was higher in this period.

Before performing RQA it is necessary to determine the time delay and the embedding dimension that define the reconstructed state space and different methods have been developed for this purpose (the interested reader is referred to the books of Abarbanel [36] and Kantz and Shreiber [37]). In this work we used the first minimum of the average mutual information (AMI) to calculate the time delay [38] and the False Nearest Neighbours (FNNs) [39] to compute the embedding dimension. Since we were interested in comparing all recorded time series at different periods (DIV), we used average values and obtained $\tau = 2$ and $d_E = 5$.

Another important parameter for the recurrence quantification analysis is the cutoff radius ε. If ε is too small, almost no recurrence points exist; conversely if ε is too large, almost every point is close to every other point. Hence, a compromise for the ε value has to be found. Some "rules of thumb" can be applied [35].

(1) ε should not exceed 10% of the mean or the maximum phase space diameter [40].

(2) ε should be such that the recurrence point density in RP is approximately 1% [41].

(3) To avoid problems related to noise, ε has to be chosen such that it is five time larger than the standard deviation of the observational noise, that is, $\varepsilon > 5\sigma$ [42].

We applied the second rule to the first set of time series with enough points (at DIV 10) and we kept the same values for the subsequent recording periods. Even though in that case the rule is not true any longer, obviously it is necessary to use the same parameters for allowing the intercomparison between the spike interval time series.

An example of Recurrence Plot is presented in Figure 4 where both the x- and y-axis report the interspike interval (isi) after reconstruction; that is, each point is $\{\tau, isi\text{-}\tau, isi\text{-}2 * \tau, isi\text{-}3 * \tau, isi\text{-}4 * \tau\}$ because the spike trains embedded dimension and delay are 5 and 2, respectively, and dark points correspond to those for which the Euclidean distance is lower than the cutoff ratio (i.e., $\varepsilon = 1300$).

Only channels with sufficient activity (at least a spiking rate of 0.33 spikes/s) were considered for recurrence quantification analysis (Figure 5). The evolution of RQA parameters over the entire experiment was possible only on four channels (namely, 35, 58 67, and 74; channel name is in accordance with MCS MEA chip layout) that were highly active since the early days of the experiment. Linear correlation analysis of the spike interval time series showed that these four channels were low correlated. The maximum R value, found at DIV 42, was 0.26 ($P < .0005$) between channels 58 and 67. The shifting of the spike interval time series, using the cross correlation function, did not improve the maximum correlation between these channels.

To investigate whether there was high correlation between channels for which we did not have enough data to carry out recurrence quantitative analysis, the raw signals (one minute, i.e., 60000 points, five minutes after the start

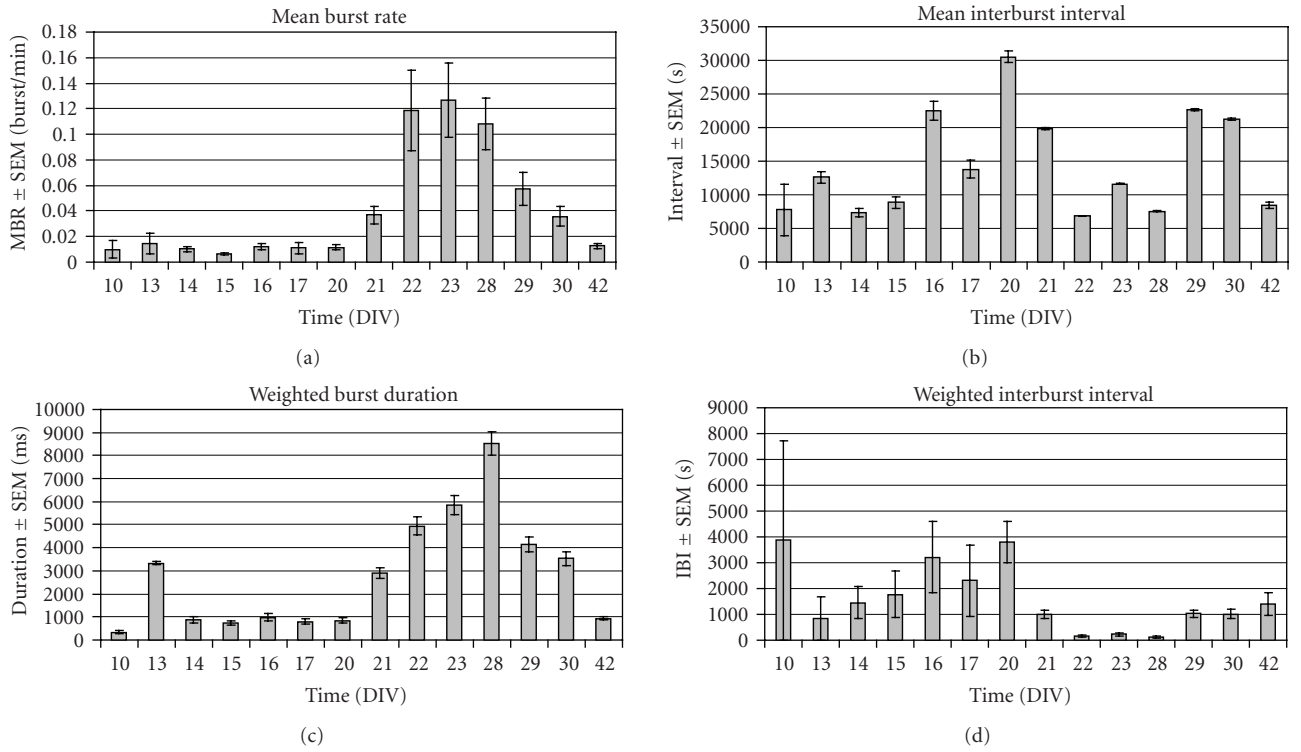


Figure 3: Mean Burst Rate, mean Burst Duration (weighed value), mean Interburst Interval, and the weighted mean Interburst Interval. Bursting behavior starts being recordable after the second week in vitro. Until the third week, bursts are generated at low frequency and low amplitude (i.e., the number of spikes belonging to the same burst). The mean burst duration shows less variation during network development. The weighted burst duration shows that the network is most active during the third week. The mean interburst interval (at channel level) is variable, while the weighted interburst interval (interburst interval/number of active bursting channels) shows the lowest values during the third week and confirms that the rate of burst event is higher in this period.

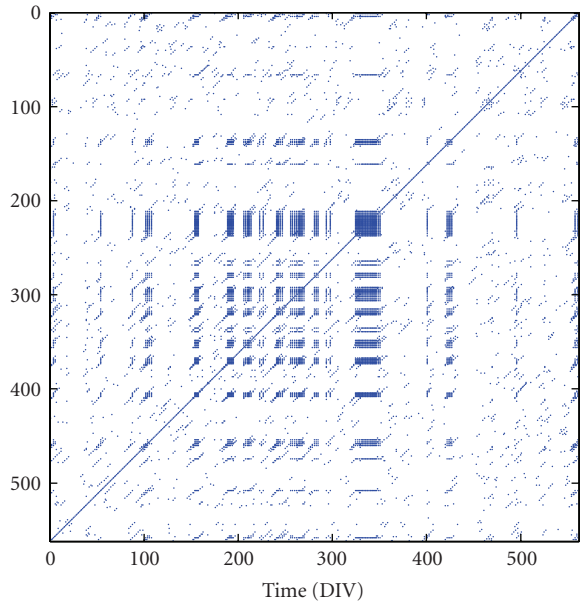


Figure 4: Recurrence plot of the interspike interval time series channel 35, DIV 15. Reconstruction parameters: $\tau = 2$ and $d_E = 5$, $\varepsilon = 1300$ (using Euclidean absolute distance). Both x- and y-axis report the interspike interval (*isi*) after reconstruction and dark points correspond to those for which the Euclidean distance is lower than the cutoff ratio.

of the recording) for all channels at DIV 9 and DIV 14 were analyzed. The maximum correlation coefficient between the 60 recorded channels at DIV 9 was 0.0786 between channels 46 and 57, whereas for DIV 14 was 0.26 between channels 76 and 77. In general, there was a slight increase in the average correlation coefficient between both days. However the values were quite low. Also in this case shifting the time series using CCF did not improve the correlation results.

4. Discussion

In vitro neuronal networks of cortical cells grown on Multielectrode Array (MEA) chips are becoming a widely used means to investigate basic neuronal properties (neural coding) [4–8], higher properties (learning and plasticity) [2, 9], and electrophysiological response to pharmacological manipulation [13, 17, 18].

During development, neuronal networks change in morphology and express different activity patterns [7, 20, 30]: activity that ranges from sporadic spiking (when the network presents immature types of synapses and a very low synaptic density) [43, 44] to complex synchronous activity packages (when neurons start exploiting "far" connections and the network manifests its collective behavior, i.e., network burst) [4, 7, 30, 45, 46].

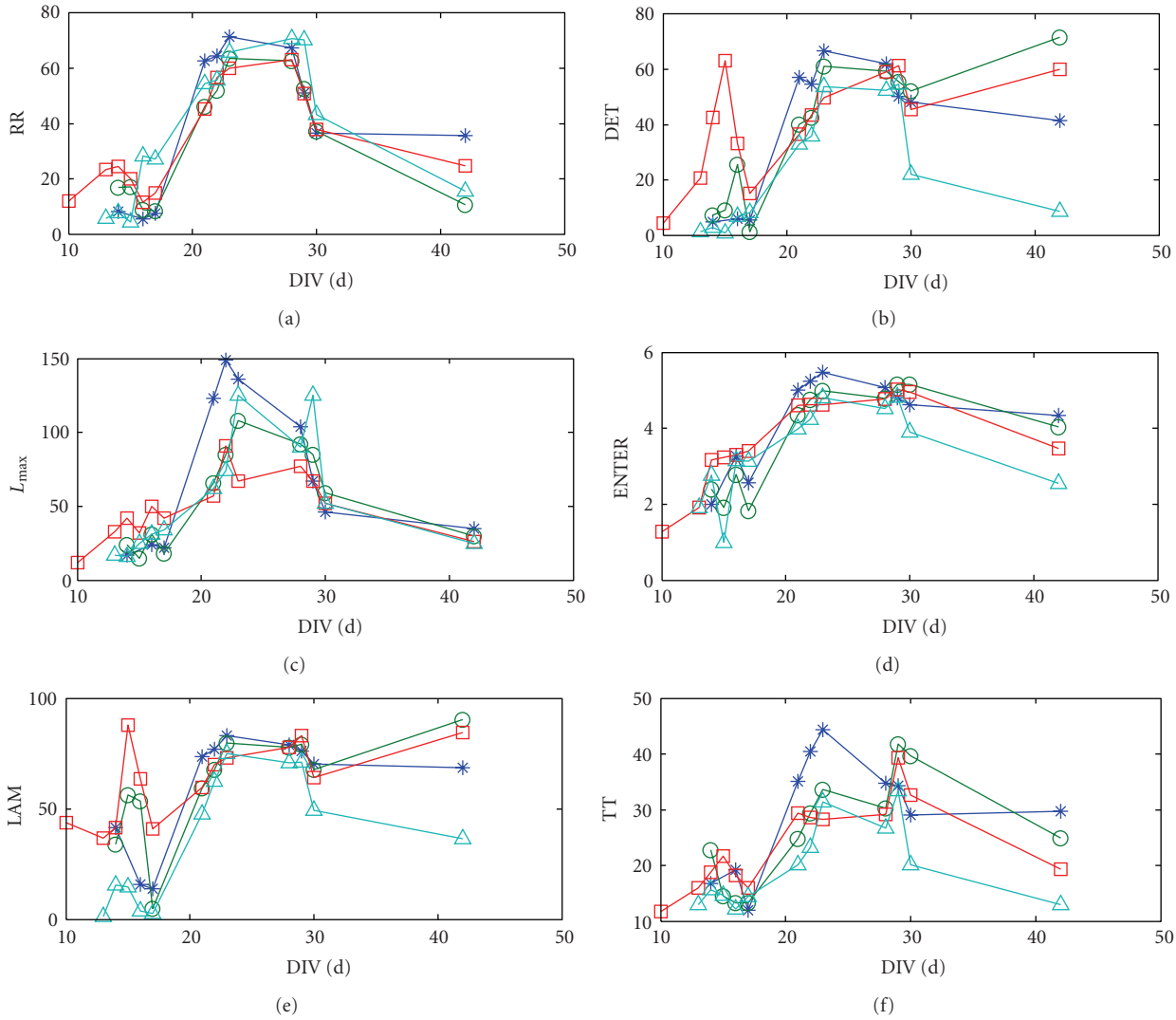


Figure 5: Percentage of recurrence (RR), determinism (DET), L_{max}, entropy (ENTR), and laminarity (LAM), and Trapping time (TT) for the channels no. 35 (∗), 58 (o), 67 (□), and 74 (Δ).

While spike trains can be accurately extracted from MEA recordings, for example, [21, 22], identification, description, and prediction of changes in electrophysiological dynamic are less accessible by means of standard processing techniques for neuronal electrophysiology. The introduction of Recurrence Plot [25] and Recurrence Quantification Analysis [24] has made available innovative tools for investigating in complex dynamic of nonstationary systems. As far as we know, this paper is the first attempt to investigate the changes in the spontaneous electrophysiological activity of a growing neuronal network of mouse cortical cells coupled to a multielectrode array chip in terms of Recurrent Quantification Analysis.

The structures of the RPs obtained point out at the occurrence of several regime shifts and transitions in which the system stays during a certain period in a certain state (dark zones) and after moves to another state [43]. This behavior occurs several times for each channel during the analyzed period at each DIV. Concerning the evolution of the MEA chip during the experiment, it is possible to observe that the percentage of recurrence (RR) follows the same trend in all channels: it sharply increases after DIV 20, and it reaches a plateau period between DIV 20 and DIV 30, when it starts decreasing. DET and LAM reveal a transition from less to more laminar states in channel 67 that occurs around DIV 15 and with a lower amplitude and slightly delayed at channel 58. This is an example of a transition between two states as it is shown by Marwan et al. [35] using the logistic map and DET and LAM to detect periodic-chaos/chaos-periodic transitions. The absence of correlation in stochastic processes produces RPs with none or very short diagonals which are the signature of a predetermined trajectory in phase space. Although DET does not really reflect the determinism of the system, an increase in DET can be related to an increase in close trajectories at different times. This is confirmed by the parallel increase in L_{max} which can be interpreted as an

increase in the mean prediction time or its inverse as the maximal positive Lyapunov exponent (assuming a chaotic system which is not the case here) as well as by the increase in the entropy (ENTR) which reflects the complexity of the diagonal lines in the RP. In addition to the detection of changes by LAM, the trapping time (TT) specifies the time in which a system is trapped in a specific state which also increases after DIV 20.

In addition, in order to verify that our model was consistent with what had already been reported in literature, we processed the recorded electrophysiology to describe the network dynamic in terms of spikes and bursts.

Our recording confirmed that the behavior of the network changed in the third week in vitro: bursting activity becomes more dominant, synchronized over all the active channels, and frequent (Figure 3) and changes do not answer to chaotic laws (Figure 5).

Besides pointing to the same results than we observed with more traditional approaches, RQA parameters provide further details about neuronal electrophysiological dynamics with a particular focus on the periodic structures and clustering properties that are difficult to determine in the original time series: the evolution of spontaneous neuronal electrophysiology is less nonstationary than one could expect. In other words it will be possible to design *in silico* models that consider and mimic those states and transitions and experimental design can benefit from models where it is possible to better predict the electrophysiological activity.

We believe that the application of RQA to in vitro electrophysiology is a very promising tool to improve the quality of the results as well as a read-out itself. In particular, the technique could be used to allow the intercomparison between results obtained from different MEA chips, providing a tool for toxicity testing standardization.

Acknowledgments

The authors wish to acknowledge Dr. Anna Price and Dr. Taina Paloosari for cell preparation and maintenance and Professor Maurice Whelan and Professor Alfons Schuster for their very helpful advice.

References

[1] A. R. Kriegstein and M. A. Dichter, "Morphological classification of rat cortical neurons in cell culture," *Journal of Neuroscience*, vol. 3, no. 8, pp. 1634–1647, 1983.

[2] Y. Jimbo, T. Tateno, and H. P. C. Robinson, "Simultaneous induction of pathway-specific potentiation and depression in networks of cortical neurons," *Biophysical Journal*, vol. 76, no. 2, pp. 670–678, 1999.

[3] G.-Q. Bi and M.-M. Poo, "Distributed synaptic modification in neural networks induced by patterned stimulation," *Nature*, vol. 401, no. 6755, pp. 792–796, 1999.

[4] E. Maeda, H. P. C. Robinson, and A. Kawana, "The mechanisms of generation and propagation of synchronized bursting in developing networks of cortical neurons," *Journal of Neuroscience*, vol. 15, no. 10, pp. 6834–6845, 1995.

[5] G. W. Gross, J. M. Kowalski, and B. K. Rhoades, "Spontaneous and evoked oscillations in cultured neuronal networks," in *Oscillations in Neural Systems*, D. Levine, V. Brown, and T. Shirey, Eds., pp. 3–29, Erlbaum, New York, NY, USA, 1999.

[6] J. M. Beggs and D. Plenz, "Neuronal avalanches in neocortical circuits," *Journal of Neuroscience*, vol. 23, no. 35, pp. 11167–11177, 2003.

[7] J. van Pelt, M. A. Corner, P. S. Wolters, W. L. C. Rutten, and G. J. A. Ramakers, "Long-term stability and developmental changes in spontaneous network burst firing patterns in dissociated rat cerebral cortex cell cultures on multielectrode arrays," *Neuroscience Letters*, vol. 361, no. 1–3, pp. 86–89, 2004.

[8] V. Pasquale, P. Massobrio, L. L. Bologna, M. Chiappalone, and S. Martinoia, "Self-organization and neuronal avalanches in networks of dissociated cortical neurons," *Neuroscience*, vol. 153, no. 4, pp. 1354–1369, 2008.

[9] E. Maeda, Y. Kuroda, H. P. C. Robinson, and A. Kawana, "Modification of parallel activity elicited by propagating bursts in developing networks of rat cortical neurones," *European Journal of Neuroscience*, vol. 10, no. 2, pp. 488–496, 1998.

[10] G. Shahaf and S. Marom, "Learning in networks of cortical neurons," *Journal of Neuroscience*, vol. 21, no. 22, pp. 8782–8788, 2001.

[11] D. Eytan, N. Brenner, and S. Marom, "Selective adaptation in networks of cortical neurons," *Journal of Neuroscience*, vol. 23, no. 28, pp. 9349–9356, 2003.

[12] A. Novellino, P. D'Angelo, L. Cozzi, M. Chiappalone, V. Sanguineti, and S. Martinoia, "Connecting neurons to a mobile robot: an in vitro bidirectional neural interface," *Computational Intelligence and Neuroscience*, vol. 2007, Article ID 12725, 13 pages, 2007.

[13] S. I. Morefield, E. W. Keefer, K. D. Chapman, and G. W. Gross, "Drug evaluations using neuronal networks cultured on microelectrode arrays," *Biosensors & Bioelectronics*, vol. 15, no. 7-8, pp. 383–396, 2000.

[14] A. Gramowski, D. Schiffmann, and G. W. Gross, "Quantification of acute neurotoxic effects of trimethyltin using neuronal network cultured on microelectrode arrays," *NeuroToxicology*, vol. 21, no. 3, pp. 331–342, 2000.

[15] E. W. Keefer, A. Gramowski, D. A. Stenger, J. J. Pancrazio, and G. W. Gross, "Characterization of acute neurotoxic effects of trimethylolpropane phosphate via neuronal network biosensors," *Biosensors & Bioelectronics*, vol. 16, no. 7-8, pp. 513–525, 2001.

[16] K. V. Gopal, "Neurotoxic effects of mercury on auditory cortex networks growing on microelectrode arrays: a preliminary analysis," *Neurotoxicology & Teratology*, vol. 25, no. 1, pp. 69–76, 2003.

[17] M. Chiappalone, A. Vato, M. Tedesco, M. Marcoli, F. A. Davide, and S. Marinoia, "Network of neurons coupled to microelectrode arrays: a neuronal sensory system for pharmacological applications," *Biosensors & Bioelectronics*, vol. 18, no. 5-6, pp. 627–634, 2003.

[18] A. Gramowski, K. Jügelt, S. Stüwe, et al., "Functional screening of traditional antidepressants with primary cortical neuronal networks grown on multielectrode neurochips," *European Journal of Neuroscience*, vol. 24, no. 2, pp. 455–465, 2006.

[19] M. Chiappalone, M. Bove, A. Vato, M. Tedesco, and S. Martinoia, "Dissociated cortical networks show spontaneously correlated activity patterns during in vitro development," *Brain Research*, vol. 1093, no. 1, pp. 41–53, 2006.

[20] D. A. Wagenaar, J. Pine, and S. M. Potter, "An extremely rich repertoire of bursting patterns during the development of cortical cultures," *BMC Neuroscience*, vol. 7, article 11, 2006.

[21] A. Maccione, M. Gandolfo, P. Massobrio, A. Novellino, S. Martinoia, and M. Chiappalone, "A novel algorithm for precise identification of spikes in extracellularly recorded neuronal signals," *Journal of Neuroscience Methods*, vol. 177, no. 1, pp. 241–249, 2009.

[22] A. Novellino, M. Chiappalone, A. MacCione, and S. Martinoia, "Neural signal manager: a collection of classical and innovative tools for multi-channel spike train analysis," *International Journal of Adaptive Control and Signal Processing*, vol. 23, no. 11, pp. 999–1013, 2009.

[23] E. N. Brown, R. E. Kass, and P. P. Mitra, "Multiple neural spike train data analysis: state-of-the-art and future challenges," *Nature Neuroscience*, vol. 7, no. 5, pp. 456–461, 2004.

[24] J. P. Zbilut and C. L. Webber Jr., "Embeddings and delays as derived from quantification of recurrence plots," *Physics Letters A*, vol. 171, no. 3-4, pp. 199–203, 1992.

[25] J. P. Eckmann, S. O. Kamphorst, and D. Ruelle, "Recurrence plots of dynamical systems," *Europhysics Letters*, vol. 4, pp. 973–977, 1987.

[26] P. Kalużny and R. Tarnecki, "Recurrence plots of neuronal spike trains," *Biological Cybernetics*, vol. 68, no. 6, pp. 527–534, 1993.

[27] V. Novák and J. Schmidt, "Changes of the inner time structures in the sequences of interspike intervals produced by the activity of excitatory and inhibitory synapses: simulation with Gaussian input processes," *Physiological Research*, vol. 46, no. 6, pp. 497–505, 1997.

[28] N. Marwan and A. Meinke, "Extended recurrence plot analysis and its application to ERP data," *International Journal of Bifurcation and Chaos*, vol. 14, no. 2, pp. 761–771, 2004.

[29] A. Bergner, M. C. Romano, J. Kurths, and M. Thiel, "Synchronization analysis of neuronal networks by means of recurrence plots," in *Lectures in Supercomputational Neuroscience: Dynamics in Complex Brain Networks*, P. Graben, C. Zhou, M. Thiel, and J. Kurths, Eds., pp. 177–191, Springer, Berlin, Germany, 2008.

[30] M. Chiappalone, A. Novellino, I. Vajda, A. Vato, S. Martinoia, and J. van Pelt, "Burst detection algorithms for the analysis of spatio-temporal patterns in cortical networks of neurons," *Neurocomputing*, vol. 65-66, pp. 653–662, 2005.

[31] S. J. Orfanidis, *Optimum Signal Processing: An Introduction*, Prentice-Hall, Englewood Cliffs, NJ, USA, 2nd edition, 1996.

[32] C. L. Webber Jr. and J. P. Zbilut, "Dynamical assessment of physiological systems and states using recurrence plot strategies," *Journal of Applied Physiology*, vol. 76, no. 2, pp. 965–973, 1994.

[33] L. L. Trulla, A. Giuliani, J. P. Zbilut, and C. L. Webber Jr., "Recurrence quantification analysis of the logistic equation with transients," *Physics Letters A*, vol. 223, no. 4, pp. 255–260, 1996.

[34] N. Marwan, N. Wessel, U. Meyerfeldt, A. Schirdewan, and J. Kurths, "Recurrence-plot-based measures of complexity and their application to heart-rate-variability data," *Physical Review E*, vol. 66, no. 2, Article ID 026702, 8 pages, 2002.

[35] N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, "Recurrence plots for the analysis of complex systems," *Physics Reports*, vol. 438, no. 5-6, pp. 237–329, 2007.

[36] H. D. I. Abarbanel, *Analysis of Observed Chaotic Data*, Springer, New York, NY, USA, 1996.

[37] H. Kantz and T. Shreiber, *Nonlinear Time Series Analysis*, Cambridge University Press, Cambridge, UK, 1997.

[38] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," *Physical Review A*, vol. 33, no. 2, pp. 1134–1140, 1986.

[39] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical construction," *Physical Review A*, vol. 45, no. 6, pp. 3403–3411, 1992.

[40] M. Koebbe and G. Mayer-Kress, "Use of recurrence plots in the analysis of time-series data," in *Proceedings of SFI Studies in the Science of Complexity*, M. Casdagli and S. Eubank, Eds., vol. 21, pp. 361–378, Adison-Wesley, Reading, Mass, USA, 1992.

[41] J. P. Zbilut, J.-M. Zaldívar-Comenges, and F. Strozzi, "Recurrence quantification based Liapunov exponents for monitoring divergence in experimental data," *Physics Letters A*, vol. 297, no. 3-4, pp. 173–181, 2002.

[42] M. Thiel, M. C. Romano, J. Kurths, R. Meucci, E. Allaria, and F. T. Arecchi, "Influence of observational noise on the recurrence quantification analysis," *Physica D*, vol. 171, no. 3, pp. 138–152, 2002.

[43] M. Ichikawa, K. Muramoto, K. Kobayashi, M. Kawahara, and Y. Kuroda, "Formation and maturation of synapses in primary cultures of rat cerebral cortical cells: an electron microscopic study," *Neuroscience Research*, vol. 16, no. 2, pp. 95–103, 1993.

[44] K. Muramoto, M. Ichikawa, M. Kawahara, K. Kobayashi, and Y. Kuroda, "Frequency of synchronous oscillations of neuronal activity increases during development and is correlated to the number of synapses in cultured cortical neuron networks," *Neuroscience Letters*, vol. 163, no. 2, pp. 163–165, 1993.

[45] S. Marom and G. Shahaf, "Development, learning and memory in large random networks of cortical neurons: lessons beyond anatomy," *Quarterly Reviews of Biophysics*, vol. 35, no. 1, pp. 63–87, 2002.

[46] J. van Pelt, I. Vajda, P. S. Wolters, M. A. Corner, and G. J. A. Ramakers, "Dynamics and plasticity in developing neuronal networks in vitro," *Progress in Brain Research*, vol. 147, pp. 171–188, 2005.

Hindawi Publishing Corporation
Advances in Artificial Intelligence
Volume 2010, Article ID 392868, 10 pages
doi:10.1155/2010/392868

Review Article

Simulation of Human Episodic Memory by Using a Computational Model of the Hippocampus

Naoyuki Sato[1,2] and Yoko Yamaguchi[2]

[1] *Department of Complex Systems, School of Systems Information Science, Future University—Hakodate, 116–2 Kamedanakano, Hakodate, Hokkaido 041–8655, Japan*
[2] *Laboratory for Dynamics of Emergent Intelligence, RIKEN Brain Science Institute, 2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan*

Correspondence should be addressed to Naoyuki Sato, satonao@fun.ac.jp

Received 25 August 2009; Accepted 2 November 2009

Academic Editor: Alfons Schuster

The episodic memory, the memory of personal events and history, is essential for understanding the mechanism of human intelligence. Neuroscience evidence has shown that the hippocampus, a part of the limbic system, plays an important role in the encoding and the retrieval of the episodic memory. This paper reviews computational models of the hippocampus and introduces our own computational model of human episodic memory based on neural synchronization. Results from computer simulations demonstrate that our model provides advantage for instantaneous memory formation and selective retrieval enabling memory search. Moreover, this model was found to have the ability to predict human memory recall by integrating human eye movement data during encoding. The combined approach between computational models and experiment is efficient for theorizing the human episodic memory.

1. Introduction

In 1982, Marr [1] argued the importance of computational theory for understanding the information processing in the brain and presented "three levels at which any machine carrying out an information-processing task must be understood (p. 25)" as follows.

(i) *Computational theory.* What is the goal of the computation, why is it appropriate, and what is the logic of the strategy by which it can be carried out?

(ii) *Representation and algorithm.* How can this computational theory be implemented? In particular, what is the representation for the input and output, and what is the algorithm for the transformation?

(iii) *Hardware implementation.* How can the representation and algorithm be realized physically?

As an example, consider a brain function of "associative memory of visual stimulus A and B." In the level of the computational theory, it is asked what relationship between stimulus A and B results in the memory; for example, a correlation coefficient of presentation sequences of A and B will indicate a strength of association between A and B. On the level of representation and algorithm, the visual stimuli can be understood by an N-dimensional binary vector pattern where an overlap between stimulus A and B will be an important parameter for their association. A correlation of vector patterns will be represented by a $N \times N$ matrix denoting the connection strength between ith and jth units and the matrix will be formed by the Hebb rule with a repetitive presentation of the stimulus. On the level of hardware implementation, it is asked what neuronal activation and dynamics are used for implementing the above algorithm; for example, neuronal synchronization dynamics might play an important role in the synaptic plasticity under the Hebb rule. The above three levels of understanding can be separately considered, while all levels are necessary for a complete understanding of the function of visual associative memory.

In the case of the memory function, the main problem is how to theorize the memory function; for example, a simple record and playback scenario is not perfect and there is a problem on the level of the computational theory. For example, how does the brain organize experiences into memory that can be applicable to novel situations? The models of artificial intelligence focus on the level of computational theory, and the models of neuroscience further address the representation-algorithm level and the implementation level. The final goal is to have a computational theory of the memory function that can be common between artificial intelligence models and neuroscience models, while the neuroscience models are advantageous in the theorization of the memory function in cooperation with experimental evidences.

This paper reviews computational models of human episodic memory that are associated with the personal history and contextual information of the environment. Section 2 summarizes the functional aspects of the episodic memory and the contribution the hippocampus makes to this memory. Section 3 investigates computational models of the hippocampus. Sections 4 and 5 describe our computational model of the human episodic memory and its application to the simulation of the human memory by using eye movement data. Section 6 summarizes the paper and provides future directions.

2. What Is Episodic Memory?

2.1. Episodic Memory in the Hippocampus. The bilateral hippocampal damaged patient H.M. [2] clearly demonstrated a significant role of the hippocampus in the formation of new memories. Patient H.M. had a normal IQ score and normal language skills and procedural memory, while H.M. had great difficulty in recognizing the current location and time (e.g., events where H.M.'s own conduct had occurred several minutes earlier). This kind of memory is categorized as "episodic memory" [3] and known to be maintained by the hippocampus. Even if damage to the hippocampus occurs in childhood, patients with damage to the hippocampus show difficulty in the formation and maintenance of the episodic memory [4]. This is one of the reason why the hippocampus is considered an essential structure for maintaining episodic memory.

In 1983, Tulving [5] proposed that the episodic memory can be modeled by an association of information among "what," "where," and "when." In relationship to this proposal, a simplified version of the episodic memory model, an object-place association model, is often used in experiments involving humans [6–9], monkeys [10, 11], and rats [12]. In a task, participants are asked to remember identities and locations of objects on a table during a short period. After a short delay period, the participants are asked to retrieve identities of the objects and reconstruct the arrangement of the objects. When the hippocampus is damaged, patients have great difficulty in performing such task [6–9]. This evidence suggests that the hippocampus uses the object-place representation as part of the episodic memory.

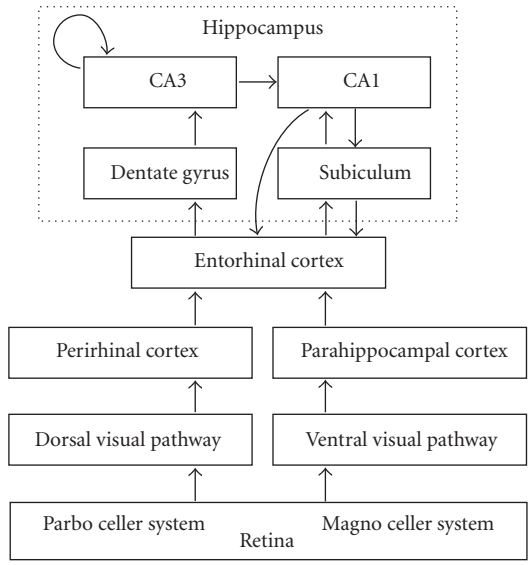


Figure 1: Structure of the hippocampus and adjacent regions.

Anatomically, the hippocampus is known to receive a convergent projection of the information of object and space through the parahippocampal region [13] (Figure 1). The object information starts from the perbocellar system with color information. It then forms a ventral visual pathway, converges to the perirhinal cortex in the parahippocampal region, and then enters the hippocampus. The space information starts from the magnocellar system with a wide visual field and then forms a dorsal visual pathway, converges to the parahippocampal cortex, and enters the hippocampus. This anatomical structure is reasonable in relationship to the object-place memory of the hippocampus; so the object-place memory paradigm is a good tool for evaluating the neural mechanism of the episodic memory in the hippocampus.

2.2. Neural Dynamics of the Hippocampus. The hippocampus is part of the limbic system and characterized by a closed loop circuit [14] (Figure 1). The cortical input enters from the superficial layer of the entorhinal cortex and is then sequentially transmitted to the dentate gyrus, the CA3 and the CA1 regions, and returns back to the deeper layer of the entorhinal cortex. The hippocampus has been considered to implement an associative memory [15] and the CA3 region including massive recurrent connections is considered to be a major network of the hippocampal memory [16]. These structures are similar between the hippocampus of rodents and primates, so that a common principle of the memory function is expected [17].

In the CA1 and CA3 regions of the rat hippocampus, many neurons were found to show a selective activation during passing through a specific portion of the environment [18]. Such neurons are called "place cells" and are also found in monkeys [19] and humans [20]. The hippocampus is known to represent a map of the environment called "cognitive map," and therefore the place cells are considered

Table 1: Classification of the models of the hippocampus with input overlap and asymmetry of connection weights.

		Overlap of input vector		
		Discrete	Continuous	Discrete and continuous
CA3 connection weights	Symmetric	Associative network	Cognitive map network	Event-space associative network
	Asymmetric	Sequence memory network	Cognitive map network for navigation	Object-place hierarchical cognitive map network

a neuronal basis of the cognitive map [18]. In the case of monkeys, other neuronal selectivity is further reported. This selectivity is called "view cells" and encodes information about the spatial location at which a monkey is looking into the environment [10]. Interestingly, the activation is not determined by a specific visual feature. Thus, this activity is considered a result of the information integration among body motion, head direction, and self-location. Both place cells and view cells are considered to contribute to the spatial navigation in the environment.

In the rat hippocampus, the local field potential of 4–12 Hz (theta-band) oscillations appears prominently during moving in the environment and the place cell firings are known to be synchronized with the local field potential (LFP) theta [21]. Moreover, the phase of the firing with the LFP theta cycle is found to gradually advance as the rat passes through the environment [22]. This phenomenon is called "theta phase precession." Each place cell firings have different phases according to their entering time of the place field, which then results in a sequential place cell firing in a theta cycle that represents a temporally compressed sequence of place field activation [23]. More important, the time difference of the sequential firings agrees with an asymmetric time window of a modified Hebb rule [24, 25]. The firing pattern of the theta phase precession is expected to contribute to the formation of the cognitive map in the hippocampus.

3. Computational Models of the Hippocampus

In this section, we review models of the hippocampus by using a classification with input overlap and the asymmetry of CA3 connection weights (Table 1). The CA3 region has been considered a major region for maintaining the hippocampal memory [16]; so the classification is applicable for many models of the hippocampus. Although each model has its own advantages in specific problems and the use of the CA3 network highly depends on the dynamics of units and other adjacent systems, the classification is meaningful for looking over the function and dynamics of the hippocampus models.

The CA3 region is regarded as a center of the memory function and modeled as an associative network [26–29]. In the associative network, multiple vector patterns can be stored into the CA3 connection weights and one of the patterns can be recalled through mutual activation among units. The memory encoding is implemented by the Hebb rule in which the connections between simultaneously activated units increase. The recall is implemented by mutual unit activations through the connections, where the stored vector pattern can be self-organized and completed from an initial activation of a part of the vector pattern. The performance of the pattern completion becomes better when the overlap of the arbitrary pairs of vector patters is small and random. In agreement with this model, an experimental study involving rats demonstrated that the CA3 region is essential for pattern completion [30].

In the above associative network, the connection weights are symmetric, while the connection weights can be asymmetric according to the Hebb rule with an asymmetric time window [25]. Models with asymmetric CA3 connections revealed that a sequence of vector patters can be stored and recalled with mutual unit activations [31, 32]. It is important that these models can deal with the information of the time with asymmetric connections. Moreover, the temporal compression with phase precession has been demonstrated to have an advantage in the sequence memory formation [33–35].

In cognitive map theory [18], the map of the environment is represented by a network of place cells, where population activity of the place cells gradually changes as the rat passes through the environment. Such neuronal activation was modeled by a "continuous attractor network" [36], where the overlap of the positional input vector is given by a function of spatial geometry (e.g., input vectors of neighboring positions have a large overlap and input vectors of distant positions have a small overlap) and CA3 connections are given by symmetric connections. This model demonstrated that the population activity of place cells representing a location in an environment can be self-organized from an initial state of random unit activation. When asymmetric connections are introduced to the CA3 network, the models are further able to show the ability of spatial navigation where the sequential activation to a goal location is evoked from arbitral location in the environment [37–40].

A combination of discrete and continuous input vectors was also used to represent an environment consisting of objects. Rolls et al. [41] proposed a unified network between discrete and continuous attractors, where both discrete and continuous patterns are associated with symmetric connections that implement the pattern completion including both discrete and continuous patterns. Byrne et al. [42] proposed

a network model including the medial temporal system and the parietal system where the CA3 region represents both object and space information. Interestingly, this model implements the mental imagery of navigation by integrating movement signals into a proper population activation of place cells. The authors proposed a model of a cognitive map for object-place associations [43]. The overlap of input is similar to the above models, while asymmetric connections according to phase precession are introduced. The model can store multiple object-place associations in a hierarchical structure of this network with asymmetric connections that represent inclusion relationships among visual features. In such a structure, a set of object-place associations can be recalled sequentially.

Let us consider the relationship between the models and the episodic memory. According to Tulving's proposal [5], the episodic memory is modeled by an association of information among "what," "where," and "when." The models with asymmetric connection with discrete input and continuous input can deal with "when" information. On the other hand, the models with discrete-continuous input can deal with "what-where" associations that are often used as an experimental model of the episodic memory in animals [10–12]. In order to understand comprehensively the episodic memory, it seems to be necessary to investigate the integration between "when" and "what-where" in future studies. In that case, the dynamics of phase precession can be a strong candidate for integrating "what-where-when," because the model already demonstrated to be able to encode each "when" and "what-where" information. In the next section, we review a model of "what-where" association by using theta phase precession.

4. A Computational Model of the Episodic Memory Based on Neural Synchronization

In this section, a computational model of the episodic memory based on neural synchronization of phase precession [43] is reviewed.

4.1. Representation of Object and Scene Information. Figure 2(a) shows the information flow of the model that follows experimental proposals [13, 17]. Retinal information produces two visual pathways that converge on the parahippocampal region, in which the perirhinal cortex receives object information in the ventral visual pathway and the parahippocampal cortex receives space information in dorsal visual pathways. Subsequently, the object and space information converge on the hippocampus that stores object-place associations in the connection weights in the CA3 network.

In the model, a one-dimensional environment with a grayscale pattern including multiple objects with different colors was assumed (Figure 2(b)). The object information is represented by color features at the center of the visual field that produces a discrete vector pattern. The scene information is represented by spatial frequency components of an object-centered gray-scale pattern in a 120 degree-wide visual field that represents a spatially continuous vector pattern. In these representations, the scene information plays an essential role in the binding among multiple object-place associations in an environment; that is, overlap between scene information works as a tab for combining two scenes and their orientation and distance can be obtained by calculating a shift of the two visual patterns (Figure 2(c)).

Multiple object-place associations are encoded by a sequence with "saccadic" eye movement where one of the objects is successively caught at the center of the visual field. Since a size of saccade is found less than 10 degree [44], the scene vector pattern will have a large overlap with a subsequent scene vector and the object vector pattern will drastically change at a subsequent saccade. Thus, the eye movement produces a sequence consisting of discrete and continuous vector sequences. It should be noted that the eye movement sequence was assumed to "randomly" catch the object and it is not like the scan path theory [45] in which a stereotyped eye movement repeats when seeing a picture.

4.2. Memory Encoding Based on Neural Synchronization. The visual input sequence of object and scene information is stored into the CA3 connection weights by using theta phase precession that has a computational advantage in the encoding of the sequence [34, 35], the spatiotemporal patterns [46], and the map of the environment [40].

In the model, the visual input sequence is translated into a phase precession pattern at the entorhinal cortex, where each neural unit shows an oscillatory activity according to an excitatory visual input and its oscillatory frequency is assumed to gradually increase when receiving a persistently excitatory input [34]. Phase-locking dynamics between the units' oscillation and a global theta rhythm results in a gradual phase advancement of the units' oscillation with the theta cycle. The CA3 region receives the pattern that is stored into the CA3 connection weights according to the Hebb rule with an asymmetric time window. The connections between a simultaneously activated object and scene units at each eye fixations can increase, while an additional effect appears in the phase precession; earlier and persistently activated units have activations at later phases and other intermittently activated units can only have activations at earlier phases. Subsequently, the modified Hebb rule with an asymmetric time window results in the formation of asymmetric connections from persistently activated units to intermittently activated units (Figure 2(d)).

The activation duration of each unit can vary according to the eye movement sequence, while on average the larger overlaps of scene input vectors would produce a longer activation of scene units than object units. The resultant network appears to include unidirectional connections from scene to object units from a random eye movement sequence. In cooperation with symmetric autoassociative connections, the network is characterized by a layered structure of symmetric connections and interlayer asymmetric connection from scenes to object units (Figure 2(e)). We refer to

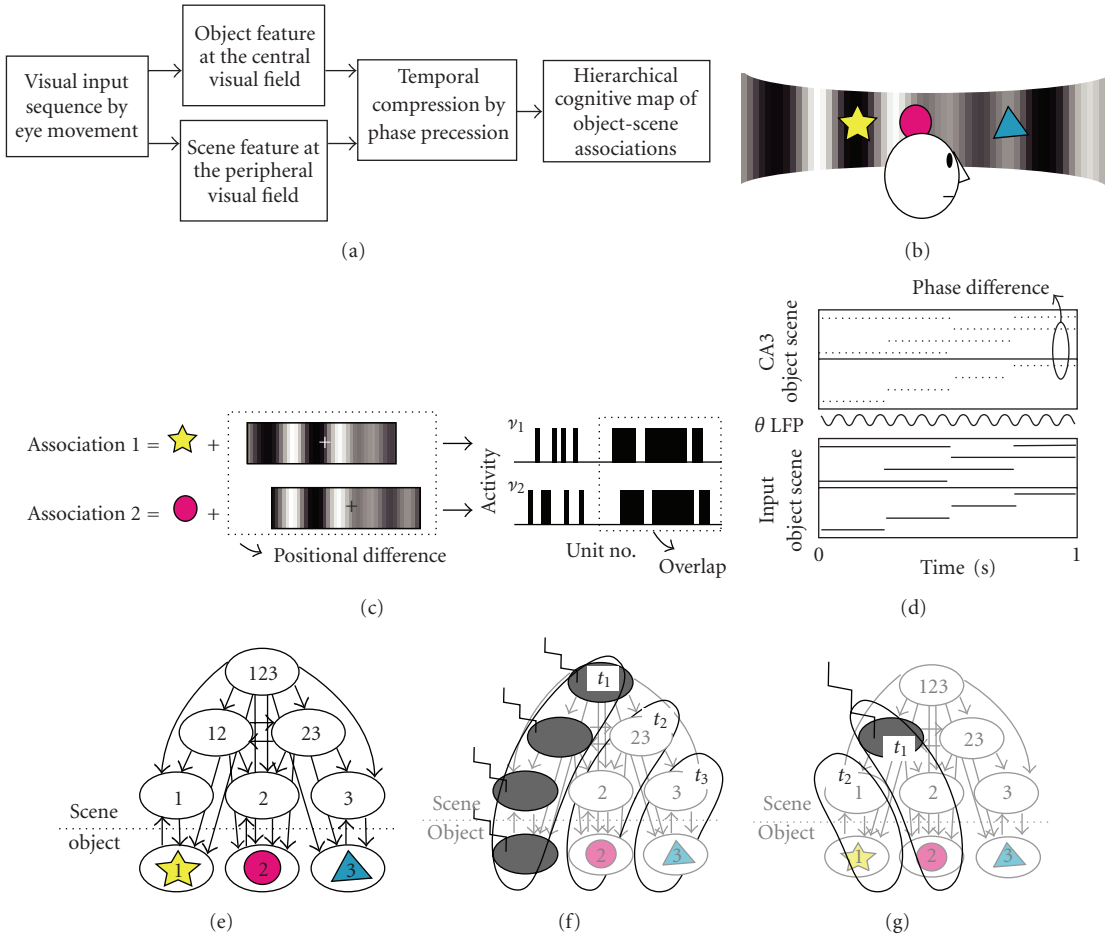


FIGURE 2: Model of object-place association by using neural synchronization [43]. (a) Basic information flow of the model. (b) One-dimensional visual environment including three colored objects and grayscale background. (c) Neuronal representation of object and scene information. An association is represented by a combination of object feature and object-centered background that are encoded by a set of discrete and continuous vector patterns. (d) Encoding of visual input sequence with theta phase precession. In the plot, black bars indicate input and CA3 activities. Difference in input durations between objects and scenes results in a robust object-scene phase difference. (e) A schematic graph of resultant CA3 connection weight. A node denotes a population of CA3 units having same selectivity of visual features. The numbers in a node denote the selectivity of scene units of the node. For example, "23" indicates units activated during fixating on objects 2 and 3, while these are not activated during fixating on an object 1. Lines with an arrowhead indicate directional connections between unit populations. ((f), (g)) Retrieval in the hierarchical network. Gray nodes denote units that are activated initially, and t_1, t_2, and t_3 denote the time after giving initial activations. A set of object-place associations is sequentially activated as an activity propagation in the network.

the structure as a hierarchical cognitive map for object-place associations [43]. Interestingly, the network represents "inclusion" relationships of visual features in multiple spatial scales and it can be organized in an encoding period of several seconds. This structure is expected to contribute to an efficient memory storage of the global environment, as demonstrated in psychological studies of human cognitive maps [47, 48].

4.3. Memory Retrieval. The hierarchical cognitive map of object-place associations has several advantages for memory retrieval. When the CA3 units of an object-scene association are activated as an initial cue, the units of other associations are automatically activated through the CA3 recurrent connections. Since the network is organized asymmetrically from the top to the bottom layers, the activity propagation accordingly occurs from the top to the bottom layers (Figure 2(f)). During the activity propagation, the asymmetric connections between object and scene units support a synchronized activation between corresponding object and scene units. Then a set of object-place associations is recalled where individual object-place associations appear in a sequence [43]. Such a simultaneous activation of multiple associations is an advantage of the hierarchical network.

When a small part of the hierarchical network is activated as an initial cue, an interesting retrieval appears [49]. By using a global inhibition of the network, initial activation of units at the top layer results in a sequential retrieval of all object-place associations. On the other hand, the

activation of units at the middle layer results in a constrained retrieval where multiple associations including the visual feature of the initial cue only are activated (Figure 2(g)). Such a selective retrieval will relate to the memory search, where course scene information can evoke a set of possible object-place associations. In the network, asymmetric connections are formed to represent inclusion relationships of visual features; therefore any initial cue of a partial feature is considered to evoke a set of possible object-place associations. This property is important for understanding the memory search mechanism in the hippocampus that maintains a large memory content.

4.4. Experimental Support for the Model. The model predicts a positive correlation among saccade rate, EEG theta power, and memory recall performance. We have evaluated the prediction by using brain signal analysis of human participants (see [50] for review). In the EEG measurement during object-place memory encoding, the EEG 7.0 Hz power and coherence at central region showed to significantly correlate with subsequent successful recall [51]. The coherence between EEG theta power and saccade rate was also found to correlate with the subsequent successful recall [52]. These results indicate that the EEG theta-related neural dynamics plays an important role in the memory encoding with eye movement.

Moreover, the results of an EEG-fMRI simultaneous measurement showed that scalp EEG theta power during object-place memory encoding is correlated with BOLD responses in the medial prefrontal, medial posterior, and right parahippocampal regions [53]. This result did not show a direct link between the hippocampus and theta dynamics, but it does suggest that the medial temporal memory system, consisting of the hippocampus and the parahippocampal region, uses theta dynamics for memory encoding.

5. Simulation of the Episodic Memory Based on the Computational Model

It has been shown that memory recall performance of object-place associations can be predicted by either EEG theta power [43] or BOLD responses [55] during encoding. This fact leads to the prediction that the computational model integrating experimental data could have an excellent ability in the prediction of a subsequent recall. At the same time, it produces a good validation of the computational model; for example, if the brain really uses the dynamic of this model, then the model should have predictability; otherwise the model will be rejected.

This section now reviews an application of the computational model of object-place associations to the prediction of human subsequent recall by using eye movement data during encoding [54, 56]. In the analysis, the eye movement data of our previous report [51] were used that consists of 350 trials of object-place memory encoding from eleven subjects. During the task, the participants were asked to remember identities and locations of four objects in a 3×3 grid during 8 seconds. Afterwards, the participants were asked to reconstruct the arrangement of the objects by using a mouse on the display after a 10-second delay period that contains a secondary task of randomly targeted saccades to inhibit the memory rehearsals (data were also used to calibrate eye cameras). Both temporal parameters are sensitive to participants' correct recall rates, while the temporal parameters were determined to make the correct recall rate at around 50%. Each participant performed 30 trials of the encoding task. Trials which failed to record any eye movement were discarded in the analysis. The interobject saccade rate of remaining data appeared in normal range (579.7 milliseconds) and almost all fixations appeared on each object.

To apply the model to the experimental data analysis, the visual features of the model were adapted to include object shapes used in the experiment and multiscaled receptive fields for location of eye fixation. In the model, a visual input at a fixation location was represented by 9 object units and 36 scene unit activations. A sequence of eye movement was translated to a visual input sequence and is stored into a $45 \times \times 45$ CA3 connection matrix by using phase precession and the Hebb rule with an asymmetric time window and then connection matrices were varied for trials using identical stimulus (it should be noted that the eye movements were not stereotyped).

In the statistical procedure, the individual correlation coefficient of a predictor and human recall were calculated (Figure 3(a)) and these were averaged over participants. In order to evaluate the importance of the hierarchal structure in the recall prediction, following four computational predictors and three traditional experimental predictors were used. The computational predictors are (1) the connection weight sum, $\sum_i \sum_j w_{ij}$, (2) the asymmetric connection weight sum, $\sum_i \sum_j |w_{ij} - w_{ji}|$, (3) the hierarchical connection weight sum, $\sum_i \sum_j (h_i - h_j)(w_{ij} - w_{ji})$, and (4) computational recall evoked by an initial input to the top layer, where w_{ij} denotes a CA3 connection weight from the jth to the ith unit and h_i indicates the hierarchy of the ith unit in the hierarchical network. The experimental predictors are (5) blink rate, (6) saccade rate, and (7) EEG 7 Hz power at a central region. The forthcoming results section will discuss the meaning of these predictors in more detail.

The results are shown in Figure 3(b). Only three predictors, the hierarchical connection weight sum ($r = 0.1154$, $P = .0309 < .05$), the computational recall ($r = 0.1183$, $P = .0269 < .05$), and EEG theta power ($r = 0.1226$, $P = .0178 < .05$), were found to significantly correlate with the human recall. This indicates that the computational model receiving eye movement data has similar predictability with the EEG theta power. On the other hand, the experimental predictors of the blink rate ($r = -0.0293$, $P = .5849 > .05$) and the saccade rate ($r = 0.0992$, $P = .0638 > .05$) did not show a significant correlation with the human recall. These results indicate that the computational network somehow extracted a memory-dependent component from the eye movement during encoding. Together with the result of no significant correlation between other computational predictors and the human recall (the sum connection weights, $r = 0.0802$, $P = .1343 > .05$; the asymmetric connection weights, $r = 0.0984$,

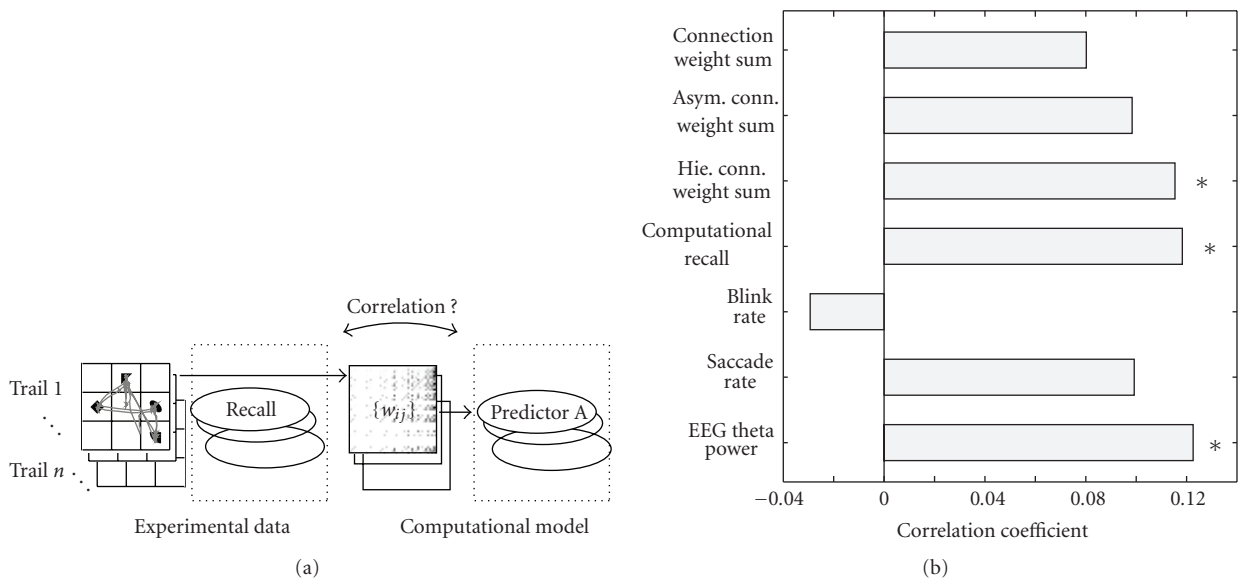


Figure 3: Prediction of human subsequent recall by using the computational model and other behavioral parameters [54]. (a) A computational model-based analysis of eye movement data during encoding. Computational predictors based on the computational model storing the visual input sequence were compared with human subsequent recalls. (b) Correlation coefficients between the predictors and human subsequent recall. Asterisks denote significant correlations ($P < .05$).

$P = .0659 > .05$), the hierarchal structure itself is considered to be an important factor to predict the human recall.

From a computational point of view, the hierarchical connection weight sum can increase when the eye movement occurs to evenly catch neighboring and distant pairs of objects with a saccade interval of more than 250 milliseconds. In order to experimentally evaluate that point, we further tested other experimental predictors (e.g., the variance of fixation duration of individual objects, etc.), while we have not found a suitable experimental predictor (data not shown). It is considered that more complicated memory-related components, such as order of fixated objects, might be extracted by the model. These results suggest that the model dynamics exists in the human brain and work during object-place memory encoding and retrieval.

6. Summary and Future Directions

The computational models of the episodic memory in the hippocampus and a simulation of the human episodic memory based on a computational model are reviewed. The hippocampus has a clear functional role in the episodic memory with a beautiful anatomical organization; thus many models have been proposed. A computational model of the hippocampus based on neural synchronization of phase precession [43] produces neural dynamics of the episodic memory formation that is characterized by the one trial learning of multiple object-place associations and the selective retrieval realizing memory search. The model was further applied to experimental data analysis, where a neural network organized by human eye movement data was found to have the ability to predict human object-place memory recall. This suggests that the model's dynamics exists in the brain and works during memory encoding and retrieval. This also indicates the importance of bridging between the computational model and experimental studies for theorizing the human episodic memory (Figure 4). In the following section, questions for future research are discussed.

6.1. Neural Mechanism of Memory Retrieval. Section 4.3 identified that the retrieval of the computational network is constrained by the initial cue, while the definition of the cue is a problem for understanding the human episodic memory. The initial cue could be modeled in the context of a situation, such as task demand and intentional effort. Such context information is proposed to be represented in the prefrontal region [57, 58], and the framework of the computational model of the hippocampus is developing to include the prefrontal and other regions. Recent simultaneous recordings of the prefrontal region and the hippocampus in rats are becoming possible [59], and these data give insight to a new framework of the episodic memory.

6.2. Representation of the Episodic Memory. The representation of the episodic memory is still an important issue. In the computational models of cognitive maps in rats, the representation of cortical inputs to the hippocampus has been discussed. Hartley et al. [60] proposed a boundary vector cell (BVC) as a component of the cortical inputs leading the place cell properties. De Araujo et al. [61] proposed an angular combination of visual cues and showed that the size of the visual field is critical for forming the place and view cell properties in rats and primates. Among Marr's three levels of understanding, the level of representation and algorithm is key to a combination between computational

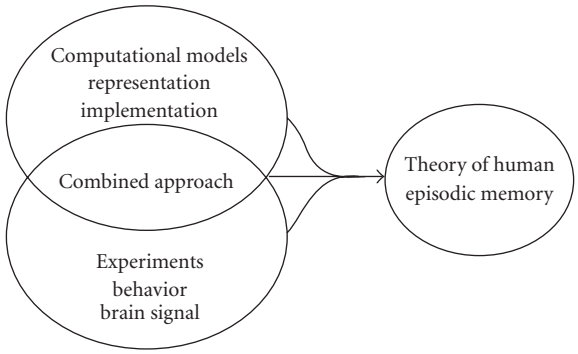


Figure 4: Computational model, experiments, and their combined approach are necessary for theorizing the human episodic memory.

models and experiments. Although there are few computational proposals on the representation of human episodic memory, virtual maze experiments in humans [62, 63] might produce essential data for linking human experience and episodic memory. Moreover, it will be a great step toward understanding the human episodic memory outside of the laboratory.

In addition to the above discussion, it is necessary to integrate "what" information with "what-where" association models as discussed. In Section 4, we reviewed a computational model of object-scene association by using theta phase precession. The model was also shown to be able to encode and recall the temporal sequence through asymmetric connections [34, 35], while it might require some balance in the usage of asymmetric connections for representing both "what" and "what-where" association. Further evaluation is necessary in terms of both representation and dynamics for the comprehensive understanding of the episodic memory.

6.3. Computational Models-Experiments-Combined Approach. The computational models have been applied to experimental data analysis of fMRI measurements. Tanaka et al. [64] applied the temporal difference (TD) learning algorithm to the BOLD signal analysis and detected a topographical map of time scales of reward predictions. Anderson et al. [65] applied their information processing model to the analysis of BOLD responses and evaluated functional roles of their region-of-interests. Section 5 indicated that our computational model of the hippocampus was applied to analyze human eye movement data and showed its prediction ability for human memory recall. The model is also applicable to the brain signal analysis and its performance is now under evaluation. These studies demonstrated the efficacy of computational model-based analyses for understanding system-level brain functions. Recently methods of fMRI signal decoding have been developed to read the perceptual state of an observer [66]. The computational models mentioned in this text should contribute to brain signal decoding to validate the existence of their dynamics within the brain.

Acknowledgment

This study was supported by the MEXT KAKENHI (20220003).

References

[1] D. Marr, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, W. H. Freeman, New York, NY, USA, 1982.
[2] L. R. Squire, "The legacy of patient H. M. for neuroscience," *Neuron*, vol. 61, no. 1, pp. 6–9, 2009.
[3] M. A. Wheeler, "Episodic memory and autonoetic awareness," in *The Oxford Handbook of Memory*, E. Tulving and F. I. M. Craik, Eds., Oxford University Press, Oxford, UK, 2005.
[4] F. Vargha-Khadem, D. G. Gadian, K. E. Watkins, A. Connelly, W. Van Paesschen, and M. Mishkin, "Differential effects of early hippocampal pathology on episodic and semantic memory," *Science*, vol. 277, no. 5324, pp. 376–380, 1997.
[5] E. Tulving, *Elements of Episodic Memory*, Clarendon Press, Oxford, UK, 1983.
[6] M. L. Smith and B. Milner, "The role of the right hippocampus in the recall of spatial location," *Neuropsychologia*, vol. 19, no. 6, pp. 781–793, 1981.
[7] C. B. Cave and L. R. Squire, "Equivalent impairment of spatial and nonspatial memory following damage to the human hippocampus," *Hippocampus*, vol. 1, no. 3, pp. 329–340, 1991.
[8] J. A. King, N. Burgess, T. Hartley, F. Vargha-Khadem, and J. O'Keefe, "Human hippocampus and viewpoint dependence in spatial memory," *Hippocampus*, vol. 12, no. 6, pp. 811–820, 2002.
[9] K. Stepankova, A. A. Fenton, E. Pastalkova, M. Kalina, and V. E. D. Bohbot, "Object-location memory impairment in patients with thermal lesions to the right or left hippocampus," *Neuropsychologia*, vol. 42, no. 8, pp. 1017–1028, 2004.
[10] E. T. Rolls, "Spatial view cells and the representation of place in the primate hippocampus," *Hippocampus*, vol. 9, no. 4, pp. 467–480, 1999.
[11] D. Gaffan, "Scene-specific memory for objects: a model of episodic memory impairment in monkeys with fornix transection," *Journal of Cognitive Neuroscience*, vol. 6, no. 4, pp. 305–320, 1994.
[12] M. J. Eacott and G. Norman, "Integrated memory for object, place, and context in rats: a possible model of episodic-like memory?" *Journal of Neuroscience*, vol. 24, no. 8, pp. 1948–1953, 2004.
[13] M. Mishkin, W. A. Suzuki, D. G. Gadian, and F. Vargha-Khadem, "Hierarchical organization of cognitive memory," *Philosophical Transactions of the Royal Society of London B*, vol. 352, no. 1360, pp. 1461–1467, 1997.
[14] D. Johnston and D. Amaral, "Hippocampus," in *The Synaptic Organization of the Brain*, G. Shepherd, Ed., Oxford University Press, Oxford, UK, 4th edition, 2004.
[15] D. Marr, "Simple memory: a theory for archicortex," *Philosophical Transactions of the Royal Society of London B*, vol. 262, no. 841, pp. 23–81, 1971.
[16] B. L. McNaughton and R. G. M. Morris, "Hippocampal synaptic enhancement and information storage within a distributed memory system," *Trends in Neurosciences*, vol. 10, no. 10, pp. 408–415, 1987.
[17] L. R. Squire, "Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans," *Psychological Review*, vol. 99, no. 2, pp. 195–231, 1992.

[18] J. O'Keefe and L. Nadel, *The Hippocampus as a Cognitive Map*, Clarendon Press, Oxford, UK, 1978.
[19] T. Ono and H. Nishijo, "Active spatial information processing in the septo-hippocampal system," *Hippocampus*, vol. 9, no. 4, pp. 458–466, 1999.
[20] A. D. Ekstrom, J. B. Caplan, E. Ho, K. Shattuck, I. Fried, and M. J. Kahana, "Human hippocampal theta activity during virtual navigation," *Hippocampus*, vol. 15, no. 7, pp. 881–889, 2005.
[21] B. H. Bland, "The physiology and pharmacology of hippocampal formation theta rhythms," *Progress in Neurobiology*, vol. 26, no. 1, pp. 1–54, 1986.
[22] J. O'Keefe and M. L. Recce, "Phase relationship between hippocampal place units and the EEG theta rhythm," *Hippocampus*, vol. 3, no. 3, pp. 317–330, 1993.
[23] W. E. Skaggs, B. L. McNaughton, M. A. Wilson, and C. A. Barnes, "Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences," *Hippocampus*, vol. 6, no. 2, pp. 149–172, 1996.
[24] W. B. Levy and O. Steward, "Temporal contiguity requirements for long-term associative potentiation/depression in the hippocampus," *Neuroscience*, vol. 8, no. 4, pp. 791–797, 1983.
[25] G.-Q. Bi and M.-M. Poo, "Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type," *Journal of Neuroscience*, vol. 18, no. 24, pp. 10464–10472, 1998.
[26] W. B. Levy, "A computational approach to hippocampal function," in *Computational Models of Learning in Simple Neural Systems*, R. D. Hawkins and G. H. Bower, Eds., pp. 243–305, Academic Press, San Diego, Calif, USA, 1989.
[27] A. Treves and E. T. Rolls, "Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network," *Hippocampus*, vol. 2, no. 2, pp. 189–199, 1992.
[28] M. E. Hasselmo, E. Schnell, and E. Barkai, "Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3," *Journal of Neuroscience*, vol. 15, no. 7, pp. 5249–5262, 1995.
[29] R. C. O'Reilly and J. W. Rudy, "Conjunctive representations in learning and memory: principles of cortical and hippocampal function," *Psychological Review*, vol. 108, no. 2, pp. 311–345, 2001.
[30] K. Nakazawa, M. C. Quirk, R. A. Chitwood, et al., "Requirement for hippocampal CA3 NMDA receptors in associative memory recall," *Science*, vol. 297, no. 5579, pp. 211–218, 2002.
[31] W. B. Levy, "A sequence predicting CA3 is a flexible associator that learns and uses context to solve hippocampal-like tasks," *Hippocampus*, vol. 6, no. 6, pp. 579–590, 1996.
[32] O. Jensen and J. E. Lisman, "Theta/gamma networks with slow NMDA channels learn sequences and encode episodic memory: role of NMDA channels in recall," *Learning & Memory*, vol. 3, no. 2-3, pp. 264–278, 1996.
[33] O. Jensen and J. E. Lisman, "Hippocampal CA3 region predicts memory sequences: accounting for the phase precession of place cells," *Learning & Memory*, vol. 3, no. 2-3, pp. 279–287, 1996.
[34] Y. Yamaguchi, "A theory of hippocampal memory based on theta phase precession," *Biological Cybernetics*, vol. 89, no. 1, pp. 1–9, 2003.
[35] N. Sato and Y. Yamaguchi, "Memory encoding by theta phase precession in the hippocampal network," *Neural Computation*, vol. 15, no. 10, pp. 2379–2397, 2003.
[36] A. Samsonovich and B. L. McNaughton, "Path integration and cognitive mapping in a continuous attractor neural network model," *Journal of Neuroscience*, vol. 17, no. 15, pp. 5900–5920, 1997.
[37] K. I. Blum and L. F. Abbott, "A model of spatial map formation in the hippocampus of the rat," *Neural Computation*, vol. 8, no. 1, pp. 85–93, 1996.
[38] N. Burgess, M. Recce, and J. O'Keefe, "A model of hippocampal function," *Neural Networks*, vol. 7, no. 6-7, pp. 1065–1081, 1994.
[39] A. D. Redish and D. S. Touretzky, "The role of the hippocampus in solving the Morris water maze," *Neural Computation*, vol. 10, no. 1, pp. 73–111, 1998.
[40] H. Wagatsuma and Y. Yamaguchi, "Cognitive map formation through sequence encoding by theta phase precession," *Neural Computation*, vol. 16, no. 12, pp. 2665–2697, 2004.
[41] E. T. Rolls, S. M. Stringer, and T. P. Trappenberg, "A unified model of spatial and episodic memory," *Proceedings of the Royal Society of London B*, vol. 269, no. 1496, pp. 1087–1093, 2002.
[42] P. Byrne, S. Becker, and N. Burgess, "Remembering the past and imagining the future: a neural model of spatial memory and imagery," *Psychological Review*, vol. 114, no. 2, pp. 340–375, 2007.
[43] N. Sato and Y. Yamaguchi, "On-line formation of a hierarchical cognitive map for object-place association by theta phase coding," *Hippocampus*, vol. 15, no. 7, pp. 963–978, 2005.
[44] K. Rayner and A. Pollatsek, "Eye movements and scene perception," *Canadian Journal of Psychology*, vol. 46, no. 3, pp. 342–376, 1992.
[45] D. Noton and L. Stark, "Eye movements and visual perception," *Scientific American*, vol. 224, no. 6, pp. 35–43, 1971.
[46] Z. Wu and Y. Yamaguchi, "Input-dependent learning rule for the memory of spatiotemporal sequences in hippocampal network with theta phase precession," *Biological Cybernetics*, vol. 90, no. 2, pp. 113–124, 2004.
[47] A. Stevens and P. Coupe, "Distortions in judged spatial relations," *Cognitive Psychology*, vol. 10, no. 4, pp. 422–437, 1978.
[48] T. P. McNamara, J. K. Hardy, and S. C. Hirtle, "Subjective hierarchies in spatial memory," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 15, no. 2, pp. 211–227, 1989.
[49] N. Sato and Y. Yamaguchi, "Spatial-area selective retrieval of multiple object-place associations in a hierarchical cognitive map formed by theta phase coding," *Cognitive Neurodynamics*, vol. 3, no. 2, pp. 131–140, 2009.
[50] N. Sato, "Theta phase coding in human hippocampus: a combined approach of computational model and human brain activity analyses," in *Dynamic Brain*, M. Marinaro, S. Scarpetta, and Y. Yamaguchi, Eds., vol. 5286 of *Lecture Notes in Computer Science*, pp. 13–27, Springer, Berlin, Germany, 2008.
[51] N. Sato and Y. Yamaguchi, "Theta synchronization networks emerge during human object-place memory encoding," *NeuroReport*, vol. 18, no. 5, pp. 419–424, 2007.
[52] N. Sato and Y. Yamaguchi, "EEG theta regulates eye saccade generation during human object-place memory encoding," in *Advances in Cognitive Neurodynamics*, R. Wang, F. Gu, and E. Shen, Eds., pp. 429–434, Springer, Berlin, Germany, 2008.
[53] N. Sato, T. Ozaki, Y. Someya, et al., "Subsequent memory-dependent EEG θ correlates to parahippocampal BOLD response," *NeuroReport*. In press.
[54] N. Sato and Y. Yamaguchi, "Computational model-based prediction of human episodic memory performance based on eye movements," *IEICE Transactions on Communications*, vol. E91-B, no. 7, pp. 2142–2143, 2008.
[55] T. Sommer, M. Rose, C. Weiller, and C. Büchel, "Contributions of occipital, parietal and parahippocampal cortex to encoding

of object-location associations," *Neuropsychologia*, vol. 43, no. 5, pp. 732–743, 2005.

[56] N. Sato and Y. Yamaguchi, "An evidence of a hierarchical representation of object-place memory based on theta phase coding: a computational model-human experiment combined analysis," in *Proceedings of the Neuroscience Meeting Planner (NMP '06)*, Society for Neuroscience, Atlanta, Ga, USA, 2006, program No. 366.25, CD-ROM.

[57] M. E. Hasselmo and H. Eichenbaum, "Hippocampal mechanisms for the context-dependent retrieval of episodes," *Neural Networks*, vol. 18, no. 9, pp. 1172–1190, 2005.

[58] S. M. Polyn and M. J. Kahana, "Memory search and the neural representation of context," *Trends in Cognitive Sciences*, vol. 12, no. 1, pp. 24–30, 2008.

[59] M. W. Jones and M. A. Wilson, "Theta rhythms coordinate hippocampal-prefrontal interactions in a spatial memory task," *PLoS Biology*, vol. 3, no. 12, article e402, 2005.

[60] T. Hartley, N. Burgess, C. Lever, F. Cacucci, and J. O'Keefe, "Modeling place fields in terms of the cortical inputs to the hippocampus," *Hippocampus*, vol. 10, no. 4, pp. 369–379, 2000.

[61] I. E.T. de Araujo, E. T. Rolls, and S. M. Stringer, "A view model which accounts for the spatial fields of hippocampal primate spatial view cells and rat place cells," *Hippocampus*, vol. 11, no. 6, pp. 699–706, 2001.

[62] N. Burgess, E. A. Maguire, and J. O'Keefe, "The human hippocampus and spatial and episodic memory," *Neuron*, vol. 35, no. 4, pp. 625–641, 2002.

[63] M. J. Kahana, "The cognitive correlates of human brain oscillations," *Journal of Neuroscience*, vol. 26, no. 6, pp. 1669–1672, 2006.

[64] C. S. Tanaka, K. Doya, G. Okada, K. Ueda, Y. Okamoto, and S. Yamawaki, "Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops," *Nature Neuroscience*, vol. 7, no. 8, pp. 887–893, 2004.

[65] J. R. Anderson, M. V. Albert, and J. M. Fincham, "Tracing problem solving in real time: fMRI analysis of the subject-paced Tower of Hanoi," *Journal of Cognitive Neuroscience*, vol. 17, no. 8, pp. 1261–1274, 2005.

[66] K. N. Kay and J. L. Gallant, "I can see what you see," *Nature Neuroscience*, vol. 12, no. 3, pp. 245–246, 2009.

Hindawi Publishing Corporation
Advances in Artificial Intelligence
Volume 2010, Article ID 521606, 12 pages
doi:10.1155/2010/521606

Research Article

Application of Game Theory to Neuronal Networks

Alfons Schuster[1,2] and Yoko Yamaguchi[2]

[1] *School of Computing and Mathematics, Faculty of Computing and Engineering, University of Ulster, Shore Road, Newtownabbey, Co. Antrim BT37 0QB, Northern Ireland*
[2] *Laboratory for Dynamics of Emergent Intelligence, RIKEN Brain Science Institute, Wako-shi, Saitama 351-0198, Japan*

Correspondence should be addressed to Alfons Schuster, a.schuster@ulster.ac.uk

Received 28 August 2009; Accepted 2 October 2009

Academic Editor: Naoyuki Sato

The paper is a theoretical investigation into the potential application of game theoretic concepts to neural networks (natural and artificial). The paper relies on basic models but the findings are more general in nature and therefore should apply to more complex environments. A major outcome of the paper is a learning algorithm based on game theory for a paired neuron system.

1. Introduction

Individual neurons are the building blocks for more complex neural circuits. In natural systems these more complex neural circuits interact with other components in a manifold of ways thereby generating the compellingly sensual world of behavior around us. Although tireless and tedious efforts in various disciplines culminated in fundamental insights in the field, there are still many unknowns about individual neurons and the processes in which individual neurons interact and organize themselves in neural circuits (e.g., [1]).

Recently, game theory has obtained some attention in the field of neuroscience. The field of neuroeconomics, for instance, combines the two fields in experiments with human and nonhuman players in order to better understand human decision-making (e.g., [2]). This paper has a different motivation and proposes a neural network model under a concept of game theory where individual neurons are assumed to optimally behave with a given payoff matrix. The paper theoretically analyzes a paired neuron system and critically specifies that the value game theory may have as an organizing principle for such a system (in the sense of a guiding principle or mechanism involved in neural communication, organization, and synchronization). The paper also specifies a learning algorithm based on game theory for a paired neuron system, which is a major contribution in this text.

In the remainder of this text, Section 2 summarizes the motivation for this paper and validates an intuitively appealing (though not unproblematic) relationship between game theory and biological/artificial neurons. Sections 3 and 4 investigate this relationship, the theory, and the major concepts and challenges involved in more detail, concentrating, among other things, on static and dynamic games of complete/perfect information. Section 5 applies game theoretic constructs to artificial neural networks and presents a learning algorithm based on game theory for network learning. The discussion in Section 6 revolves around related work and Section 7 ends the paper with a summary.

2. Game Theory, Biological Neurons, and Artificial Neural Networks

Our previous work in various areas (e.g., artificial intelligence, soft computing, reasoning under uncertainty, and neuroscience) identified that many cooperations between two agents (artificial or natural) can be interpreted or bear some of the characteristic features of a game. For example, the main concepts in a game are the players in a game, a set of rules by which the game is played, and an outcome in the form of a reward or a punishment (more generally referred to as a payoff) for the players in the game. In addition,

a so-called payoff matrix is a common scheme to represent the dynamic behavior of a game.

Figure 1 applies these key concepts to a coupled neuron system where the neurons are modeled to calculate their strategies according to their individual payoff matrix. (The scopes for game theory and neural networks are extremely wide. The paper therefore uses several abstractions and simplifications (e.g., the neuronal circuit models presented in this text are relatively basic, and in terms of game theory this paper concentrates on static games and dynamic games of complete/perfect information). At large, the paper does not suffer from this reductionism as the findings mentioned in the paper are relevant in a wider sense. London and Häusser [1], for instance, emphasize that the contribution of single neurons to computation in the brain has long been underestimated and that there is a need to investigate novel mechanisms that allow individual neurons to implement elementary computations.) Imagine that the two neurons in Figure 1(a) shall generate the following global behavior: if Neuron-1 fires, then Neuron-2 shall fire, and if Neuron-1 is at rest (not firing), then Neuron-2 shall be at rest (it is possible to assume an information exchange, unidirectional or bidirectional, via biochemical substances or electrical signals between Neuron-1 and Neuron-2). Figure 1(b) presents this behavior in a payoff matrix. The payoff matrix assigns a payoff (illustrated as a reward R or a punishment P) to each neuron for each combination of strategies (Fire, Rest). For instance, if Neuron-1 fires and Neuron-2 also fires, then each neuron obtains a rewarding payoff. (Traditionally, the payoff for Neuron-1 would be the left value in a matrix cell, and the payoff for Neuron-2 would be the right value in a cell. Note also that the payoffs in a cell need not be identical.) If the two neurons correspond with different strategies (e.g., Neuron-1 fires and Neuron-2 remains at rest or vice versa), then each neuron receives a punishment payoff P. Thus, if the goal for the two neurons in Figure 1(a) is to eventually demonstrate the global behavior Fire/Fire, Rest/Rest, then it is possible to assume the following: (i) if the two neurons demonstrate the desired behavior (Fire/Fire, Rest/Rest), then no action is required, and (ii) in case the two neurons do not demonstrate this desired way of interaction, then some corrective action has to be taken to achieve the desired global behavior. Again, this paper is not interested in the exact description of the biochemical processes (which are not known in their entirety anyway) that may achieve this mode of operation in biological neurons—the motivation here is to describe this global interaction via game theoretic concepts, perhaps involving additional models and abstractions for the two neurons in Figure 1(a). (The following book by Purves et al. [3] provides a comprehensive account of the state of art of neuroscience, and Chapter 1 of this book, which is dedicated to neural signaling, is particularly informative about many of the issues mentioned in this text.) On the other hand, it is crucial to understand that the payoff matrix in Figure 1 is a crude generalization. In reality, it is very difficult to find and specify *exactly* a payoff function for a game, which is a critical task in game theory (i.e., approximations are the norm rather than the exception).

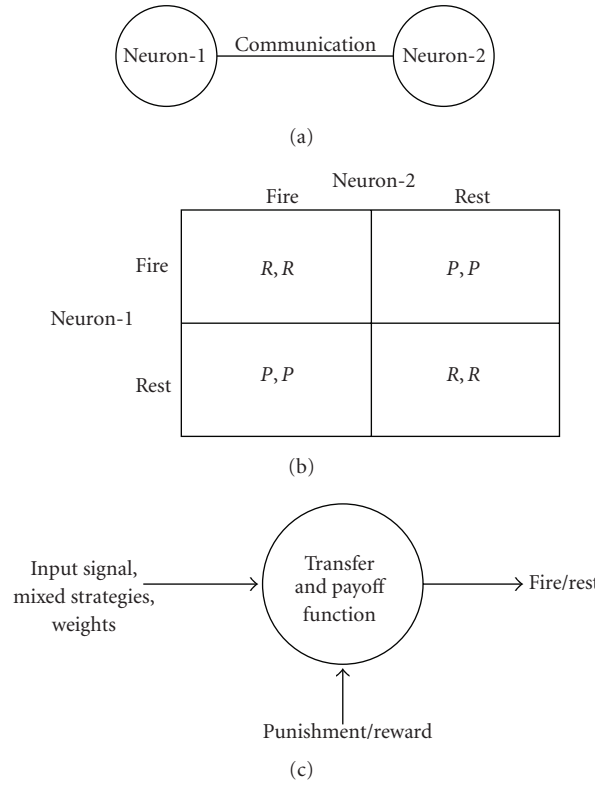


Figure 1: Relationships between (a) biological neurons, (b) game theory, and (c) artificial neurons.

Laying this issue aside, it is possible to provide a rather straightforward mathematical description for the modeling of the global behavior desired for the two neurons in Figure 1(a). To begin with Figure 1(a), it is necessary to understand that the communication between the two neurons in Figure 1(a) is a relatively simple, one-dimensional, linearly separable, and supervised learning classification task. Neuron-1 can either fire or be at rest, and Neuron-2 has to respond accordingly. It is possible to imagine a function $f(x)$ where a value $x \in R$ above a certain threshold value $t \in R$ represents the firing state for Neuron-1 and, a value $x \leq t$ represents the resting state for this neuron (1) as

$$f(x) = \begin{cases} \text{Fire} & \text{if } x > t, \\ \text{Rest} & \text{otherwise.} \end{cases} \tag{1}$$

Collectively, it is possible to think of Neuron-1 and Neuron-2 as a simple input-output unit that behaves similar to a switch. In terms of its global behavior, a perceptron can be interpreted exactly in the same way. (It is not necessary to elaborate on the perceptron learning algorithm in great detail as this information is widely available in the neural network literature (e.g., in [4, pages 43–54]).) This does not mean, however, that the payoff matrix in Figure 1(b) can be implemented by a traditional perception. Figure 1(c) illustrates a model that is similar to a perceptron but incorporates elements from game theory that may allow this model to demonstrate the behavior

illustrated by the payoff matrix in Figure 1(b). It is clear from Figure 1(c) that the decision-making process for this model involves some form of an input, an output, a transfer function, and a reward/punishment mechanism, all based on concepts from game theory. The forthcoming Section 5 provides a more detailed description for this model and the relationship illustrated in Figure 1 at large. The current focus is to describe the intuitive relationship between game theory, biological neurons, and artificial neural networks just mentioned in more detail and to elaborate on the various (fundamental) challenges involved in this relationship.

3. Game Theoretic Interpretations

In order to appreciate the forthcoming sections and to avoid unnecessary confusion, it is helpful to understand that game theory distinguishes between different types of games. At large, there are *static* games or *dynamic* games with *complete* information or *incomplete* information. If the payoffs and strategies available to other players are known and common knowledge to each player, as in Figure 1(b), then a game has complete information; otherwise, the game is classified as a game of incomplete information. Crucially, in a static game, players take their decisions simultaneously (individually and independently), they then move (not necessarily simultaneously but bound to the decisions they took) and then receive their payoffs. That is, the players in a static game are unaware about the strategies the other players in the game may choose but any player may hypothesize on the strategies other players may choose. (Marriage vows couples exchange to each other during a wedding ceremony may be a good example; the decisions are taken independently and the further proceedings of the ceremony unfold upon these decisions.) In a dynamic game, decisions are taken sequentially. In such a game, a player A may choose and act a particular strategy, and another player B who has observed player A may use this information for an appropriate response. (Chess is a typical example for such a game.)

It is tempting now to immediately view and deal with Figure 1 as a dynamic game with complete information where the payoffs in the matrix are common knowledge between the players, and Neuron-2 reacts (sequentially) to the signal arriving from Neuron-1 (perhaps with other processes going on bidirectionally). There are several reasons, however, to initially treat Figure 1 as a static game with complete information. For one thing, Figure 1 is a rather extreme reduction and it is relatively easy to envisage more complex scenarios. The two neurons in the figure could be exchanged with the brains of two humans or, for that matter, with the complete computer simulation of such two brains, which is the dream of the Blue Brain Project at EPFL (École Polytechnique Fédérale de Lausanne). Another reason involves understanding and learning; it is better to begin with (somewhat simpler) games of complete information and then to move on to more challenging games (in terms of the theory involved). In any case, the forthcoming text benefits from this bottom-up approach as it helps to specify, more clearly, some of the subtleties involved in this investigation. In terms of these subtleties, it is important to understand that several of the fundamental assumptions in game theory can be challenged intellectually with relative ease. Some of the reasons for this not only relate to the current example but also reach out deeper into the heart of game theory. These more sensitive (interrelated) concepts include *rationality*, *simultaneity*, *equilibrium*, and *mixed strategies*.

Rationality. Many of the formalisms in traditional game theory imply a degree of rationality by the players/agents involved in a game. As crucial as the notion of rationality is for the theory, the term rationality is not without problems. For one thing, the term rationality is not universally defined, and for another thing, human agents are often not the hyper-rational agents the theory requires them to be. Many applications of game theory therefore involve abstractions and simplifications to various degrees. For instance, this happens when game theory is applied to the modeling of interactions in genes, viruses, or cells, as is the case in *evolutionary game theory* [5]. (Evolutionary game theory is an extension to classical game theory motivated by some of the more problematic issues discussed in this section. Though very interesting and with some relevance to this work, evolutionary game theory has not been dealt with in this text mainly for the sake of brevity.) Another interesting contribution to this discussion may come from the observation that people usually associate biological brains with higher cognitive functions such as learning or rational decision-making. As true as this may be, many people also carry the common misconception that such a task can only be achieved by organisms with highly developed nervous systems, i.e., with brains, which is incorrect. For example, there are instances of predictive behavior within microbial genetic networks where bacteria anticipate changing environments [6]. Bacteria, however, have no brains or nervous systems. Instead, these microbes experience and learn through evolutionary changes in their complex networks of interacting genes and proteins (i.e., the problem-solving potential is encoded, in part, in the architectural configuration of the system) [7]. Although the specific mechanisms for this problem-solving ability are largely unknown today, many would agree that such tasks should involve some form of memory. The recent euphoria devoted to so-called memristors (memory resistors) may shed some light on this topic in the future. In electronics, a memristor is a fundamental basic circuit element [8]. Importantly, through this element, nature seems to provide a form of memory for free. Naturally, the value memristors have for neural networks has been identified in some of the aforementioned and other works already (e.g., [9]).

Simultaneity and Equilibrium. These terms are problematic too and can quickly lead into a deep philosophical discussion. A root problem in Figure 1(a) seems to relate to the larger problem of existence and timing. The typical development process for artificial neural networks relates to this problem quite well too. The learning process for such networks usually starts with a network configuration and a random weight assignment. But how does nature determine the configuration for a network or the degree of connectivity?

And how does the network know about the point in time when operation begins? Are these tasks performed by a monitoring supervisory unit or do the neurons involved act with a degree of autonomy (and rationality)? A more distant view magnifies this point even more. An outside observer looking at the complete neural activity of a human being, or a human being in its entirety for that matter, witnesses a multitude of processes running in parallel/simultaneously, and it is not clear at all to this observer how these processes may relate to each other or how they are coordinated in detail. A full discussion of this problem is beyond the scope of this preliminary investigation but it is worthwhile to describe how the concept of equilibrium emerges in this context. It is difficult to imagine an observer that is able to grasp a human being in its entirety. It is possible, however, to imagine an observer witnessing the object under observation in a particular higher-level, abstract global *state*. Assume a state of equilibrium (e.g., defined by an energy minimum or some other form of optimization or stabilization). In nature, a system may naturally strive or converge for such an equilibrium. Game theory provides the concept of equilibrium too—the agents in a game acquire this equilibrium through rational thought. Whether such an equilibrium is a *law* in nature (e.g., similar to the concept of entropy in physics) is only a thought that shall be laid aside here.

Mixed Strategies. Imagine that for some reason Neuron-1 and Neuron-2 in Figure 1 have cooperated well over time. In this case the likelihood that Neuron-2 fires when Neuron-1 fires could be rather high. On the other hand, if for some reason their cooperation was relatively poor in the past, then the likelihood of a correct response may be low. It is important to understand that in both cases positive as well as negative responses are still possible (e.g., a relatively good cooperation over time may not entirely prevent undesired responses). Game theory uses mixed strategies for the modeling of such likelihoods, and from a purely theoretical point of view, they are rather important in game theory. For example, in any game where a player has to outguess the behavior (strategy) of any other player involved in the game (e.g., in poker or in the childhood game rock-paper-scissors), there is no *Nash equilibrium* [10, pages 29–33]. In such a game a player may select a strategy according to some likelihood (e.g., motivated by a hint, a tipoff, or some other piece of information that may be difficult to quantify). Game theory expresses a mixed strategy for a player as a probability distribution over some or all strategies available to a player (p_i) in a game. It is clear that in many cases probability distributions may not be available and that the exact quantification of likelihoods is a point of weakness in game theory. In such cases the term *uncertainty* is often more appropriate. This term, however, opens the door for various theories dedicated to the field of management of uncertainty and ultimately adds a touch of vagueness to the rigorous formal underpinnings game theory provides. Hampton et al. [11], for instance, present several update rules for mixed strategies in a neuroscience-related study with human players and the paper mentions several other sources where this has happened in the past. Anyhow, Figure 2 illustrates a case with mixed strategies $(r, 1-r)$ for Player-1 and $(q, 1-q)$ for Player-2. The hypothetical mixed strategy $p_2 = (q, 1-q) = (0.8, 0.2)$ for Player-2 may then be interpreted as Player-1's uncertainty that Player-2 may play strategy Fire with probability/likelihood 0.8 and strategy Rest with probability/likelihood 0.2. (Note that the terms player and neuron can be used interchangeably in the figure. In addition, the payoff matrix in Figure 2 with its numeric values is less general than that of Figure 1(b). This is for demonstration purposes only and does not impair the general conclusions presented in the forthcoming sections.)

		Player-2 (Neuron-2): Fire (q)	Player-2 (Neuron-2): Rest ($1-q$)
Player-1 (Neuron-1): Fire	(r)	1,1	0,0
Player-1 (Neuron-1): Rest	($1-r$)	0,0	1,1

Figure 2: A static game with complete information and mixed strategies.

The remaining text in Section 3 analyzes the static game with complete information illustrated in Figure 2 in more detail and starts with Player-1's point of view of the game. (Gibbon's [10] book on game theory is a major resource in this work and those readers wishing to get further information about the game theoretic elements mentioned in this text are referred to that text.)

3.1. Player-1's (Neuron-1's) Point of View. For simplicity, Figure 3 illustrates Player-1's (Neuron-1's) view only. Player-2's payoff is irrelevant in this view; that is why it is omitted in Figure 3.

According to Figure 3, given that Player-1 believes that Player-2 will play the mixed strategy $(q, 1-q)$, then the expected payoff for Player-1 for playing the pure strategy Fire is

$$f^*(q) = q \cdot (1) + (1-q) \cdot (0) = q. \tag{2}$$

Similarly, the expected payoff for Player-1 for playing the pure strategy Rest is

$$g^*(q) = q \cdot (0) + (1-q) \cdot (1) = 1-q. \tag{3}$$

Figure 4 illustrates (2) and (3) in a single diagram. In order to understand the forthcoming arguments, it is important to always bear in mind that the main goal for each player is to obtain a *maximum payoff* in a game. Figure 4 illustrates that if $q > 1/2$, then $f^*(q) > g^*(q)$ in which case Player-1 should play strategy Fire (see also Figure 3). On the other hand, if $q < 1/2$, then $g^*(q) > f^*(q)$ in which case Player-1 should adopt strategy Rest. A special case exists for

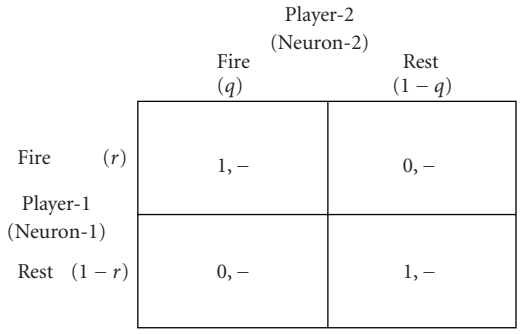


Figure 3: Viewpoint of Player-1 (Neuron-1).

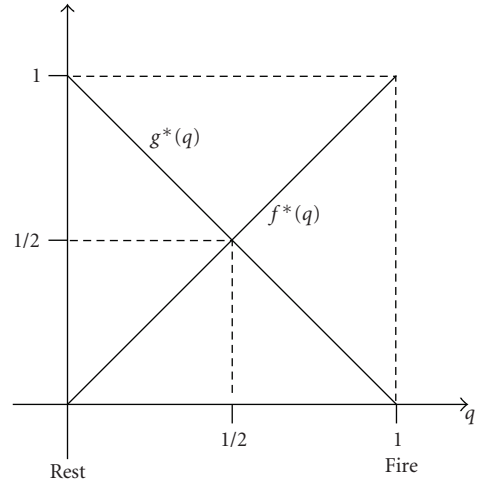


Figure 4: Decision-making support for Player-1 if Player-1 believes that Player-2 plays the mixed strategy $(q, 1-q)$.

$q = 1/2$, which is the point where the two straight lines $f^*(q)$ and $g^*(q)$ intersect. In this case Player-1 is indifferent about which strategy to play.

It is also possible to consider mixed strategy responses by Player-1. Player-1's expected payoff $r^*(q)$ from playing the mixed strategy $(r, 1-r)$ when Player-2 plays the mixed strategy $(q, 1-q)$ is the weighted sum of the expected payoff for each of the pure strategies (Fire, Rest) where the weights are the probabilities $(r, 1-r)$. According to Figure 3 this payoff amounts to

$$\begin{aligned} r^*(q) &= r \cdot q \cdot (1) + r \cdot (1-q) \cdot (0) + (1-r) \cdot q \cdot (0) \\ &\quad + (1-r) \cdot (1-q) \cdot (1) \\ &= r \cdot q + (1-r) \cdot (1-q) \\ &= 1 - q + r(2q-1). \end{aligned} \tag{4}$$

What exactly is at stake here? At stake is the goal to maximize the payoff for Player-1 expressed by (4). The mixed strategy $(r, 1-r)$ is the parameter that provides Player-1 with a handle to work towards this maximum. Consider three cases: $q = 0$, $q = 1$, and $q = 1/2$ (i.e., the problem is to determine which values for r maximize $r^*(q = 0)$, $r^*(q = 1)$, or $r^*(q = 1/2)$ for Player-1). For $q = 0$, (4) gives $r^*(q = 0) = 1 - r$. In this case $r = 0$ maximizes the term $1 - r$. For $q = 1$, (4) gives $r^*(q = 1) = r$, in which case $r = 1$ provides the maximum. Finally, for $q = 1/2$, (4) yields $r^*(q = 1/2) = 1/2$. This term is independent of r and indicates that any response by Player-1 is a best response to Player-2's assumed strategy. Figure 5 summarizes all best responses by Player-1 if Player-2 plays mixed strategy $(q, 1-q)$, and mixed strategy $(r, 1-r)$ is available to Player-1.

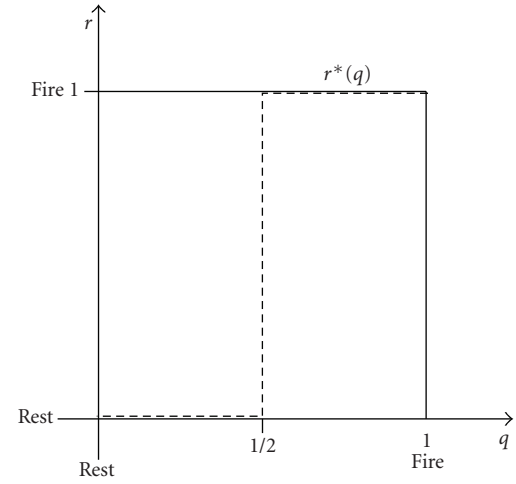


Figure 5: Player-1's best response (maximizing the expected payoff $r^*(q)$) from playing $(r, 1-r)$ when Player-2 plays $(q, 1-q)$. (The additional information on the vertical axis (r, and strategies Fire, Rest) aims to support the interpretation of this figure.)

All in all, Figure 5 indicates that if Player-2 plays mixed strategy $(q, 1-q)$, then Player-1's best response is to play (i) strategy Fire if $q > 1/2$, (ii) strategy Rest if $q < 1/2$, and (iii) any strategy if $q = 1/2$.

3.2. Player-2's (Neuron-2's) Point of View. This section describes Player-2's view from Figure 2. Overall, the steps are similar to those steps performed in the previous section. Given that Player-2 believes that Player-1 will play the mixed strategy $(r, 1-r)$, then the expected payoff for Player-2 when playing strategy Fire is

$$f^*(r) = r \cdot (1) + (1-r) \cdot (0) = r. \tag{5}$$

The expected payoff for Player-2 for playing the pure strategy Rest is

$$g^*(r) = r \cdot (0) + (1-r) \cdot (1) = 1 - r. \tag{6}$$

Figure 6 illustrates (5) and (6) in a single diagram. The interpretation of Figure 6 is similar to that of Figure 4. It is, however, important to carefully look at the labeling on the coordinate system axes. In Figure 6 the two straight lines $f^*(r)$ and $g^*(r)$ intersect at $r = 1/2$, indicating that for $r = 1/2$, Player-2 is indifferent about which strategy to play. Figure 6 then illustrates that Player-2 should play strategy Fire for $r > 1/2$ (because $f^*(r) > g^*(r)$) and strategy Rest for $r < 1/2$ (because $g^*(r) > f^*(r)$).

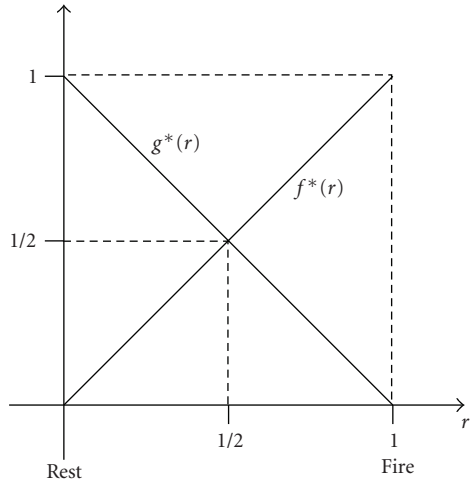


Figure 6: Decision-making support for Player-2 if Player-2 believes that Player-1 plays the mixed strategy $(r, 1-r)$.

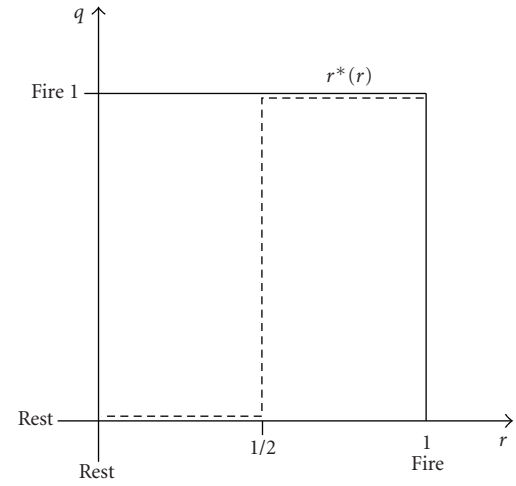


Figure 7: Player-2's best response (maximizing the expected payoff $r^*(r)$) from playing $(q, 1-q)$ when Player-1 plays $(r, 1-r)$. (The additional information on the vertical axis (q, and strategies Fire, Rest) aims to support the interpretation of this figure.)

Further, Player-2's expected payoff $r^*(r)$ from playing the mixed strategy $(q, 1-q)$ when Player-1 plays the mixed strategy $(r, 1-r)$ is (see Figure 2)

$$\begin{aligned} r^*(r) &= q \cdot r \cdot (1) + q \cdot (1-r) \cdot (0) + (1-q) \cdot r \cdot (0) \\ &\quad + (1-q) \cdot (1-r) \cdot (1) \\ &= q \cdot r + (1-q) \cdot (1-r) \\ &= 1 - r + q(2r-1). \end{aligned} \tag{7}$$

The interpretation of (7) is similar to that for (4). Here, Player-2 has the mixed strategy $(q, 1-q)$ at his disposal in order to maximize (7). Consider the following three cases: $r = 0$, $r = 1$, and $r = 1/2$. For $r = 0$, (7) gives $r^*(r = 0) = 1 - q$, and $q = 0$ generates the maximum for this term. Next, $r = 1$ gives $r^*(r = 1) = q$, and $q = 1$ provides the maximum. Finally, $r = 1/2$ establishes $r^*(r = 1/2) = 1/2$. This term is independent of q and so any response by Player-2 is a best response to Player-1's proposal. Figure 7 summarizes all best responses by Player-2 if Player-1 plays the mixed strategy $(r, 1-r)$.

Figure 7 illustrates that if Player-1 plays mixed strategy $(r, 1-r)$, then Player-2's best response is to play (i) strategy Fire if $r > 1/2$, (ii) strategy Rest if $r < 1/2$, and (iii) any strategy if $r = 1/2$.

3.3. Nash Equilibrium for Player-1 (Neuron-1) and Player-2 (Neuron-2). Figures 5 and 7 are quite similar and it is possible to combine both figures in a single diagram. Figure 8 emerges if Figure 7 is put on top of Figure 5 and additionally Figure 7 is flipped and rotated.

The interesting features in Figure 8 include those points where $r^*(q)$ and $r^*(r)$ intersect (i.e., points (0, 0), (1/2, 1/2), and (1, 1)). What makes these three points important is that for each of these three points the strategy chosen by any of the two players involved is a best response to the strategy chosen by the other player, and this is the definition of a

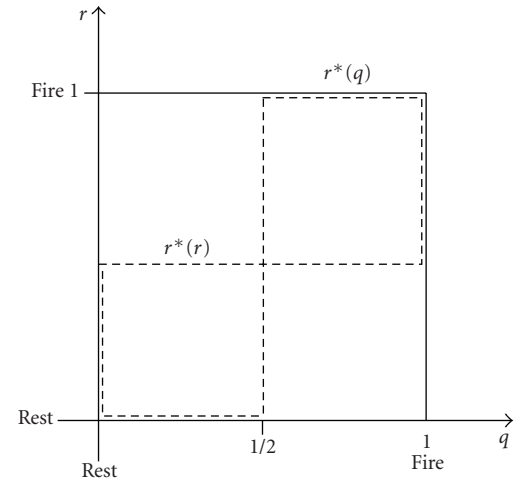


Figure 8: Combined view of best responses for Player-1 and Player-2. The three intersections between $r^*(q)$ and $r^*(r)$ are the Nash equilibriums in the game.

Nash equilibrium. Crudely, in a game played by n players, the strategies $(s_1, \ldots, s_n)$ are in a Nash equilibrium if for each player i in a game strategy s_i is a best response to the strategies $(s_1, \ldots, s_{i-1}, s_{i+1}, \ldots, s_n)$ specified for the $n-1$ other players in the game (e.g., see [10, pages 8–12 and 33–48]).

In the communicating neuron context of Figure 1(a), this means that if Neuron-1 fires then Neuron-2's best response is to fire too. If Neuron-1 is at rest, then Neuron-2's best response is to be at rest too. An interesting situation exists for point (1/2, 1/2). This situation may be interpreted as if Neuron-2 is unaware about the state (strategy) of Neuron-1, then Neuron-2 may play either strategy, and vice versa (i.e., the situation for each neuron/player is similar to the tossing of a coin).

At this moment, it may be useful to take a step back and to evaluate the results mentioned before a bit more carefully. The results are derived from a purely formal investigation of the (arbitrary) game illustrated in Figure 2. As discussed above, a neural behavior can be modeled under these game theoretic concepts. Whether these concepts can be theoretically generalized to neural systems with other arbitral payoff matrices is a question of debate. For example, the assumption that natural systems organize themselves according to the predictions of game theory (e.g., converge to or exploit Nash equilibriums) rather quickly leads back to the problems mentioned earlier in Section 3 (simultaneity, rationality, etc.). Consider a newly created or evolving biological neural network where new neurons emerge frequently (e.g., thousands of new neurons arise in the adult brain every day [12]). Some of these new neurons may be required to establish a way of communication with other neurons and it is difficult to imagine how this may work if there is no previous history between these neurons. Theoretically, for artificial neural networks, the situation is similar. Imagine a supervised learning scenario and an untrained network just provided with an initial random weight assignment. How does such a network know about a correct/incorrect classification outcome in the first place? The simple answer is that it knows from its supervisor (the network designer, developer, programmer, etc.). But who is the supervisor in nature? In nature, scientists often search for a guiding principle or law. This text does not suggest at all that game theory provides such a guiding principle, but it is necessary to create an awareness of the wider issues this work touches upon. Forthcoming sections relate back to some of the problems mentioned in this section but for the moment this text moves on to dynamic games.

4. Dynamic Games and Neural Circuit Dynamic

This section concentrates on dynamic games with complete and perfect information. Such games have three distinctive features: (i) the moves in the game occur sequentially, (ii) a sort of move history exists (i.e., all previous moves are observed before a next move is chosen), and (iii) the payoffs in the payoff matrix are known to all players in the game. Remember, a game has complete information if the content of the payoff matrix is common knowledge to all players in the game. A game has *perfect* information if every player has a record of the complete history of the game so far; otherwise, the game has *imperfect* information. *Backwards induction* is a general problem-solving strategy for such games and in many situations a *game tree* is a useful representation for a dynamic game. The game tree in Figure 9 represents an arbitrary dynamic two-move game played by two players (indicated as 1 and 2 in the figure).

The strategies for the players in Figure 9 are Left (L) and Right (R) for player one and Up (U) and Down (D) for player two. The numbers at the leave nodes at the bottom of the tree represent the payoffs for the players after traversing a particular route through the tree. The top number represents the payoff for player one, and the bottom number represents the payoff for player two. The game follows three rules; and

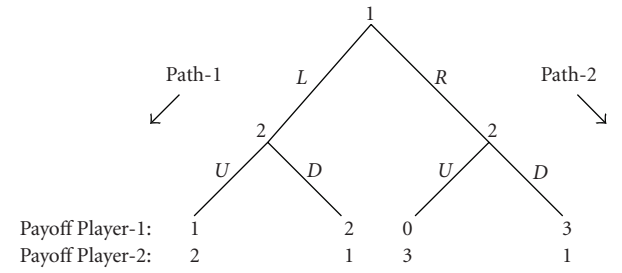


Figure 9: A game tree for a simple two-move game. There are two players (1 and 2) and the numbers at the bottom of the tree represent the payoff for each player traversing a particular path.

taken together, these rules are referred to as the *extensive-form* representation of the game.

(1) Player one decides on one of the available strategies (here, L or R).

(2) Player two observes this decision and decides on an appropriate strategy response (here, U or D).

(3) The players receive their payoffs.

Backwards induction works its way up from the bottom of the tree. Assume the position at the bottom of Path-1 where player one has decided to play strategy L and player two, who has observed this decision, is contemplating a response. The best response for player two is to play strategy U in which case player two receives the payoff 2 (instead of payoff 1), and player one recieves the payoff 1 (instead of payoff 2). Per definition, all information in the tree is available to all players (i.e., player one is aware that the response of player two is U if player one decides to play strategy L). Now assume the position of player two at the bottom of Path-2. In this case the best response for player two is again to play strategy U in which case player two receives the payoff 3 and player one the payoff 0. Player one can do some reasoning too. Between the two paths, and expecting best response decisions by player two, player one can expect a payoff of 1 for Path-1 (L) and a payoff of 0 for Path-2 (R). Each player aims for a maximum payoff and so player one decides to play strategy L. For player two, who is rational and aware of this thinking, the best response for this choice is to play strategy U. The pair (L, U) of best responses for player one and player two is referred to as the *backwards-induction outcome* of the game. This text mentioned earlier that there are different types and definitions for Nash equilibrium. In the type of dynamic game that just investigated the backwards-induction outcome of the game is the Nash equilibrium for the game (note that a game may have more than one Nash equilibrium).

Figure 10 applies these notions to the neuron communication example (see Figure 1(a) and the payoff matrix in Figure 2). The number 1 in the figure represents Neuron-1 and the number 2 stands for Neuron-2. The strategies for both neurons are Fire (F) and Rest (R).

For the game tree in Figure 10, backwards induction produces two backwards-induction outcome pairs, namely,

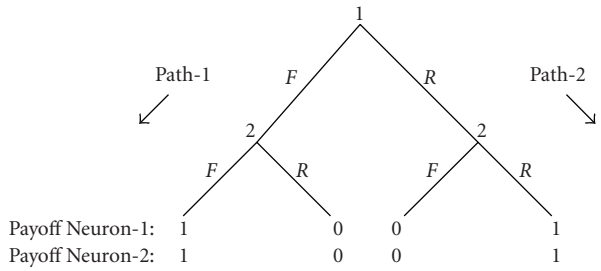


Figure 10: Game tree for the communicating neuron example (Figures 1 and 2). Two neurons (1 and 2) and their payoffs for traversing a particular path.

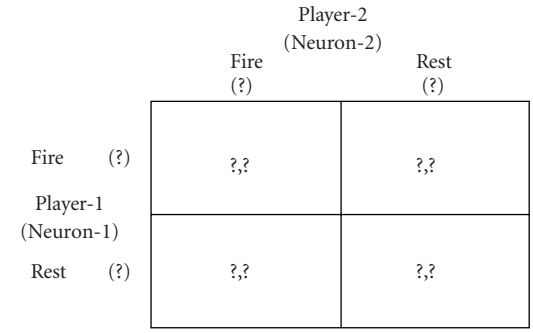


Figure 11: A learning algorithm for an artificial neural network based on game theory may exploit the payoffs in the matrix (i.e., the payoff function) and the mixed strategies.

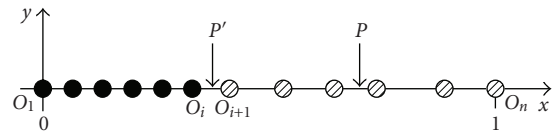


Figure 12: A one-dimensional, linearly separable, and supervised learning classification task.

the pair (F, F) and the pair (R, R). Both pairs are a Nash equilibrium for the game. This result is not so surprising and correlates with those results produced in the previous Section 3. If Neuron-1 fires, then the best response for Neuron-2 is to fire too, and if Neuron-1 is at rest, then the best response for Neuron-2 is to be at rest too. It is necessary now to mention that game theory provides several possible extensions to the type of games presented in this section. A simple extension is games with longer sequences (perhaps an infinite number) of moves and more than two players. A complete treatment of all these features is well beyond the scope of this paper, and the reference section in this paper may direct the interested reader to further relevant information on these topics. Overall, however, the section provides several important insights. First, the findings in this section associate game theory and neural network dynamic intuitively well, and second, the issue of repetitive, longer sequences involving updates naturally leads to the issue of learning.

5. Game Theory and Neural Network Learning

In order to acquire a capacity for decision-making, a network has to evolve from an unorganized state to an organized (synchronized) state with the latter state demonstrating the desired problem-solving potential. The mechanism that drives artificial neural networks from an unorganized state to an organized state is typically realized by a learning algorithm. This section describes a learning algorithm based on game theory for artificial neural networks. The question marks in Figure 11 indicate that game theory provides two possible access points for a learning algorithm: (i) the payoffs in the payoff matrix (i.e., the payoff function), and (ii) the values for the mixed strategies.

5.1. Algorithm. For the algorithm, imagine a one-dimensional, linearly separable, and supervised learning classification task. Figure 12 illustrates such a task. The classification scenario in Figure 12 takes place in an arbitrary real-valued x, y coordinate system. The classification scenario involves n objects and together these objects represent the training set for the learning algorithm (e.g., an object may represent a measurement of membrane potential in a neuroscience experiment and indicate whether a neuron is firing or in a resting state). The values measured for these objects have been normalized such that for every object i yields $x_i \in [0, 1]$. Let the black dots in Figure 12 represent objects of *Class 1,* and let the lined circles represent objects of *Class 2*; and let *Class 1* indicate the resting state of a neuron and *Class 2* indicate the firing state of a neuron. The two points P' and P in the figure are division points. In their current positions, P' correctly separates all objects into their corresponding classes, whereas P incorrectly classifies three *Class 2* objects. At the start of a learning scenario, P may have been positioned randomly and in successive steps the learning algorithm may have moved this starting point (through various other points) until it finished in location P', which is a solution to the problem. Figure 13 projects these ideas into a game theoretic context.

The figures in Figure 13 are similar to Figure 4 and represent Neuron-1's point of view. Remember, the mixed strategy $(q, 1 - q)$ represents Neuron-1's uncertainty about Neuron-2 and the task for Neuron-1 is to establish (in a learning process) a model about the expected behavior (mixed strategy $(q, 1 - q)$, payoff function) for Neuron-2. Further, every figure in Figure 13 includes two lines f_0 and either f_Q or f_R, which are all payoff functions. (Note that the forthcoming discussion now focuses on Figure 13(a) to 13(c).) Line f_0 is fixed and always remains unaltered during the learning process. In addition, f_0 represents the payoff function for *Class 1* and so, per definition, the resting state for Neuron-1. The second line f_Q is determined by the angle Q, where $0 \leq Q \leq 90$ degree. This line represents the payoff function for *Class 2* (i.e., the firing state for Neuron-1). The angle Q is derived by the function $m : q = [0, 1] \rightarrow Q = [0°, 90°]$ (e.g., the value $q = 0.5$ corresponds to an angle $Q = 45°$). The learning process for Figure 13 is similar to the scenario mentioned for Figure 12. Figure 13(a) represents an initial random assignment for Q.

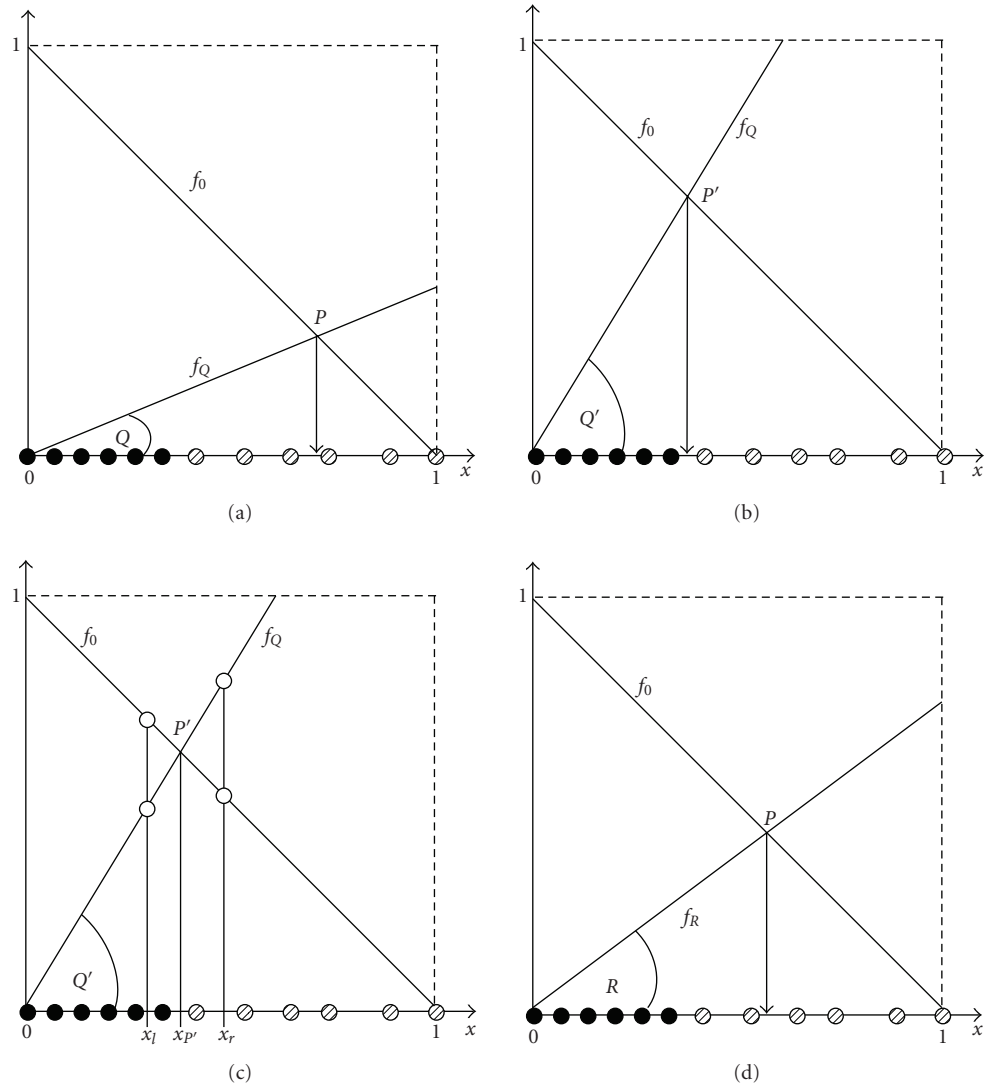


Figure 13: The game theory: based learning algorithm at work. Neuron-1's point of view: (a), (b), and (c). Neuron-2's point of view (d).

Point P in this figure is at the intersection of f_0 and f_Q. The learning algorithm will find out in the training phase that this point does not separate the two classes correctly and take appropriate action. In this case, the algorithm will increase the angle Q, which moves the intersection point further to the left. There may be several such steps until the algorithm arrives at point P' in Figure 13(b), which is a solution to the problem. (Note that from this position it is possible to determine (i) the mixed strategy $(q, 1-q)$ and (ii) the payoff functions (f_0 and f_Q) for fire/rest for Neuron-1, which is the goal for the learning algorithm.) Anyway, Algorithm 1 illustrates the learning algorithm in pseudocode, where the positive constant η in the algorithm represents the well-known learning rate. (Of course, the problems of seize of learning rate, outliers, stopping criteria, etc. apply to this algorithm too. These problems, however, can be neglected in this text as they are well commented elsewhere and not really relevant to the focus of this work.)

Algorithm Game Theory Neural Learning
  Start with a randomly chosen angle Q_0;
  Let $k = 1$;
  While there exist misclassified objects by Q_{k-1} **do**
    Let o_j be a misclassified object;
    Update the angle to $Q_k = Q_{k-1} \pm \eta$;
    Increment k;
  end-While;

Algorithm 1: Pseudocode for the game theory-based learning algorithm.

Figure 13(c) captures how the learning algorithm classifies unknown objects (i.e., objects that were not involved in the training phase). Any unknown object x_l to the left of point P' produces two intersections, one at f_Q and one at

f_0 (white circles in the figure). However, any of these points yields $f_0(x_l) > f_Q(x_l)$. That is, the payoff for $f_0(x_l)$ (rest) is always larger than the payoff for $f_Q(x_l)$ (fire). Therefore, Neuron-1 chooses to stay at rest for any such value. For similar reasons, for any object x_r to the right of P', Neuron-1 chooses to fire, because for any such value, the payoff $f_Q(x_r) > f_0(x_r)$. Equation (8) formalizes this outcome as follows:

$$g(x) = \begin{cases} \text{Fire} & \text{if } x > x_{P'}, \\ \text{Rest} & \text{otherwise}, \end{cases} \tag{8}$$

where $x_{P'}$ is the x coordinate of intersection point P' and in general the separation point determined by the learning algorithm. For the sake of completeness, Figure 13(d) illustrates a possible scenario from the viewpoint of Neuron-2. This scenario is similar to Figure 13(a) but in this figure it is Neuron-2 that has just received an initial random assignment for the angle R. The task for the learning algorithm now is to establish a model for Neuron-2 about the expected behavior (mixed strategy $(r, 1 - r)$, payoff function) for Neuron-1. It is not necessary to provide a detailed description for these processes for Neuron-2 because of the general symmetry of the system. (Note that this does not mean necessarily that Neuron-1 and Neuron-2 learn on the same data. Many of the examples in this text are high-level abstractions of natural systems where (i) information exchange between two neurons can be unidirectional, bidirectional, inhibitory, excitatory, and effect neuronal differentiation, (ii) unconventional neurotransmitters can provide signaling from postsynaptic cells back to presynaptic cells, or (iii) chemical signaling is not limited to synapses only (e.g., signaling may involve the secretion of chemical signals onto a group of nearby target cells). Thus, a measurement of data (e.g., a particular molecular concentration or a particular biochemical or electrical signal) at Neuron-1 related to q/Q or r/R may correspond to a related event involving the same or different components at Neuron-2. (For more detail, see [3, Unit 1, Neural Signalling]).) It is important, however, to understand what Section 5 achieved. The section formalized a learning algorithm in the game theoretic framework such that a paired neuron system can establish a synchronized way of communication. The learning algorithm determines the payoff functions for the payoff matrix as well as the mixed strategies for the neurons involved. This is an interesting and novel outcome according to our current understanding of the field. (It is clear that the presented algorithm shares many similarities with traditional neural network algorithms (e.g., the perceptron learning algorithm). The presented model, however, goes beyond traditional models where a neuron is modeled as an accumulator of multiple inputs (e.g., such as the McCulloch-Pitts neuron, which has been a basis of neural networks for some time). The paper mentioned already that, in reality, there are still many unknowns about individual neurons communicating with other neurons (e.g., an individual neuron is not just a neuronal membrane; for instance, it includes complex molecular circuits and well-organized structures, such as dendritic trees [1]). It is necessary, therefore, to develop the theoretical concept of individual neurons beyond the accumulative neuron model (e.g., by proposing a neural network model where individual neurons are assumed to optimally behave according to concepts from game theory). In addition, although the proposed model is for a paired neuron system only, the model has the potential to be expanded for the utilization to more complex networks. For example, the angles Q, R, etc. in Figure 13 lead to trigonometric functions (e.g., the division point $x_{P'}$ in Figure 13(c) can be determined from $\cos(Q')$ and P'). Learning algorithms for more complex multilayer networks (e.g., the backpropagation algorithm) rely heavily on derivatives (e.g., those of a transfer function). The derivatives for trigonometric functions are easy to obtain and this is certainly beneficial for potential expansions of the proposed approach to more complex network structures. A treatment of such potential expansions, however, is outside the scope of this paper.)

6. Related Work

This section initially repeats an important fact that has been mentioned several times in this text already, namely, that the scopes for neuroscience and game theory are quite rich and rather complex in their own right, and that this paper, consequently, can only present a condensed view of the many challenges involved in the wider context of this investigation. A second important statement in this section is the finding that although there is work combining game theory and neuroscience, according to our understanding, the two fields have not been combined in the way presented in this paper. For example, the relatively young field of neuroeconomics combines the two fields in experiments with human and nonhuman players (e.g., see Sanfey et al. [13] for a somewhat briefer introduction to neuroeconomics or Krüger et al. [2] who reviews this topic quite well). One assumption in the field is that one of the tasks of the human nervous system is to facilitate successful interaction in complex environments and that the process in essence is a decision-making process. Körding [14] describes that the value decision theory, which is formally well defined, may have for generating a better understanding of the processes going on in the nervous system during these interactions. The paper introduces the basic concepts of decision theory and emphasizes Bayesian decision theory because this theory, according to Körding, provides a compact and elegant formalism and contains other properties (e.g., its ability to handle uncertainty) that may suit studies in neuroscience well. Works by Sanfey [13, 15] or Hampton et al. [11] indicate other interesting research directions in neuroeconomics. A common feature in these papers is studies in which decision-making is based on game theoretic models that are mathematically well understood (e.g., Prisoners' Dilemma, Trust Game, or Ultimatum Game) and where the neural activity of participating players is recorded via established methods (e.g., functional magnetic resonance imaging). A major goal in these studies is to relate brain areas and fundamental brain mechanisms with decision-making tasks. Interesting results include those findings where outcomes disagree with theoretical predictions as is the case when emotions such

as anger, frustration, or greed, which are generally difficult to quantify and to describe mathematically, come into play because such findings may challenge basic game theoretic assumptions and definitions. For example, one study [16] measuring activation in the anterior insula (a brain region involved in emotional processing) of players participating in the so-called Ultimatum Game contradicts the concept of rationality mentioned in Section 3. The results from this study indicate that players may act irrationally (in a game theoretic sense) if other players act in an antisocial or unacceptable way (e.g., a player may not accept an indecent, unfair, or greedy offer). An outcome may also deviate from a predicted outcome if nonhuman players are involved (e.g., a program running on a desktop PC or a robot-like device), which may be interesting for people working in human computer interaction.

The possible application of game theory to fields such as human computer interaction indicates that game theory has long left its traditional environment—economy and human decision-making (the famous mathematician John Forbes Nash was awarded, jointly, the Nobel Prize in Economics in 1994 for his work in game theory). Today, the theory is widely applied in the natural sciences for the modeling of a rich variety of biological games involving agents of various types. Indeed, the principles of the theory are general enough to attract cutting-edge research in artificial intelligence or systems biology in applications where web-based intelligent agents or robots may have to wrestle with complex decision-making problems [17] or where evolutionary game theory investigates the interplay between evolutionary dynamics and biological games [5]. For this work it is important to understand that the term *rational* may not be utilized with ease in these domains and the term uncertainty often softens stricter demands (e.g., those coming from probability theory). Applications in artificial intelligence and evolutionary game theory therefore are permeated by techniques from soft computing (genetic algorithms, fuzzy logic, etc.), which makes it tempting to foresee the inclusion of some of these techniques into the model proposed in this work.

Although it is clear that several other interesting studies could be mentioned here, this review section wants to draw to an end by commenting, briefly, on the timing of games. This paper dealt with static and dynamic games in a separate way and this treatment may have given the impression that a system, over time, always sticks to one type of game, which is questionable. Consider the timing of games in a different context. Take a tournament where the teams A and B are two teams among several other teams. Imagine not only that team A and team B meet in the early qualifying stages of the tournament, that team A beats team B during these qualifying stages, but also that both teams survive qualifying and later meet again in the final, which is won by team B (e.g., in the 2008 Olympic Games, this was the case for the women's softball teams of Japan and the US. Team of Japan lost in the early stages against the team from the US but won the gold medal in the final against team of US.) Anyhow, if the team coaches elaborate on their strategies in the qualifying stages, then this analysis may have the form of a static game, whereas in the final, both teams have met before and so the coaches find themselves as game theoretic dynamic game analysts. How does this relate to neural networks? Take the case of an untrained neural network (natural or artificial) again. If the network is untrained (without history), then preliminary assumptions may come from a static game perspective. At a later point in time, some neurons in the network may have cooperated in the past in some way, and for their further interaction, dynamic game concepts may be applicable. A further treatment of this line of thought is beyond the scope of this paper but we feel that the accumulated information in this review section at large provides several pointers for further research.

7. Summary

The paper presented a novel concept for describing individual neurons under the game theoretic framework. The paper created a firm understanding about some of the fundamental problems in game theory and emphasized that these problems are not unique to the domain of neural systems, but that these problems reach out more deeply into game theory, science, and the world around us. The paper demonstrates that various strategic game theoretic concepts and calculations seem to be naturally suitable for the modeling of the behavior of a paired neuron system (and possibly for more complex networks). This finding was further solidified through the specification of a novel learning algorithm based on game theory for the purpose of neural learning.

Acknowledgment

The first author gratefully acknowledges the support of Japan Society for the Promotion of Science (JSPS Fellowship no. S09168).

References

[1] M. London and M. Häusser, "Dendritic computation," *Annual Review of Neuroscience*, vol. 28, pp. 503–532, 2005.

[2] F. Krüger, J. Grafman, and K. McCabe, "Neural correlates of economic game playing," *Philosophical Transactions of the Royal Society B*, vol. 363, no. 1511, pp. 3859–3874, 2008.

[3] D. Purves, G. J. Augustine, D. Fitzpatrick, et al., *Neuroscience*, Sinauer, Sunderland, Mass, USA, 3rd edition, 2004.

[4] K. Mehrotra, C. Mohan, and S. Ranka, *Elements of Artificial Neural Networks*, MIT Press, Cambridge, UK, 1997.

[5] M. A. Nowak and K. Sigmund, "Evolutionary dynamics of biological games," *Science*, vol. 303, no. 5659, pp. 793–799, 2004.

[6] T. Saigusa, A. Tero, T. Nakagaki, and Y. Kuramoto, "Amoebae anticipate periodic events," *Physical Review Letters*, vol. 100, no. 1, Article ID 018101, 4 pages, 2008.

[7] I. Tagkopoulos, Y.-C. Liu, and S. Tavazoie, "Predictive behavior within microbial genetic networks," *Science*, vol. 320, no. 5881, pp. 1313–1317, 2008.

[8] D. Strukov, G. Snider, D. Stewart, and R. Williams, "The missing memristor found," *Nature*, vol. 453, no. 7191, pp. 80–83, 2008.

[9] Y. Pershin, S. La Fontaine, and M. Di Ventra, "Memristive model of amoeba's learning," *Physical Review E* , vol. 80, no. 2, 6 pages, 2009.

[10] R. Gibbons, *A Primer in Game Theory*, Financial Times Prentice-Hall, Upper Saddle River, NJ, USA, 1992.

[11] A. N. Hampton, P. Bossaerts, and J. P. O'Doherty, "Neural correlates of mentalizing-related computations during strategic interactions in humans," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 18, pp. 6741–6746, 2008.

[12] T. J. Shors, "Saving new brain cells," *Scientific American*, vol. 300, no. 3, pp. 41–48, 2009.

[13] A. G. Sanfey, G. Loewenstein, J. D. Cohen, and S. M. McClure, "Neuroeconomics: cross-currents in research on decision-making," *Trends in Cognitive Sciences*, vol. 10, no. 3, pp. 108–116, 2006.

[14] K. Körding, "Decision theory: what "should" the nervous system do?" *Science*, vol. 318, no. 5850, pp. 606–610, 2007.

[15] A. G. Sanfey, "Social decision-making: insights from game theory and neuroscience," *Science*, vol. 318, no. 5850, pp. 598–602, 2007.

[16] M. Van't Wout, R. S. Kahn, A. G. Sanfey, and A. Aleman, "Affective state and decision-making in the ultimatum game," *Experimental Brain Research*, vol. 169, pp. 564–568, 2006.

[17] M. Tennenholtz, "Game theory and artificial intelligence," in *Foundations and Applications of Multi-Agent Systems*, M. d'Inverno, M. Luck, M. Fisher, and C. Preist, Eds., vol. 2403 of *Lecture Notes in Computer Science*, pp. 34–52, Springer, Berlin, Germany, 2002.

Hindawi Publishing Corporation
Advances in Artificial Intelligence
Volume 2010, Article ID 845723, 6 pages
doi:10.1155/2010/845723

Review Article

Constraints of Biological Neural Networks and Their Consideration in AI Applications

Richard Stafford

Department of Natural and Social Sciences, University of Gloucestershire, Cheltenham, GL50 4AZ, UK

Correspondence should be addressed to Richard Stafford, richardstafford@yahoo.co.uk

Received 30 August 2009; Accepted 7 November 2009

Academic Editor: Naoyuki Sato

Biological organisms do not evolve to perfection, but to out compete others in their ecological niche, and therefore survive and reproduce. This paper reviews the constraints imposed on imperfect organisms, particularly on their neural systems and ability to capture and process information accurately. By understanding biological constraints of the physical properties of neurons, simpler and more efficient artificial neural networks can be made (e.g., spiking networks will transmit less information than graded potential networks, spikes only occur in nature due to limitations of carrying electrical charges over large distances). Furthermore, understanding the behavioural and ecological constraints on animals allows an understanding of the limitations of bio-inspired solutions, but also an understanding of why bio-inspired solutions may fail and how to correct these failures.

1. Introduction

A common misconception of evolutionary biology is that natural selection acts to produce organisms perfectly adapted to their environment. This notion has perhaps best been challenged by Gould and Lewontin [1] with their analogy of the "the spandrels of San Marco," where they claim that evolution normally acts on existing structures and body plans, making the best use of them as possible. In general, new structures are not produced and large changes, such as changes in the body plan of arthropods, do not occur (the analogy being that the highly decorated spandrels have the primary function of supporting the structure of the cathedral, and only the secondary function of aesthetic beauty—nevertheless many visitors assume the highly decorated spandrels are there for the artwork they display).

While the structures of neurons are flexible, for example in terms of their length or number of dendrites or synaptic connections, many biochemical, physiological, behavioural, and ecological constraints still apply to their form and function, and to the ability of their form and function to adapt through evolution.

Understanding these constraints in a holistic manner is essential for the effective design of artificial neural networks. In some cases, somewhat paradoxically, the constraints can produce more efficient solutions than fully flexible artificial networks might (see section on energy constraints below). However, in many cases, rigidly following the constraints of biological networks is neither productive nor necessary for artificial intelligence applications (see sections on information transfer and ecological constraints below). This brief review gives an overview of the biological constraints of neural processing and where and when they might be important to consider in the design of artificial neural networks.

2. The Evolutionary Origins of Neurons

All animals with the exception of the phylum Porifera (the sponges) possess some sort of neurons, from the loose neural nets of the Cnidaria to the highly developed brains of the cephalopod molluscs and vertebrates [2]. Neurons are specialised forms of cells, and share many common features with other cells in the body. In particular, all cells have numerous proteins that allow transport of particles across the cell membrane, either through (facilitated) diffusion or by active transport against a concentration gradient.

In neurons, charged ions are the particles that move across the cell membrane, altering the voltage between the inside and outside of the cell [2].

The evolution of neurons from "normal" cells gives an insight into one of their major constraints; diffusion occurs across cell membranes (and the proteins embedded in them) and therefore essentially ions "leak" out of neurons. Neurons are not electrical wires, and a voltage or membrane potential at one end of a neuron will not travel far before degradation. In fact information—in terms of voltage—rapidly degrades in these "graded potential neurons" over distances of less than 1 mm [3]. Essentially, unmodified cells are not a good way of conducting electricity (through charged ions), and neurons have evolved mechanisms to cope with this, such as spiking. These mechanisms also have their own constraints and are further discussed below.

While neurons transmit electronic information, synapses are where the processing of information occurs [2]. Initially, the conversion of electrical information into chemical information in the synapse, and the reconversion back to electronic information in the postsynaptic neuron appears an anarchic process, yet the role of neurotransmitters can be highly varied in the postsynaptic role (producing various excitatory and inhibitory postsynaptic responses) (reviewed by [4]). Synaptic transfer, however, is a slow process in relation to the speed information passes through the nervous system [2]. Processes with many synapses, therefore, will be unable to process information very rapidly, as might be needed for escape behaviours, and classic examples of escape behaviours generally involve at most, a moderate number of synapses (see reviews in [5, 6]). While the functional evolutionary origins of synapses are unclear, recent studies have demonstrated that the genes required to produce proteins necessary for synaptic transmission are found in the genomes of sponges, which lack nervous systems [7]. Therefore, it is probable that synaptic transmission has its origins in exploiting proteins produced for another purpose. While synaptic processing is responsible for the successful functioning of animal nervous systems, it is developed from evolutionary modification of a "best available" solution, and in some cases may be constrained by the slow transmission rates of the process.

Artificial neural networks, however, do not need to model the complexity of synaptic transmission, unless their goal is to model and understand the biological processes that occur. Neither the "leak" of charge, nor the time of synaptic processing that occurs in real neurons needs to be a constraint of these networks, although computational time of processing an artificial synapse may still effectively limit the size of such a network.

3. Information Transfer by Neurons

As stated above, information rapidly degrades over short distances in simple graded potential neurons. The action potential, or spike, is therefore the common means of transfer of electronic information in neurons [2]. Essentially the spike rate or number of spikes in a given time period arriving at a synapse relates to the information available, and this information can, in some cases, be directly correlated to observable behaviours [6, 8–10]. Some studies have indicated that it is not just the spike rate and duration of spikes that contain information, but also the patterns of spikes that can contain information [11]. However, most studies suggest that spike patterns are not detected and contain no transferable information across synapses [9, 10, 12].

Conversion of membrane potential to spikes is analogous to differences between analogue and digital, and in the same manner that greater information can be transferred using analogue, graded potential neurons have the ability to transmit up to five times more information than spiking neurons [13, 14]. Essentially, this loss of information though spiking is due to the constraints of "leaky" neurons and is a biochemical/physiological constraint of information processing. Artificial neural networks (whether realised in hardware or software) are not subject to the same constraints, nevertheless numerous studies have studied or developed artificial neural networks consisting of spiking neurons, when it is highly likely that it would be simpler, and able to transmit more information, to create artificial networks of graded potential neurons. While studies investigating information limits of spiking networks, or the greater amount of information that could be passed across synapses if patterns of spikes could be recognised, should be encouraged, neural networks designed for a particular application should take heed of simpler and more effective approach of analogue or graded potential processing.

4. Energetic Constraints

In order to transfer electrical information in neurons, ions need to be moved across cell membranes to create a voltage or potential difference across the membrane, and to restore the neuron to its resting potential [2]. Much of this movement of ions is related to the work of ion pumps such as the sodium-potassium pump—which uses energy to maintain the ion balance. Therefore, complex biological neural networks used to achieve complex tasks use more energy than simple neural networks. The energetic constraints are not negligible. Larger brains and more neurons incur a higher metabolic cost than smaller brains [15] and have been shown to result in shorter life spans and lower fecundity in late life (e.g., in *Drosophila*, [16]).

In nature, evolutionary trade-offs have developed. Essentially these are where animals clearly show suboptimal behaviours, for example in terms of sex allocation, where it seems that constraints on perception may mean that optimal male to female ratios are not always produced [17]. While there is currently no firm evidence to indicate that these imperfect behaviours are related to neuronal constraints (but see evidence in [17, 18]), it is known that neural networks are bias in the type of environmental information they obtain and do not provide a full knowledge of the environment, as predicted by classic optimality models [19, 20]. Work is currently ongoing to determine the precise neural constraints in operation during the determination

of sex ratios, but it is likely to be related to the higher energy requirements required to develop a larger neural network capable of processing this information. Evolutionary trade-offs, therefore seem to occur between fitness gains from optimal behaviours and developmental costs of establishing the necessary neural pathways to sense and act on various environmental stimuli. This indicates that the energetic costs of neural processing are important in nature [21–24].

Energetic constraints are an example of where artificial neural networks could learn from biology. While the resultant behaviours of neurally constrained animals are not optimal, they are "good enough" for individuals to survive and reproduce. Most software based neural networks will not be run on computers capable of massive parallel processing, therefore the processing time for any given task will increase with the size of the network. Furthermore, for hardware embodied artificial networks, a direct analogy with power consumption can be drawn, which could affect the operating range of, for example, autonomously navigating robots. In these situations, designing a perfect solution, over something that is simply "good enough" may be ultimately disadvantageous. However, the seriousness of task may play the deciding role in the size of the network implemented. The example, for instance, of designing an analogue VLSI neural network to be deployed in cars to visually detect possible colliding objects [25], would greatly benefit from overengineering to create a perfect behaviour, even if such a behaviour is not present in the real biological system (see below). The power consumption of such a device would be minimal in comparisons with the overall power consumption of the car, and suboptimal behaviours (i.e., not detecting a collision, or falsely detecting a collision when no colliding objects are present) would have serious consequences.

5. Ecological Constraints

The ideas above have concentrated on the bottom up effect of neural processing; that is, the actual constrains on information transfer by neurons. One aspect that is often forgotten in considering information processing is the top down constraints. These are the processes that indicate the effectiveness of the neural processing—in keeping an organism in an optimal habitat to survive and reproduce.

In the above discussion on sex ratios, I discussed how the energetic costs of neural processing might outweigh the benefits obtained from capturing and acting on perfect data. In some cases, the costs of deferring from an optimal behaviour may not be large and a trade-off may be reached. In other cases, further evolutionary constraints might prevent optimal neural systems being required. For example, there is no need for an organism such as a snail to develop sensory systems to tell when it is about to be crushed by a human foot. Such a process will cause death to the snail, so the costs of being crushed are very high. However, even if an information capture and processing system did exist, it would not aid the snail. The mode of locomotion is slow, and however well it captured the information, it would not be able to avoid being crushed. This example seems obvious, but in terms of bio-inspired artificial neural networks, considering the behavioural ecology of the organism whose neural network you are trying to copy is something frequently forgotten [26] (see below). Especially when considering invertebrate neural networks and behaviours, if the animal has no need to behave in a certain way or capture a certain type of information, then it is unlikely to possess neural circuits to do so.

A key example of not considering ecology in sufficient detail can be given by work on the locust Lobula Giant Movement Detector (LGMD) neural network. Locusts possess a pair of neurons known as the Descending Contralateral Movement Detectors (DCMDs), which spike in a one-to-one ratio with the presynaptic LGMD [27]. Initial work was conducted on the neuron, largely because it was easy to record from, and responded well to visual movement stimuli [27–29]. It was found that it responded most vigorously to looming stimuli (objects approaching on a direct collision course) and was therefore considered a suitable neuron for detecting and avoiding collisions [30]. However, little consideration was given to what locusts might need to avoid colliding with. Insects, for example, as anyone who has seen a fly trapped in a room with many windows can confirm, are perfectly capable of colliding at high speed with objects and suffering little in the way of injury.

Finally, behavioural observations showed that flying locusts would briefly stop flying and perform a "gliding" manoeuvre to looming objects that best resembled the speed and size movements of predatory birds [8, 31]. These glides resulted in a rapid drop in height and would prevent the locust being eaten by the bird. The glides also occurred at peak DCMD spike frequencies [8, 9]. Objects that were larger or slower moving (in terms of a size: speed ratio) than predatory birds did not result in these gliding behaviours occurring. Thus, the collision sensor was in fact a predator avoidance mechanism, with a clear evolutionary and ecological advantage, rather than the collision detection mechanism previously proposed.

While the above example is an understandable progression of science, many bio-inspired neural networks were based on the LGMD and its ability to operate as a collision detector [32–34]. The LGMD, in fact, exploits a unique property of the predator, in that small, fast moving bird predators that catch locusts on the wing produce a mathematically unique signature of image expansion over the eye of the locust [26]. Larger or slower moving objects are much harder to detect using image expansion over the eye [26]. Thus, unmodified models of the LGMD neural network never produced reliable collision detectors, normally, since the sensitivity of the system needed to be increased, by altering the synaptic weights, to detect collisions; there were major problems with false detections [34]. Over engineering the network, to include many other bio-inspired neural networks operating together, did however, result in more reliable collision detection systems, in general, able to predict car collisions and not respond falsely to noncollision events [25, 35, 36].

Table 1: A summary of constraints of biological neural networks and the possible associated constraints for designing artificial neural networks. For further examples and references of these constraints refer to the main text.

Process or structure	Constraints in biological networks	Possible constraints in artificial networks
Chemical synapses	Slow and prone to fatigue	Synaptic modification possible without these constraints or accurate representation of the biological process (although fatigue can be useful in learning)
Graded potential neurons	Electric charge (ions) leaks, so information can only be transported a short distance	Electronics or computer simulations do not present these constraints
Spiking neurons	Less information can be carried than with graded potential neurons	No need to simulate spiking neurons, as no artificial constraints on graded potential neurons exist
Size of neural networks	Energetically costly, especially in small/cognitively simple animals	Artificial networks will have higher energy costs. This may be a constraint depending on the application (i.e., importance of battery life)
Ecological function	Many well studied neural networks are related to survival behaviours—and are tightly tuned to these behaviours	Artificial networks designed to implement different behaviours may not work, as they are not performing exactly the same role
Multi-stimuli responses	No single neural network is optimal for a behaviour, and several environmental stimuli and neural pathways combine to produce a robust behaviour	Optimising a single neural network for a behaviour may be difficult. Unexpected (and possibly poor) results may occur in overly simple environments

Another warning from behavioural ecology to the application of artificial, but bio-inspired, neural networks, is that behaviour is often more complex than first thought. To some extent, this is a situation that has been brought about by the strict "cause and effect" of manipulative experiments used in behavioural ecology, that can be used to show that behaviour X arises because of stimulus Y. In practise, although behaviour X cannot occur in the absence of stimulus Y, it is also dependent on stimuli U and W. Therefore, designing a neural network to pick up stimulus Y, will not reproduce behaviour X in an efficient manner. An example of this may be homing behaviour, where an individual animal forages away from its home, but returns to the home after each foraging location [37, 38]. Many animals (as diverse as desert ants, limpets, and fiddler crabs) use a "path integration" technique to find their way home, essentially calculating the net distance and angle travelled and moving home in a direct line. Research into experimentally moving fiddler crabs when foraging (moving the substratum) shows that they use path integration to find their way back to their home and cannot find their way home without this behaviour [39]. However, in many species, path integration is not foolproof, and is backed up by techniques such as trail following (following outgoing trails from the home position) and using features of the landscape to "reset" the path integration mechanism [38, 40, 41]. As such, designing a path integration neural network is useful, but may not make an application that behaves exactly as expected.

Testing and developing neural networks for creating robotic "behaviour" are also often conducted in simplified environments—either in laboratories with few "distracting" visual features, or in simulation environments such as Webots. In the above example of homing behaviour, it can be seen how simplified environments can be problematic—if there are no visual landscape features to detect, then this behaviour cannot occur, and path integration mechanisms cannot occur. In fact, recent work into aggregation in intertidal snails shows that simulations of the snails' behaviour need to contain complex information about the environment (particularly the persistence of mucus trails) to accurately mimic what occurs in real situations. Both simulations and real behavioural data conducted in simplified environments did not produce such effective aggregation behaviour [42].

6. Constraints on Human Cognition

The evolutionary origins of neural networks and the constraints played by neuronal processing are not only present in the invertebrates considered in the above examples. Constraints of neural processing also affect human cognition. An example of human neural processing constraints can be given by psychological tests on human perception. Many indicate that rapid decision-making is prone to errors. For example, tests showing images of soldiers either carrying a machine gun or an umbrella, and asking participants to rapidly assess the risk of the situation, often produced inaccurate responses [43]. Essentially this is an evolutionary trade-off between speed and accuracy of information processing, not different to the neural constraints faced in many nonhuman animals [44].

Equally the concept of evolutionary adaptations of neural networks is present in humans. Stress responses of the brain are thought to have arisen through the "flight or flight" mechanism to avoid predators and capture prey [45], although alternative theories do exist (e.g., [46]). The rapid

change in the human ecological niche over the past few centuries has resulted in negative implications for these stress responses. This has strong analogies with problems of artificial neural networks operating in slightly different ecological niches to which they originally evolved (see above example of locust collision detection networks). If the "ecology" of the artificial network is not identical to that of the mimicked network, it may not perform as predicted.

Although humans are far more cognitively complex than most invertebrates discussed in this paper, their neural systems still show the same constraints as discussed elsewhere, and building artificial intelligence applications to mimic human intelligence should still consider constraints such as, information transfer in spiking neurons, the evolutionary origins of the aspect of intelligence, or behaviour being mimicked.

7. Conclusions

This paper has reviewed some key biological constraints of neurons (Table 1), and indicated that many of these do not need to be constraints for developing artificial neural networks. Furthermore, it has reviewed ecological and behavioural constraints of some animals, which are unique to their particular ecological niches. These ecological and behavioural constraints will shape the neural networks used by animals to collect and process information, and are therefore vital to consider if developing artificial networks to try to mimic some or all of an animal's behaviour.

Acknowledgments

The author would like to thank Dr. Roger Santer, Dr. Di Catherwood, and two anonymous referees, whose comments greatly improved this manuscript.

References

[1] S. J. Gould and R. C. Lewontin, "The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme," *Proceedings of the Royal Society of London B*, vol. 205, no. 1161, pp. 581–598, 1979.

[2] M. F. Bear, B. W. Connors, and M. A. Paradiso, *Neuroscience: Exploring the Brain*, Lippincott, Williams and Wilkins, Baltimore, Md, USA, 3rd edition, 2007.

[3] J. H. van Hateren and S. B. Laughlin, "Membrane parameters, signal transmission, and the design of a graded potential neuron," *Journal of Comparative Physiology A*, vol. 166, pp. 437–448, 1990.

[4] G. Burnstock, "Neurotransmitters and trophic factors in the autonomic nervous system," *Journal of Physiology*, vol. 313, pp. 1–35, 1981.

[5] D. H. Edwards, W. J. Heitler, and F. B. Krasne, "Fifty years of a command neuron: the neurobiology of escape behavior in the crayfish," *Trends in Neurosciences*, vol. 22, no. 4, pp. 153–161, 1999.

[6] P. J. Simmons and D. Young, *Nerve Cells and Animal Behaviour*, Cambridge University Press, Cambridge, UK, 2nd edition, 1999.

[7] O. Sakarya, K. A. Armstrong, M. Adamska, et al., "A post-synaptic scaffold at the origin of the animal kingdom," *PLoS One*, vol. 2, article e506, 2007.

[8] R. D. Santer, P. J. Simmons, and F. C. Rind, "Gliding behaviour elicited by lateral looming stimuli in flying locusts," *Journal of Comparative Physiology A*, vol. 191, pp. 61–73, 2005.

[9] R. D. Santer, F. C. Rind, R. Stafford, and P. J. Simmons, "Role of an identified looming-sensitive neuron in triggering a flying locust's escape," *Journal of Neurophysiology*, vol. 95, pp. 3391–3400, 2006.

[10] R. Stafford and F. C. Rind, "Data mining neural spike trains for the identification of behavioural triggers using evolutionary algorithms," *Neurocomputing*, vol. 70, no. 4–6, pp. 1079–1084, 2007.

[11] E. Arabzadeh, S. Panzeri, and M. E. Diamond, "Deciphering the spike train of a sensory neuron: counts and temporal patterns in the rat whisker pathway," *Journal of Neuroscience*, vol. 26, no. 36, pp. 9216–9226, 2006.

[12] F. Rieke, D. Warland, R. R. de Ruyter van Steveninck, and W. Bialek, *Spikes: Exploring the Neural Code*, MIT Press, Cambridge, UK, 1999.

[13] R. R. de Ruyter van Steveninck and S. B. Laughlin, "The rate of information transfer at graded-potential synapses," *Nature*, vol. 379, pp. 642–645, 1996.

[14] P. J. Simmons and R. R. de Ruyter van Steveninck, "Reliability of signal transfer at a tonically transmitting, graded potential synapse of the Locust Ocellar Pathway," *Journal of Neuroscience*, vol. 25, pp. 7529–7537, 2005.

[15] K. Islar and C. P. van Schaik, "Metabolic costs of brain size evolution," *Biology Letters*, vol. 2, pp. 557–560, 2006.

[16] J. M. Burger, M. Kolss, J. Pont, and T. J. Kawecki, "Learning ability and longevity: a symmetrical evolutionary trade-off in Drosophila," *Evolution*, vol. 62, no. 6, pp. 1294–1304, 2008.

[17] S. A. West and B. C. Sheldon, "Constraints in the evolution of sex ratio adjustment," *Science*, vol. 295, no. 5560, pp. 1685–1688, 2002.

[18] D. M. Shuker, S. E. Reece, A. Lee, A. Graham, A. B. Duncan, and S. A. West, "Information use in space and time: sex allocation behaviour in the parasitoid wasp *Nasonia vitripennis*," *Animal Behaviour*, vol. 73, no. 6, pp. 971–977, 2007.

[19] M. Enquist and A. Arak, "Selection of exaggerated male traits by female aesthetic senses," *Nature*, vol. 361, no. 6411, pp. 446–448, 1993.

[20] C. R. Tosh, A. L. Jackson, and G. D. Ruxton, "The confusion effect in predatory neural networks," *The American Naturalist*, vol. 167, no. 2, pp. E52–E65, 2006.

[21] E. A. Herre, "Optimality, plasticity and selective regime in fig wasp sex ratios," *Nature*, vol. 329, no. 6140, pp. 627–629, 1987.

[22] G. A. Parker and J. Maynard Smith, "Optimality theory in evolutionary biology," *Nature*, vol. 348, pp. 27–33, 1990.

[23] K. Safi, M. A. Seid, and D. K. N. Dechmann, "Bigger is not always better: when brains get smaller," *Biology Letters*, vol. 1, no. 3, pp. 283–286, 2005.

[24] R. Stafford, "Exaptation and emergence as mechanisms to cross fitness valleys during evolution: an example using simulated homing behaviour," Nature Precedings, 2008, http://precedings.nature.com/documents/2172/version/1.

[25] J. Cuadri, G. Linan, R. Stafford, M. S. Keil, and E. Roca, "A bioinspired collision detection algorithm for VLSI implementation," in *Bioengineered and Bioinspired Systems II*, vol. 5839 of *Proceedings of SPIE*, pp. 238–248, Sevilla, Spain, May 2005.

[26] R. Stafford, R. D. Santer, and F. C. Rind, "The role of behavioural ecology in the design of bio-inspired technology," *Animal Behaviour*, vol. 74, no. 6, pp. 1813–1819, 2007.

[27] F. C. Rind and P. J. Simmons, "Seeing what is coming: building collision-sensitive neurones," *Trends in Neurosciences*, vol. 22, no. 5, pp. 215–220, 1999.

[28] C. H. F. Rowell, "The orthopteran descending movement detector (DMD) neurones: a characterisation and review," *Zeitschrift für Vergleichende Physiologie*, vol. 73, pp. 167–194, 1971.

[29] R. B. Pinter, "Visual discrimination between small objects and large textured backgrounds," *Nature*, vol. 270, no. 5636, pp. 429–431, 1977.

[30] P. J. Simmons and F. C. Rind, "Orthopteran DCMD neuron: a reevaluation of responses to moving objects. II. Critical cues for detecting approaching objects," *Journal of Neurophysiology*, vol. 68, no. 5, pp. 1667–1682, 1992.

[31] F. C. Rind and R. D. Santer, "Collision avoidance and a looming sensitive neuron: size matters but biggest is not necessarily best," *Proceedings of the Royal Society B*, vol. 271, pp. S27–S29, 2004.

[32] F. C. Rind and D. I. Bramwell, "Neural network based on the input organization of an identified neuron signaling impending collision," *Journal of Neurophysiology*, vol. 75, no. 3, pp. 967–984, 1996.

[33] M. Blanchard, F. C. Rind, and P. F. M. J. Verschure, "Collision avoidance using a model of the locust LGMD neuron," *Robotics and Autonomous Systems*, vol. 30, no. 1, pp. 17–38, 2000.

[34] S. Yue, F. C. Rind, M. S. Keil, J. Cuadri, and R. Stafford, "A bio-inspired visual collision detection mechanism for cars: optimisation of a model of a locust neuron to a novel environment," *Neurocomputing*, vol. 69, no. 13–15, pp. 1591–1598, 2006.

[35] R. Stafford, R. D. Santer, and F. C. Rind, "A bio-inspired visual collision detection mechanism for cars: combining insect inspired neurons to create a robust system," *BioSystems*, vol. 84, no. 2-3, pp. 164–171, 2007.

[36] S. Yue and F. C. Rind, "Visual motion pattern extraction and fusion for collision detection in complex dynamic scenes," *Computer Vision and Image Understanding*, vol. 104, no. 1, pp. 48–60, 2006.

[37] S. B. Cook, "Experiments on homing in the limpet Siphonaria normalis," *Animal Behaviour*, vol. 17, no. 4, pp. 679–682, 1969.

[38] M. Collett, T. S. Collett, S. Chameron, and R. Wehner, "Do familiar landmarks reset the global path integration system of desert ants?" *Journal of Experimental Biology*, vol. 206, no. 5, pp. 877–882, 2003.

[39] S. Cannicci, S. Fratini, and M. Vannini, "Short-range homing in fiddler crabs (Ocypodidae, genus Uca): a homing mechanism not based on local visual landmarks," *Ethology*, vol. 105, no. 10, pp. 867–880, 1999.

[40] T. S. Collett, P. Graham, and V. Durier, "Route learning by insects," *Current Opinion in Neurobiology*, vol. 13, no. 6, pp. 718–725, 2003.

[41] T. S. Collett and P. Graham, "Animal navigation: path integration, visual landmarks and cognitive maps," *Current Biology*, vol. 14, no. 12, pp. R475–R477, 2004.

[42] R. Stafford, M. S. Davies, and G. A. Williams, "Robustness of self-organised systems to changes in individual level behaviour: an example from real and simulated self-organised snail aggregations," Nature Precedings, 2009, http://precedings.nature.com/documents/3922/version/1.

[43] G. Edgar, D. Catherwood, H. Edgar, D. Nikolla, C. Alford, and D. Brookes, "Use it or lose it: selection of Information in decision-making," in *Proceedings of the 6th International Conference on Thinking*, Venice International University, Venice, Italy, August 2008.

[44] L. Chittka, P. Skorupski, and N. E. Raine, "Speed-accuracy tradeoffs in animal decision making," *Trends in Ecology & Evolution*, vol. 24, no. 7, pp. 400–407, 2009.

[45] B. McEwen and E. N. Lasley, *The End of Stress as We Know It*, Joseph Hendry Press, Washington, DC, USA, 2002.

[46] L. Hadany, T. Beker, I. Eshel, and M. W. Feldman, "Why is stress so deadly? An evolutionary perspective," *Proceedings of the Royal Society B*, vol. 273, pp. 881–885, 2005.

Hindawi Publishing Corporation
Advances in Artificial Intelligence
Volume 2010, Article ID 478107, 9 pages
doi:10.1155/2010/478107

Review Article

Investigating the Underlying Intelligence Mechanisms of the Biological Olfactory System

Yoshinari Makino and Masafumi Yano

Research Institute of Electrical Communication, Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai 980-8577, Japan

Correspondence should be addressed to Yoshinari Makino, tei@riec.tohoku.ac.jp

Received 8 September 2009; Revised 11 November 2009; Accepted 9 December 2009

Academic Editor: Naoyuki Sato

The brain is the center of intelligence that biological systems have acquired during their evolutionary history. In unpredictably changing environments, animals use it to recognize the external world and to make appropriate behavioral decisions. Understanding the mechanisms underlying biological intelligence is important for the development of artificial intelligence. Olfaction is one of the sensory modalities that animals use to locate distant objects. Because of its relative simplicity compared with other sensory modalities and the wealth of knowledge at cellular, network, system, and psychophysical levels, it is possible that the biological olfactory system would be understood comprehensively. This paper reviews our biological and computational works with a focus on the temporal aspects of olfactory information processing. In addition, the paper highlights that the "time" dimension is essential for the functioning of the olfactory information processing system in the real world.

1. Introduction

Consider animals, such as honeybees, butterflies and so forth, collecting flower nectar in fields. For simplicity, all flowers have the same visual cues (e.g., shape, size, color, etc.), but each of them has a distinct odor. Some flowers can have the same odor but are spatially distributed randomly in the field. The quantity of flower nectar is fixed according to the type of odor; for example, flowers with *odor A* have 10 mL nectar, *odor B* have 5 mL, and *odor C* have 0 mL. There might be some dangerous flowers, such as insectivorous flowers, giving off specific odors. It is important to be aware of the cognitive abilities that an animal needs to possess in order to collect nectar efficiently and avoid danger.

An animal can sense an odor at various distances from its source, implying that an odor can be sensed at different concentrations. To approach or avoid flowers with a specific odor, the animal should be able to judge the odor at different concentrations as an odor with the same quality and should be able to determine its direction based on whether the odor signal becomes stronger or weaker. Thus, the animal must have the ability to identify odors in a concentration-invariant manner.

Suppose that the animal has visited many flowers in a certain field and has experienced the relationship of odors with nectar quantities or dangers and then moves to a different field. How will the animal behave in the new field where there are no flowers with odors identical to those experienced in the previous field? Based on the experience in the previous field, the animal should visit flowers with odors similar to those associated with a large amount of nectar and avoid flowers with odors similar to those of dangerous flowers. Thus, the animal must detect similarities and differences between odors and make appropriate behavioral decisions. These examples demonstrate that, at large, olfactory information processing is a typical example of pattern recognition. Problems of pattern recognition are further complicated in the real world where unpredictable noises or perturbations occur. Even in such situations, animals use the olfactory information to make appropriate behavior decisions in order to survive in the real world.

Olfaction is the sensory modality that many animals use to locate distant objects. Structures of the olfactory system in the brain, especially those from peripheral olfactory receptors to the primary olfactory network, have common

characteristics among vertebrates, insects, terrestrial gastropods (slugs/snails), and so forth [1]. Odorant molecules are received by olfactory receptor neurons (ORNs) that are distributed in the peripheral organs (e.g., the nasal cavity in vertebrates, the antennae in insects, and the tentacles in slugs/snails) [1–4]. ORNs project axons to round-shaped structures called glomeruli in the primary olfactory network [1, 5]. Axons of ORNs that express the same odor receptor type converge in each glomerulus [6–8]. The primary olfactory network (the olfactory bulb in vertebrates and the antennal lobe, AL, in insects) consists of output neurons and local inhibitory interneurons [2]. These neurons extend their dendrites into glomeruli and receive inputs from ORNs. The glomeruli are distributed as a glomerular layer along the surface of the olfactory bulb in vertebrates [2, 3] and botryoidally arranged in the insect antennal lobe [2, 4]. Physiological studies using an optical recording technique revealed that even monomolecular odorants evoke broadly distributed activation across glomeruli [9–12], so that an odor input is first represented as a spatial pattern of glomerular activation at the primary olfactory network.

The primary olfactory network transforms the input spatial pattern into the spatiotemporal activity pattern of output neurons [2] and sends it to various brain regions [5]. In general, brain functions emerge by cooperative information processing in various brain regions. To understand how the biological olfactory system solves pattern recognition problems in the real world, it is important to clarify the following two issues.

(i) How do odors spatiotemporally activate the olfactory regions in the brain, and how does the odor information flow within the brain as a whole?

(ii) To solve the problems of pattern recognition, what kind of computational algorithm is needed? How does the biological network implement it? To perform the algorithm, how is the spatiotemporal odor representation in the primary olfactory network useful?

By understanding these issues comprehensively, the intelligence that the biological olfactory system has acquired evolutionarily to solve pattern recognition problems in the real world can be explained.

This paper reviews physiological and computational studies on olfaction from these perspectives. Section 2 describes an experimental approach to explain the flow of olfactory information in the slug's brain. Section 3 describes the computational coding scheme of olfactory information using the time dimension. Section 4 summarizes the results and provides directions for future work.

2. Information Flows in the Whole Brain

Understanding the flow of information within the brain as a whole is one of the goals that neuroscience research addresses, and for this purpose, several techniques have been developed to measure brain activities. However, it is still hard to observe the flow of information in the whole brain because of the following difficulties: there is a tradeoff between spatial and temporal resolutions in any method, and the brain is too large and complex to be easily studied. So, it is important to select appropriate animals for measurement. In cases where the object to be measured is small, the optical imaging technique has a good spatiotemporal resolution. If we can use an animal that has a good learning ability, can discriminate objects, and has a small brain that is simple in structure, the flow of information can be visualized in such a brain, and it will be possible to explain the information processing principle in the brain as a whole.

From this perspective, a slug/snail is a useful animal model. In these animals, olfaction is a dominant sensory modality for recognizing external objects (visual or auditory systems have not been developed in their brain) [1, 13, 14]. The slug/snail has a good odor learning ability [15–20]. It has a small and simple brain; so whole brain activity can be measured with good spatiotemporal resolutions [21–26]. Furthermore, there are several experimental advantages of their olfactory system. Noses of these animals are located on the tips of two pairs of tentacles (superior and inferior tentacles, STs and ITs). Several behavioral and physiological studies have revealed that there are functional differences between STs and ITs [27–31]; ST is involved in olfactory orientation [28], whereas IT is involved in learning [29] or retrieving odors [27, 31]. Their brain can be dissected and isolated as a whole without a lose in the function of these olfactory organs [23, 24, 27, 32, 33]. These features are useful for explaining the relationship between the flow of information and the emergence of brain functions.

Figure 1(a) shows the slug we used, the Japanese slug *Incilaria fruhstorferi*, and Figure 1(b) schematically illustrates the slug's brain, its cerebral ganglia (CG). The STs and ITs are connected to CG via superior and inferior tentacle nerves (STNs and ITNs), respectively. CG comprises three lobes: the procerebrum (PC), mesocerebrum (MsC), and metacerebrum (MtC). Anatomical studies indicate that after entering the body of CG, afferent fibers of the tentacle nerves segregate into several bundles and terminate in PC and MtC [1, 14, 35]. The PC is regarded as the olfactory center, and many studies have investigated its role in the processing of olfactory information [22–24, 27, 30, 36–40]. It is suggested that PC is involved in acquiring and retrieving odor memory [27, 30, 38].

In contrast, few studies have investigated olfactory information processing in MtC. Anatomically, afferent fibers from STN and ITN converge into the medial region of MtC (mMtC) (Figures 1(c) and 1(d)) [26]. The MtC is thought to collect olfactory, taste, and other sensory information and to command motor actions when appropriate signals are received [14]. The relationships between PC and MtC in the slug's brain might correspond to those between cortical and subcortical structures in the vertebrate's brain. Hence, the slug's brain is suitable for our research purpose to observe the flow of information within the whole brain.

We optically recorded the brain activity evoked by electrical stimulations of STN and ITN and obtained the following results [26].

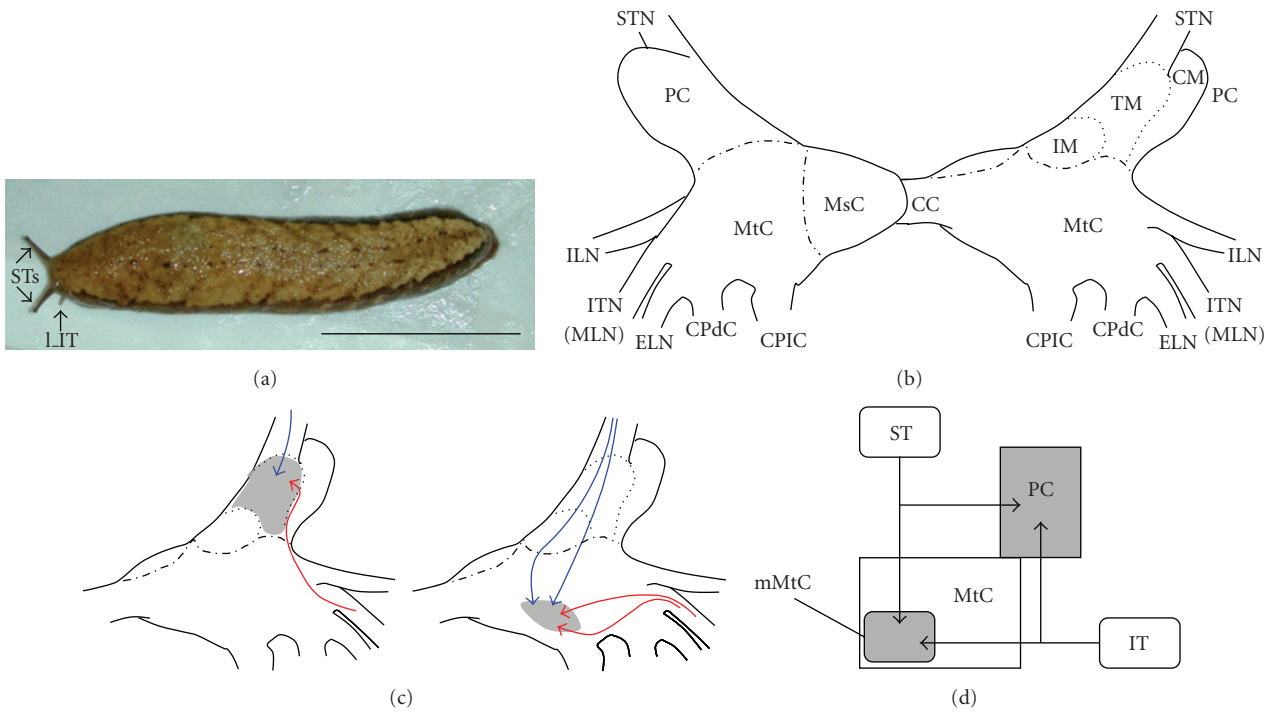


Figure 1: Slug's brain. (a) The Japanese giant terrestrial slug *Incilaria fruhstorferi*. A pair of superior tentacles (STs) can be seen at the head end. A part of the left inferior tentacle (l_IT) can be seen close to the base of the left ST. Scale bar: 5 cm. (b) Gross anatomy of the slug brain, cerebral ganglia. Left and right sides show dorsal and ventral views of the cerebral ganglia, respectively. The cerebral ganglion is divided into three lobes: the procerebrum (PC), mesocerebrum (MsC), and metacerebrum (MtC). Olfactory information received at the tips of STs and ITs reaches the brain through superior and inferior tentacle nerves (STNs and ITNs, resp.). See [26] for details. (c) Anatomical projections of fibers from STN (blue arrows) and ITN (red arrows). Both ST and IT send olfactory information into PC (left panel) and to the medial region of MtC (right panel). (d) Schematic illustration of anatomical olfactory projections from ST and IT in the brain. Figures 1(b) and 1(c) are adapted and modified, with permission, from [34] 2006 IEEE.

(1) STN and ITN stimulations activate both PC and mMtC.

(2) Regardless of STN or ITN stimulations, the mMtC response is about 50 milliseconds earlier than the PC response.

(3) STN and ITN stimulations evoked different activation patterns of mMtC: the ITN stimulation activated the lateral half of mMtC more strongly than its medial half, whereas the STN stimulation activated both halves evenly. In contrast, there seems to be no difference between the activation patterns in PC evoked by STN and ITN stimulations.

It is interesting that PC responses to STN and ITN stimulations are the same, in spite of memory functions of PC [27, 30, 38, 40] and functional differences between STs (orientation) and ITs (memory) [27–31]. This implies that direct olfactory inputs to PC would not explain the differences in memory functions between STs and ITs.

Since mMtC responds faster than PC to the nerve stimulation, there is a possibility that mMtC activations evoked by STN and ITN stimulations affect PC differently, and as a result, a functional difference between ST and IT for memory functions might emerge. We hypothesized that monoamine- (such as serotonin, dopamine, etc.) containing neurons might mediate the transfer of information from mMtC to PC because, as neuromodulators, monoamines are known to have important roles in memory functions [41–43] and in changing the functional modes of neural networks [44–46]. Therefore, we stained the dopamine-containing neurons in the slug's brain [34]. Main results are summarized in Figure 2. The dopamine-containing neuropils are intensively distributed in the central and lateral regions of MtC (Figure 2(a), dotted area), and this central MtC region seems to spatially overlap with lateral area of the mMtC region that was strongly activated by the ITN stimulation (Figure 2(a)). The dopamine containing output fibers from the central MtC region seem to project into PC. Gelperin and his colleagues physiologically suggested that dopamine might modulate the network function of the PC [47, 48]. So, the revealed morphological features suggest that the dopamine-containing neurons might act as a bridge between mMtC and PC, especially when the mMtC region is activated through the ITs [Note 1]. Dopamine-containing fibers project densely into the STN and ITN, and these come at the tentacle ganglion of ST and IT, suggesting another possibility that dopaminergic modulation of olfactory infor-

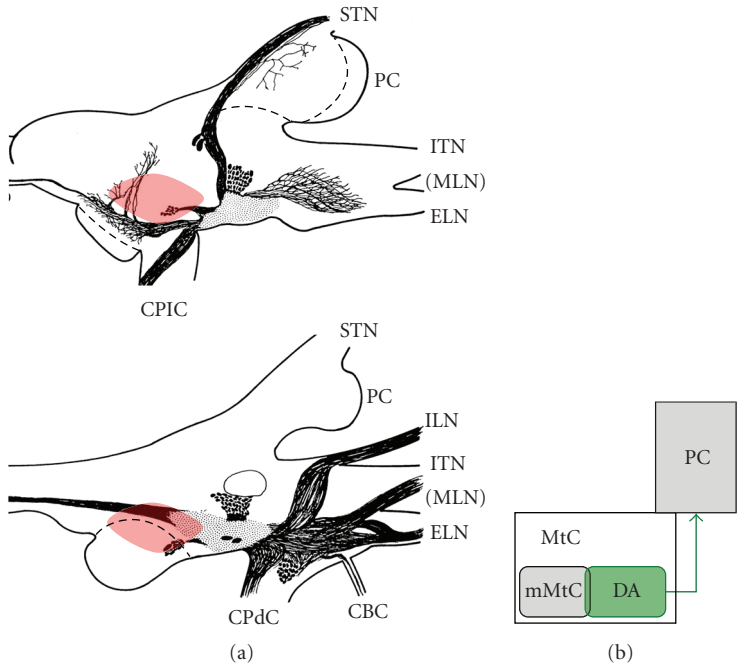


Figure 2: A major dopamine system in the slug brain. (a) Schematic illustrations of the major dopamine system. These figures are made on the basis of brain sections treated using the histofluorescent method for staining catecholamine (dopamine). Top and bottom panels are based on sections from dorsal and ventral parts, respectively. Filled ovals indicate dopamine-containing cell bodies. Dotted area indicates the area that the dopamine-containing fibers are densely distributed. The estimated mMtC region that would be activated by the STN or ITN stimulation is indicated by red shade. A large open-oval in the bottom indicates the metacerebral giant cell. (b) Function of the major dopamine system (DA) suggested by its morphology. The major dopamine system might function as a bridge between the medial region of MtC (mMtC) and the procerebrum (PC). Figure 2(a) is adapted and modified, with permission, from [34] 2006 IEEE.

mation processing might occur more peripherally through the tentacle ganglia.

Olfactory functions in the slug brain can be explained on the basis of the flow of olfactory information (Figure 3). Based on the mMtC activity, ST and IT would have different roles in olfactory functions. The medial half of mMtC (that is strongly activated by the STN stimulation) might be related to olfactory orientation. Indeed, several motor neurons related to olfactory orientation movements are known in regions close to mMtC [49–51]. The lateral half of mMtC (that is strongly activated by the ITN stimulation) might activate the monoamine system, and this system would modulate functional modes of PC to acquire or retrieve odor memory.

It is suggested that the brain (cerebral ganglia) can evaluate important odors and can command motor actions without PC, even though the odor discriminating ability of the brain without PC is not good compared to that with PC [49]. Based on this observation and data from our study [26], it can be hypothesized that mMtC might contribute to the rough evaluation of important odors. The quick response of mMtC to STN and ITN stimulations might result in quick responses of the slug/snail to important odors and limit the odors that are processed by PC for fine discrimination [26]. Such a temporal process would be ecologically advantageous, since various odors exist in the real world.

3. Information Representation Using Time Dimension Solves Pattern-Recognition Problems

There is growing evidence that biological systems use time as a dimension for sensory information coding [52–57]. Computationally, a central aspect of pattern recognition is achieving a "good" representation in which the main features of an object are simply and naturally revealed [58, 59]. Neurobiological information representation is crucial in facilitating such a computation [58, 60].

Regardless of biological or artificial systems, recognition systems must extract invariant features of objects from varying input signals depending on situations and must detect and judge similarities and differences between objects appropriately. These requirements are for a system that can identify objects and make appropriate behavioral decisions in the real world. Thus, the temporal representation of objects should satisfy certain computational criteria for it to be useful for recognition. Criterion 1 is invariance. For instance, the size or strength of sensory signals varies when the same object is sensed from different distances. Even in this situation, object representation should be invariant. Criterion 2 is similarity. The degree of similarity of objects must be reflected in their representation while also expressing subtle differences. Regardless of sensory modalities, biological recognition shows a speed/accuracy

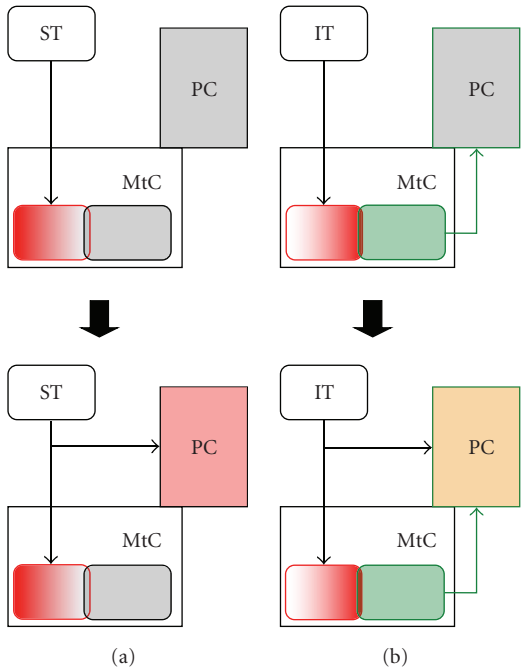


Figure 3: The flow of olfactory information in the slug brain. Brain activities evoked by superior tentacle (ST) (a) and inferior tentacle (IT) (b) stimulations in early phase (top) and late phase (bottom). Based on the activation patterns evoked by ST and IT stimulations, it was suggested that only the IT stimulation would activate the dopamine system in the early phase (a), (b) top. It may be possible to modulate the functional mode of the procerebrum (PC) activation by the activated dopamine system, so that the processing of olfactory information at PC differs between ST and IT stimulations (a), (b) bottom.

tradeoff, which is the relationship between sampling time and accuracy in object discrimination [61–63]. This psychophysical principle constrains the temporal representation format of the biological system by imposing Criterion 3, which is coarse-to-fine nature. The coarse similarity or difference of objects should be encoded in the early epoch of the temporal representation, and the information encoded in the late epoch should help in detecting subtle differences among objects.

The question arises as to what kind of information representation format would satisfy the three representation criteria. Suppose that there is a set of n feature dimensions each of which describes a specific object feature. By preprocessing of sensory signals from an object, the intensity of each feature is obtained, and consequently, the input representation of the object is an analog vector of n-dimensional feature space. Changing the signal intensity or size from the same object would only cause changes in the coefficients or bias of input analog vectors. In this case, the order of feature dimensions (in other words, the ranking of feature dimensions) determined by elemental analog values of the input vector would be scale invariant. So, if a computational algorithm transforms the order of analog values of the input vector into a temporal sequence of spike activity of neurons (or neural assemblies), each of which controls the respective feature dimension (Figure 4(a)), then the emergent temporal representation has scale invariance, satisfying Criterion 1. A temporal sequence of neural activities represents the hierarchical relationship of feature dimensions: as time advances, an object is coarsely classified on the basis of a feature represented by the first activated neuron, further subclassified by subsequent ones, and finally identified by taking all activated neurons into account (Figure 4(a), bottom), implying that the representation satisfies Criteria 2 and 3.

Using the n-dimensional feature space, a system adopting such a coding scheme can theoretically represent and discriminate $n!$ objects ("!" is the factorial operator). If the system can set a temporal range for object representation, it can control the clustering levels of objects and make appropriate behavioral decisions that are flexible according to situations.

The proposed coding scheme is highly consistent with information transformation in the primary olfactory network of biological systems, especially in that of insects and vertebrate zebra fish [52–55]. As mentioned in Section 1, odor input representation is the spatial pattern of glomerular activation [9–12]. Even monomolecular odorants evoke a broadly distributed activation across glomeruli, such that each glomerulus is regarded as a functional module detecting a particular feature of odorant molecular structure [3, 9–12, 64]. Odors at very low concentrations activate a few glomeruli above the background level. As the concentration increases, several additional glomeruli are recruited, and the glomerular activity is a logarithmic function of concentration [9, 11, 65–67]. Furthermore, similarity in the molecular structures of different odor molecules is correlated with similarity in spatial patterns of glomerular activation [10, 11, 68, 69] and with perceptual similarity [69–71]. Spatiotemporal activity patterns of the insect primary olfactory network are robust to odor concentration changes [54], whereas those of the zebra fish primary olfactory network are suggested to reflect odor similarity in a temporally coarse-to-fine manner [53]. These imply that the input spatial pattern of odor-evoked glomerular activation can be represented as an analog vector with a concentration bias [67]; the glomerular ranking that can be determined by the odor-evoked glomerular activation would be concentration invariant [72–74]; if the glomerular ranking can be transformed into temporal sequence of neural activities according to the proposed coding algorithm, the emergent spatiotemporal patterns would reflect the odor similarity as physiologically observed [72–74].

By means of biologically plausible neural components and their interactions, we constructed a primary olfactory network model that can implement the coding algorithm shown in Figure 4(a) for odor information. The model consists of glomerular modules, each consisting of one output neuron and two types (intraglomrular and interglomerular types) of inhibitory local interneurons. In the single glomerular module model, the intraglomerular inhibition helps to cause a phasic activity of the output neuron in response to a constant glomerular input, and onset latency

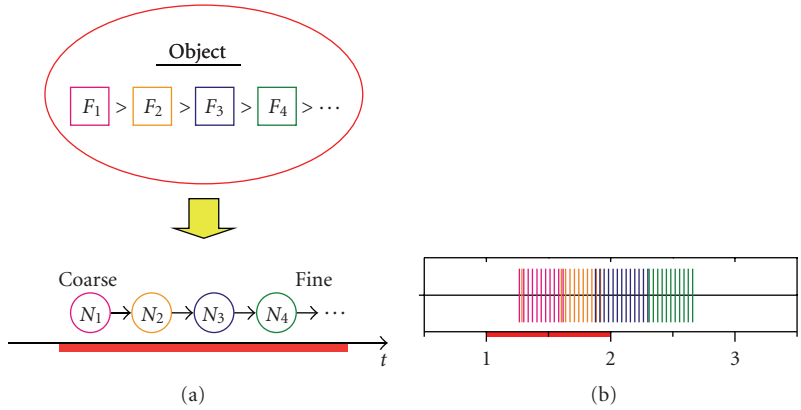


Figure 4: Object representation using temporal sequence of neural activities. (a) The order of feature intensities of an object (e.g., $F_1 > F_2 > F_3 > F_4$) is transformed into a temporal sequence of neural activities in control of respective features (e.g., $N_1 \rightarrow N_2 \rightarrow N_3 \rightarrow N_4$). The emergent temporal sequence of neural activities represents a hierarchical relationship of object features in coarse-to-fine manner on the temporal axis. (b) A typical simulated result of spike activities of output neurons of the four-glomerular module model. Spikes of each output neuron are indicated in different colors.

of the phasic response of the output neuron decreases as the input strength increases. In the network model, the glomerular modules are connected by the interglomerular inhibitory interneurons. Through these interglomerular connections, the activated output neuron inhibits other glomerular output neurons and delays their phasic activation, and consequently phasic activities of the output neurons are ordered according to the strength order of the glomerular inputs (Figure 4(b)). The network model successfully implements this spatiotemporal transformation regardless of absolute input analog values or concentration biases [72–74].

In the insect higher center network, that is, the mushroom body (MB), odors are represented by spikes sparsely distributed in space and time [53]. In our primary olfactory network model, the odor information is represented as the temporal sequence of the output neurons' activities, in which the information about the activity sequence is intensively-coded around transient overlaps of the neural activities (see Figure 4(b)). So, we further constructed a higher center model that evaluates the transient overlaps of the neural activities of the output neurons of the primal olfactory network model. The higher center model can extract odor information as a temporal sequence of sparse spikes with a coarse-to-fine nature and can store the odor information as network connectivity with the help of simple Hebbian connections [72–74].

These indicate that the essential problems of olfactory recognition can be solved by information representation using the time dimension. In the proposed coding scheme, significant features characterizing objects are temporally placed first, followed by more subtle features. This strategy is crucial for the system to survive in an unpredictably changing environment because it allows the system to respond quickly to vitally important object features. Even though it may be just one dimension mathematically, "time" is an important dimension for systems that live and survive in the real world. Thus, the information representation using the time dimension is a basis for pattern recognition that warrants further investigation in biological systems.

4. Towards an Understanding of Biological System Intelligence

In Section 2, we have shown that the olfactory system in the slug's brain consists of two pathways: one is the mMtC-pathway that probably has a role for quickly and roughly discriminating important odors; the other is the PC-pathway that is rather slow but important for finely discriminating odors. With respect to the computational model shown in Section 3, the physiological data about the spatiotemporal neural activities are mainly based on those obtained from the insect AL-MB pathway [52, 54, 55]. The insect MB is suggested to have important roles in fine discrimination of odors [75, 76] and odor learning and memory [77–79], as the slug/snail PC is suggested to have such roles [33, 38, 40]. So, the computational model in Section 3 is thought to correspond to the PC-pathway in Section 2, and we can suggest that the olfactory system of the brain uses "time" dimension doubly (i.e., the anatomically differentiated two pathways that have different temporal responses to odors, and the coarse-to-fine odor representation using the temporal sequence of neural activity in the fine discrimination pathway) to quickly and dominantly process important odors or odor-features. These temporal aspects of olfactory information processing might be essential for systems working in the real world.

To comprehensively understand the mechanisms involved in olfactory information processing, we must explain how global information flow and computational information processing are related and integrated on the systems level. In the information representation proposed in Section 3, odor is described as the permutation of odor features, each of which works like letters of an alphabet.

So, information represented by the temporal sequence of neuronal activities is regarded as symbolic information. Symbolic information is useful for identifying objects and detecting their similarities and differences. However, symbolic information of objects itself cannot help the system make a behavioral decision because it does not represent information on the values of objects. We call this "value information." Value information cannot be defined by attributes of objects alone. Value information of an object should be evaluated and determined by the system in real time on the basis of experience with the object and current conditions of the environment and the system. From invertebrates (including the slugs and snails mentioned in Section 2) to vertebrates (including primates), monoamine systems in the brain are known to be strongly related to internal and motivational states of biological systems and to play important roles in learning and memory [32, 41–43, 80–86], so that monoamine systems might be involved in determining the value information of objects. As neuromodulators, monoamines would affect the brain activity in a wide spatial and temporal range, and their spatiotemporal dynamics in the whole brain would be critical for neural network functions [84]. The challenge is to explain biological and computational mechanisms relating temporally represented symbolic information with value information that might be determined by global information flows within the brain. This approach would clarify the design principles of an intelligent system that works well in the unpredictably changing environment of the real world.

Note 1. This hypothesis predicts that the PC should respond differently between the STN and ITN stimulation by the indirect input through the mMtC. However, we could not observe such difference. In the experiment, both STN and ITN stimulations evoked transient depolarization followed by the strong and long lasting (about 10 seconds) hyperpolarization in the PC [26]. The transient depolarization of the PC is caused by the direct input from the STN and ITN, whereas the following PC hyperpolarization is probably caused by inhibitory effects of the intrinsic PC neurons [87]. In the experiment of [26], this intrinsic hyperporlaization in the PC might mask the PC response caused by the mMtC pathway.

Acknowledgments

This study was supported by a Grant no. 21500294 for Y.Makino from the Japan Society for the Promotion of Science, and by a Grant no. 17075003 for Y.Makino and M.Yano from the Japanese Ministry of Education, Culture, Sports, Science, and Technology.

References

[1] R. Chase and B. Tolloczko, "Tracing neural pathways in snail olfaction: from the tip of the tentacles to the brain and beyond," *Microscopy Research and Technique*, vol. 24, no. 3, pp. 214–230, 1993.

[2] G. Laurent, "Olfactory network dynamics and the coding of multidimensional signals," *Nature Reviews Neuroscience*, vol. 3, no. 11, pp. 884–895, 2002.

[3] K. Mori, Y. K. Takahashi, K. M. Igarashi, and M. Yamaguchi, "Maps of odorant molecular features in the mammalian olfactory bulb," *Physiological Reviews*, vol. 86, no. 2, pp. 409–433, 2006.

[4] L. B. Vosshall and R. F. Stocker, "Molecular architecture of smell and taste in *Drosophila*," *Annual Review of Neuroscience*, vol. 30, pp. 505–533, 2007.

[5] G. M. Shepherd, *Neurobiology*, Oxford University Press, New York, NY, USA, 2nd edition, 1988.

[6] L. Buck and R. Axel, "A novel multigene family may encode odorant receptors: a molecular basis for odor recognition," *Cell*, vol. 65, no. 1, pp. 175–187, 1991.

[7] R. Vassar, S. K. Chao, R. Sitcheran, J. M. Nunez, L. B. Vosshall, and R. Axel, "Topographic organization of sensory projections to the olfactory bulb," *Cell*, vol. 79, no. 6, pp. 981–991, 1994.

[8] L. B. Vosshall, A. M. Wong, and R. Axel, "An olfactory sensory map in the fly brain," *Cell*, vol. 102, no. 2, pp. 147–159, 2000.

[9] R. W. Friedrich and S. I. Korsching, "Combinatorial and chemotopic odorant coding in the zebrafish olfactory bulb visualized by optical imaging," *Neuron*, vol. 18, no. 5, pp. 737–752, 1997.

[10] C. G. Galizia, S. Sachse, A. Rappert, and R. Menzel, "The glomerular code for odor representation is species specific in the honeybee Apis mellifera," *Nature Neuroscience*, vol. 2, no. 5, pp. 473–478, 1999.

[11] B. D. Rubin and L. C. Katz, "Optical imaging of odorant representations in the mammalian olfactory bulb," *Neuron*, vol. 23, no. 3, pp. 499–511, 1999.

[12] N. Uchida, Y. K. Takahashi, M. Tanifuji, and K. Mori, "Odor maps in the mammalian olfactory bulb: domain organization and odorant structural features," *Nature Neuroscience*, vol. 3, no. 10, pp. 1035–1043, 2000.

[13] R. Chase, "The olfactory sensitivity of snails, *Achatina fulica*," *Journal of Comparative Physiology A*, vol. 148, no. 2, pp. 225–235, 1982.

[14] R. Chase, "Structure and function in the cerebral ganglion," *Microscopy Research and Technique*, vol. 49, no. 6, pp. 511–520, 2000.

[15] A. Gelperin, "Rapid food aversion learning by a terrestrial mollusk," *Science*, vol. 189, no. 4202, pp. 567–570, 1975.

[16] C. Sahley, J. W. Rudy, and A. Gelperin, "An analysis of associative learning in a terrestrial mollusc—I. Higher-order conditioning, blocking and a transient US pre-exposure effect," *Journal of Comparative Physiology A*, vol. 144, no. 1, pp. 1–8, 1981.

[17] C. L. Sahley, K. A. Martin, and A. Gelperin, "Analysis of associative learning in the terrestrial mollusc Limax maximus—II. Appetitive learning," *Journal of Comparative Physiology A*, vol. 167, no. 3, pp. 339–345, 1990.

[18] A. Yamada, T. Sekiguchi, H. Suzuki, and A. Mizukami, "Behavioral analysis of internal memory states using cooling-induced retrograde amnesia in *Limax flavus*," *Journal of Neuroscience*, vol. 12, no. 3, pp. 729–735, 1992.

[19] H. Suzuki, T. Sekiguchi, A. Yamada, and A. Mizukami, "Sensory preconditioning in the terrestrial mollusk, *Limax flavus*," *Zoological Science*, vol. 11, pp. 121–125, 1994.

[20] T. Teyke, "Food-attraction conditioning in the snail, *Helix pomatia*," *Journal of Comparative Physiology A*, vol. 177, no. 4, pp. 409–414, 1995.

[21] D. Kleinfeld, K. R. Delaney, M. S. Fee, J. A. Flores, D. W. Tank, and A. Gelperin, "Dynamics of propagating waves in

the olfactory network of a terrestrial mollusk: an electrical and optical study," *Journal of Neurophysiology*, vol. 72, no. 3, pp. 1402–1419, 1994.

[22] S. Kawahara, S. Toda, Y. Suzuki, S. Watanabe, and Y. Kirino, "Comparative study on neural oscillation in the procerebrum of the terrestrial slugs *Incilaria bilineata* and *Limax marginatus*," *Journal of Experimental Biology*, vol. 200, no. 13, pp. 1851–1861, 1997.

[23] S. Toda, S. Kawahara, and Y. Kirino, "Image analysis of olfactory responses in the procerebrum of the terrestrial slug *Limax marginatus*," *Journal of Experimental Biology*, vol. 203, no. 19, pp. 2895–2905, 2000.

[24] E. S. Nikitin and P. M. Balaban, "Optical recording of odor-evoked responses in the olfactory brain of the naive and aversively trained terrestrial snails," *Learning and Memory*, vol. 7, no. 6, pp. 422–432, 2000.

[25] S. Watanabe, S. Shimozono, and Y. Kirino, "Optical recording of oscillatory neural activities in the molluscan brain," *Neuroscience Letters*, vol. 359, no. 3, pp. 147–150, 2004.

[26] H. Makinae, Y. Makino, T. Obara, and M. Yano, "Specific spatio-temporal activities in the cerebral ganglion of *Incilaria fruhstorferi* in response to superior and inferior tentacle nerve stimulation," *Brain Research*, vol. 1231, pp. 47–62, 2008.

[27] T. Kimura, S. Toda, T. Sekiguchi, S. Kawahara, and Y. Kirino, "Optical recording analysis of olfactory response of the procerebral lobe in the slug brain," *Learning and Memory*, vol. 4, no. 5, pp. 389–400, 1998.

[28] R. Chase and R. P. Croll, "Tentacular function in snail olfactory orientation," *Journal of Comparative Physiology A*, vol. 143, no. 3, pp. 357–362, 1981.

[29] A. Friedrich and T. Teyke, "Identification of stimuli and input pathways mediating food-attraction conditioning in the snail, Helix," *Journal of Comparative Physiology A*, vol. 183, no. 2, pp. 247–254, 1998.

[30] T. Kimura, S. Toda, T. Sekiguchi, and Y. Kirino, "Behavioral modulation induced by food odor aversive conditioning and its influence on the olfactory responses of an oscillatory brain network in the slug *Limax marginatus*," *Learning and Memory*, vol. 4, no. 5, pp. 365–375, 1998.

[31] T. Kimura, A. Iwama, and T. Sekiguchi, "Contributions of superior and inferior tentacles to learned food-avoidance behavior in *Limax marginatus*," *Zoological Science*, vol. 16, no. 4, pp. 595–602, 1999.

[32] M. E. Egan and A. Gelperin, "Olfactory inputs to a bursting serotonergic interneuron in a terrestrial mollusk," *Journal of Molluscan Studies*, vol. 47, pp. 80–88, 1981.

[33] T. Teyke and A. Gelperin, "Olfactory oscillations augment odor discrimination not odor identification by Limax CNS," *NeuroReport*, vol. 10, no. 5, pp. 1061–1068, 1999.

[34] Y. Makino, H. Makinae, T. Obara, H. Miura, and M. Yano, "Observations of olfactory information flows within brain of the terrestrial slug, Inciralia fruhstorferi," in *Proceedings of IEEE International Conference on Neural Networks (IJCNN '06)*, pp. 3874–3881, Vancouver, Canada, July 2006.

[35] R. Chase and B. Tolloczko, "Interganglionic dendrites constitute an output pathway from the procerebrum of the snail *Achatina fulica*," *Journal of Comparative Neurology*, vol. 283, no. 1, pp. 143–152, 1989.

[36] R. Chase, "Responses to odors mapped in snail tentacle and brain by [^{14}C]-2-deoxyglucose autoradiography," *Journal of Neuroscience*, vol. 5, no. 11, pp. 2930–2939, 1985.

[37] A. Galperin and D. W. Tank, "Odour-modulated collective network oscillations of olfactory interneurons in a terrestrial mollusc," *Nature*, vol. 345, no. 6274, pp. 437–440, 1990.

[38] T. Kimura, H. Suzuki, E. Kono, and T. Sekiguchi, "Mapping of interneurons that contribute to food aversive conditioning in the slug brain," *Learning and Memory*, vol. 4, no. 5, pp. 376–388, 1998.

[39] A. Schütt, E. Başar, and T. H. Bullock, "Power spectra of ongoing activity of the snail brain can discriminate odorants," *Comparative Biochemistry and Physiology A*, vol. 123, no. 1, pp. 95–110, 1999.

[40] Y. Kasai, S. Watanabe, Y. Kirino, and R. Matsuo, "The procerebrum is necessary for odor-aversion learning in the terrestrial slug *Limax valentianus*," *Learning and Memory*, vol. 13, no. 4, pp. 482–488, 2006.

[41] S. A. Siegelbaum, J. S. Camardo, and E. R. Kandel, "Serotonin and cyclic AMP close single K^+ channels in Aplysia sensory neurones," *Nature*, vol. 299, no. 5882, pp. 413–417, 1982.

[42] M. Hammer, "An identified neuron mediates the unconditioned stimulus in associative olfactory learning in honeybees," *Nature*, vol. 366, no. 6450, pp. 59–63, 1993.

[43] W. Schultz, P. Apicella, and T. Ljungberg, "Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task," *Journal of Neuroscience*, vol. 13, no. 3, pp. 900–913, 1993.

[44] E. Marder and S. Hooper, "Neurotransmitter modulation of the stomatogastric ganglion of decapod crustaceans," in *Model Neural Networks and Behavior*, A. I. Selverston, Ed., pp. 319–337, Plenum Press, New York, NY, USA, 1985.

[45] R. E. Flamm and R. M. Harris-Warrick, "Aminergic modulation in lobster stomatogastric ganglion—I. Effects on motor pattern and activity of neurons within the pyloric circuit," *Journal of Neurophysiology*, vol. 55, no. 5, pp. 847–865, 1986.

[46] Y. Makino, M. Akiyama, and M. Yano, "Emergent mechanisms in multiple pattern generation of the lobster pyloric network," *Biological Cybernetics*, vol. 82, no. 6, pp. 443–454, 2000.

[47] A. Gelperin, L. D. Rhines, J. Flores, and D. W. Tank, "Coherent network oscillations by olfactory interneurons: modulation by endogenous amines," *Journal of Neurophysiology*, vol. 69, no. 6, pp. 1930–1939, 1993.

[48] L. D. Rhines, P. G. Sokolove, J. Flores, D. W. Tank, and A. Gelperin, "Cultured olfactory interneurons from Limax maximus: optical and electrophysiological studies of transmitter-evoked responses," *Journal of Neurophysiology*, vol. 69, no. 6, pp. 1940–1947, 1993.

[49] M. Peschel, V. Straub, and T. Teyke, "Consequences of food-attraction conditioning in Helix: a behavioral and electrophysiological study," *Journal of Comparative Physiology A*, vol. 178, no. 3, pp. 317–327, 1996.

[50] S. A. Prescott, N. Gill, and R. Chase, "Neural circuit mediating tentacle withdrawal in *Helix aspersa*, with specific reference to the competence of the motor neuron C3," *Journal of Neurophysiology*, vol. 78, no. 6, pp. 2951–2965, 1997.

[51] E. S. Nikitin, I. S. Zakharov, E. I. Samarova, G. Kemenes, and P. M. Balaban, "Fine tuning of olfactory orientation behaviour by the interaction of oscillatory and single neuronal activity," *European Journal of Neuroscience*, vol. 22, no. 11, pp. 2833–2844, 2005.

[52] G. Laurent, M. Wehr, and H. Davidowitz, "Temporal representations of odors in an olfactory network," *Journal of Neuroscience*, vol. 16, no. 12, pp. 3837–3847, 1996.

[53] R. W. Friedrich and G. Laurent, "Dynamic optimization of odor representations by slow temporal patterning of mitral cell activity," *Science*, vol. 291, no. 5505, pp. 889–894, 2001.

[54] M. Stopfer, V. Jayaraman, and G. Laurent, "Intensity versus identity coding in an olfactory system," *Neuron*, vol. 39, no. 6, pp. 991–1004, 2003.

[55] J. Perez-Orive, O. Mazor, G. C. Turner, S. Cassenaer, R. I. Wilson, and G. Laurent, "Oscillations and sparsening of odor representations in the mushroom body," *Science*, vol. 297, no. 5580, pp. 359–365, 2002.

[56] J. Hegdé and D. C. Van Essen, "Temporal dynamics of shape analysis in macaque visual area V2," *Journal of Neurophysiology*, vol. 92, no. 5, pp. 3030–3042, 2004.

[57] M. D. Menz and R. D. Freeman, "Stereoscopic depth processing in the visual cortex: a coarse-to-fine mechanism," *Nature Neuroscience*, vol. 6, no. 1, pp. 59–65, 2003.

[58] R. D. Luce, *Vision*, Freeman, New York, NY, USA, 1982.

[59] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, John Wiley & Sons, New York, NY, USA, 2nd edition, 2001.

[60] J. J. Hopfield, "Pattern recognition computation using action potential timing for stimulus representation," *Nature*, vol. 376, no. 6535, pp. 33–36, 1995.

[61] R. D. Luce, *Response Times*, Oxford University Press, Oxford, UK, 1986.

[62] N. M. Abraham, H. Spors, A. Carleton, T. W. Margrie, T. Kuner, and A. T. Schaefer, "Maintaining accuracy at the expense of speed: stimulus similarity defines odor discrimination time in mice," *Neuron*, vol. 44, no. 5, pp. 865–876, 2004.

[63] D. Rinberg, A. Koulakov, and A. Gelperin, "Speed-accuracy tradeoff in olfaction," *Neuron*, vol. 51, no. 3, pp. 351–358, 2006.

[64] K. Mori and G. M. Shepherd, "Emerging principles of molecular signal processing by mitral/tufted cells in the olfactory bulb," *Seminars in Cell and Developmental Biology*, vol. 5, no. 1, pp. 65–74, 1994.

[65] M. Wachowiak, L. B. Cohen, and M. R. Zochowski, "Distributed and concentration-invariant spatial representations of odorants by receptor neuron input to the turtle olfactory bulb," *Journal of Neurophysiology*, vol. 87, no. 2, pp. 1035–1045, 2002.

[66] S. Sachse and C. G. Galizia, "The coding of odour-intensity in the honeybee antennal lobe: local computation optimizes odour representation," *European Journal of Neuroscience*, vol. 18, no. 8, pp. 2119–2132, 2003.

[67] C. D. Brody and J. J. Hopfield, "Simple networks for spike-timing-based computation, with application to olfactory processing," *Neuron*, vol. 37, no. 5, pp. 843–852, 2003.

[68] S. Sachse, A. Rappert, and C. G. Galizia, "The spatial representation of chemical structures in the antennal lobe of honeybees: steps towards the olfactory code," *European Journal of Neuroscience*, vol. 11, no. 11, pp. 3970–3982, 1999.

[69] F. Guerrieri, M. Schubert, J. C. Sandoz, and M. Giurfa, "Perceptual and neural olfactory similarity in honeybees," *PLoS Biology*, vol. 3, no. 4, article e60, 2005.

[70] M. Laska, C. G. Galizia, M. Giurfa, and R. Menzel, "Olfactory discrimination ability and odor structure-activity relationships in honeybees," *Chemical Senses*, vol. 24, no. 4, pp. 429–438, 1999.

[71] C. Linster and M. E. Hasselmo, "Behavioral responses to aliphatic aldehydes can be predicted from known electrophysiological responses of mitral cells in the olfactory bulb," *Physiology & Behavior*, vol. 66, no. 3, pp. 497–502, 1999.

[72] Y. Makino, M. Yasuike, Y. Naka, H. Miura, and M. Yano, "Olfactory computation using spatiotemporal pattern of network activity for odor representation," *Neuroscience Research*, vol. 58, supplement, p. S103, 2007.

[73] Y. Makino, M. Yasuike, Y. Naka, H. Miura, and M. Yano, "Principal characteristics in odor recognition naturally emerge from spatiotemporal coding," *Neuroscience Research*, vol. 61, supplement, p. S249, 2008.

[74] Y. Makino, M. Yasuike, Y. Naka, H. Miura, and M. Yano, "A computational algorithm for odor representation using a spatiotemporal sequence," *Society for Neuroscience Abstract*, vol. 362, p. 15, 2008.

[75] K. MacLeod and G. Laurent, "Distinct mechanisms for synchronization and temporal patterning of odor-encoding neural assemblies," *Science*, vol. 274, no. 5289, pp. 976–979, 1996.

[76] M. Stopfer, S. Bhagavan, B. H. Smith, and G. Laurent, "Impaired odour discrimination on desynchronization of odour-encoding neural assemblies," *Nature*, vol. 390, no. 6655, pp. 70–74, 1997.

[77] J. S. de Belle and M. Heisenberg, "Associative odor learning in *Drosophila* abolished by chemical ablation of mushroom bodies," *Science*, vol. 263, no. 5147, pp. 692–695, 1994.

[78] J. B. Connolly, I. J. H. Roberts, J. D. Armstrong, et al., "Associative learning disrupted by impaired G_s signaling in *Drosophila* mushroom bodies," *Science*, vol. 274, no. 5295, pp. 2104–2107, 1996.

[79] J. Dubnau, L. Grady, T. Kitamoto, and T. Tully, "Disruption of neurotransmission in *Drosophila* mushroom body blocks retrieval but not acquisition of memory," *Nature*, vol. 411, no. 6836, pp. 476–480, 2001.

[80] R. Gillette and W. J. Davis, "The role of the metacerebral giant neuron in the feeding behavior of *Pleurobranchaea*," *Journal of Comparative Physiology A*, vol. 116, no. 2, pp. 129–159, 1977.

[81] K. R. Weiss and I. Kupfermann, "Homology of the giant serotonergic neurons (metacerebral cells) in *Aplysia* and pulmonate molluscs," *Brain Research*, vol. 117, no. 1, pp. 33–49, 1976.

[82] M. S. Livingstone, R. M. Harris-Warrick, and E. A. Kravitz, "Serotonin and octopamine produce opposite postures in lobsters," *Science*, vol. 208, no. 4439, pp. 76–79, 1980.

[83] C. M. Lent and M. H. Dickinson, "Serotonin integrates the feeding behavior of the medicinal leech," *Journal of Comparative Physiology A*, vol. 154, no. 4, pp. 457–471, 1984.

[84] G. Bicker and R. Menzel, "Chemical codes for the control of behaviour in arthropods," *Nature*, vol. 337, no. 6202, pp. 33–39, 1989.

[85] E. A. Kravitz, "Serotonin and aggression: insights gained from a lobster model system and speculations on the role of amine neurons in a complex behavior," *Journal of Comparative Physiology A*, vol. 186, no. 3, pp. 221–238, 2000.

[86] K. M. Crisp and K. A. Mesce, "To swim or not to swim: regional effects of serotonin, octopamine and amine mixtures in the medicinal leech," *Journal of Comparative Physiology A*, vol. 189, no. 6, pp. 461–470, 2003.

[87] S. Watanabe, S. Kawahara, and Y. Kirino, "Glutamate induces CI^- and K^+ currents in the olfactory interneurons of a terrestrial slug," *Journal of Comparative Physiology A*, vol. 184, no. 6, pp. 553–562, 1999.

Hindawi Publishing Corporation
Advances in Artificial Intelligence
Volume 2010, Article ID 765876, 20 pages
doi:10.1155/2010/765876

Research Article

Bootstrap Learning and Visual Processing Management on Mobile Robots

Mohan Sridharan

Department of Computer Science, Texas Tech University, Lubbock, TX 79409, USA

Correspondence should be addressed to Mohan Sridharan, mohan.sridharan@ttu.edu

Received 1 October 2009; Accepted 10 November 2009

Academic Editor: Alfons Schuster

A central goal of robotics and AI is to enable a team of robots to operate autonomously in the real world and collaborate with humans over an extended period of time. Though developments in sensor technology have resulted in the deployment of robots in specific applications the ability to accurately sense and interact with the environment is still missing. Key challenges to the widespread deployment of robots include the ability to learn models of environmental features based on sensory inputs, bootstrap off of the learned models to detect and adapt to environmental changes, and autonomously tailor the sensory processing to the task at hand. This paper summarizes a comprehensive effort towards such bootstrap learning, adaptation, and processing management using visual input. We describe probabilistic algorithms that enable a mobile robot to autonomously plan its actions to learn models of color distributions and illuminations. The learned models are used to detect and adapt to illumination changes. Furthermore, we describe a probabilistic sequential decision-making approach that autonomously tailors the visual processing to the task at hand. All algorithms are fully implemented and tested on robot platforms in dynamic environments.

1. Introduction

An open grand challenge in the field of robotics is to enable widespread deployment of robots in the real world, where they can operate autonomously and collaborate with humans. Addressing this grand challenge would in turn require answers to the following major questions.

(i) *Autonomous Learning and Adaptation*. How to enable a robot to autonomously learn models of environmental features based on sensory input, detect environmental changes, and adapt the learned models in response to such changes?

(ii) *Processing Management.* Given multiple sources of information, which bits of information should be processed, and what processing should be performed in order to achieve a desired goal reliably and efficiently?

(iii) *Multiagent Coordination.* How to enable a team of robots, each with possibly different capabilities and constraints, to collaborate robustly towards a shared objective despite noisy sensing and communication?

In this paper, the focus is primarily on developing probabilistic methods for *Autonomous Learning and Adaptation*, and for *Processing Management*. We propose probabilistic methods that enable a robot to use sensory inputs to learn environmental models and respond to environmental changes. Furthermore, given multiple sources of information, the robot autonomously tailors the sensory processing to the task at hand.

Mobile robots that sense and interact with the environment through a set of sensors and actuators are characterized by the following features and requirements.

(i) Features

(a) *Partial Observability*. The true state of the world is not directly observable. The robot can only update its *belief*, that is, an estimate of the world state by executing actions and observing the noisy outcomes.

(b) *Nondeterministic Actions and Observations.* The outcome of executing actions or making observations based on sensory input is nondeterministic, that is, actions and observations are unreliable.

(c) *Computational Complexity.* Many state-of-the-art sensory processing algorithms (e.g., vision algorithms) have high computational complexity.

(ii) Requirements.

(a) *Dynamic Performance.* Robots operating in the real world need to respond to the changes in their environment despite computational constraints; that is, there is a strong *real-time* requirement.

(b) *Reliability.* Though outcomes of actions and observations are nondeterministic, the robot needs to operate with a high degree of reliability, especially in critical applications such as disaster rescue or surveillance.

Developments in sensor technology [1, 2] have resulted in the deployment of mobile robots in specific applications such as disaster rescue, navigation, and medicine [3–6]. The ability to accurately sense and interact with the environment is however still lacking. The state of the art in mobile robotics is hence far from achieving autonomous operation over a range of domains. Real world environments change in ways that cannot be specified in advance, while most sensors mounted on mobile robots require a time-consuming manual calibration phase before deployment. In addition, this calibration is sensitive to environmental changes. Furthermore, a robot can process the inputs from its multiple sensors using a set of algorithms, each of which may have a different reliability and computational complexity. Processing all the information would be infeasible in dynamic domains where real-time operation is essential.

The above-mentioned challenges are all the more pronounced in the case of visual input from color cameras. A color camera provides higher bandwidth information than range sensors (laser, sonar, etc.) at a much lower cost. Visual input is however more noisy and sensitive to environmental factors such as illumination, and visual information processing algorithms are typically computationally expensive. Until recently, many mobile robot applications have therefore relied on range sensors [7, 8]. Even the approaches that consider visual input make most high-level decisions based on other sources of input [6, 9], or only use the limited information obtained from intensity images [10]. A rich source of information is hence not fully exploited.

One factor that can be utilized to offset the challenges listed above is the presence of a moderate amount of structure in many mobile robot environments. Examples of such *structure* include known positions and properties (e.g., size, shape) of unique objects in the environment, information which can be manually provided or automatically inferred by the robot. This structure can be exploited to enable autonomous operation on mobile robots. Our work represents a comprehensive effort towards such learning, adaptation, and processing management using the input from color cameras as the primary source of information. The work on autonomous learning and adaptation focuses on color as the feature of interest and illumination as the environmental factor that changes over time. The work on processing management considers several sources of information that are based on visual input. Specifically, this paper summarizes the following contributions.

(i) A probabilistic *bootstrap* learning framework that enables a robot to plan its actions in order to learn models of color distributions and illumination conditions in its environment [11]. The robot uses these learned models to detect and adapt to illumination changes [12].

(ii) A probabilistic sequential decision-making framework which enables a robot to autonomously tailor its visual information processing to the task at hand [13].

These algorithms are tested on specific robot platforms and have the potential of generalizing to other applications.

The remainder of this paper is organized as follows. Section 2.1 summarizes a typical robot vision system, while Section 2.2 describes the test platforms used. Section 3.1 describes the related work in the areas of color segmentation, color learning, and illumination invariance, followed by an overview of AI planning methods as applied to robot vision tasks (Section 3.2). Next, Section 4.1 describes our proposed approach for autonomous illumination-invariant color learning, while Section 4.2 presents the approach for visual processing management based on probabilistic sequential decision processes. Finally, Section 5 summarizes the conclusions and directions for further research.

2. Baseline Vision System and Test Platforms

In this section, we present an overview of a typical robot vision system, followed by a description of the test platforms used to evaluate the proposed algorithms.

2.1. Baseline Robot Vision System. Figure 1 shows a flowchart of the typical robot vision system that uses color information. Color segmentation is typically the first step, where the goal is to cluster image regions into *similar* groups and/or to create a mapping from pixel values to discrete color labels, that is, to create a *color map*

$$\Pi_E : \{m_1, m_2, m_3\} \longmapsto l|_{l\in[0,N-1]}, \tag{1}$$

where m_1, m_2, m_3 are the values along the color channels (e.g., R, G, B) that can take values in [min-max] (0-255 for RGB), the subscript E represents the dependence on illumination, and l refers to the numerical indices of the color labels (e.g., blue = 1, orange = 2). This mapping is typically generated by extensive manual labeling of image regions.

The (color) segmented image regions are used to find "objects" and other desired structures using heuristics and constraints based on the known properties (size, shape, color, etc.) of the target objects. The detected objects and their locations in the image can be used along with other inputs (e.g., depth map from a stereo camera) for creating

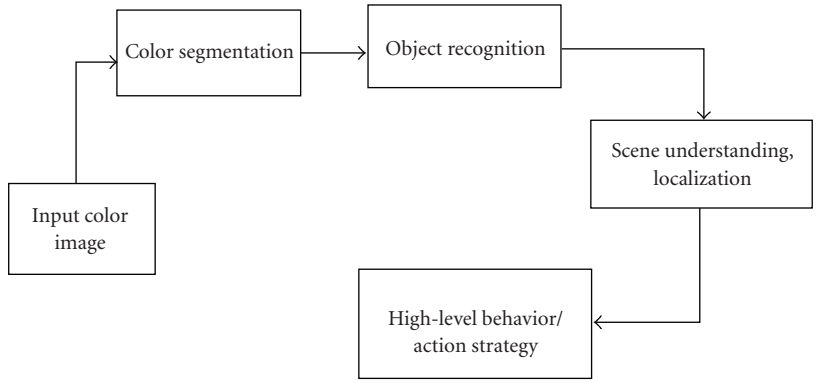


Figure 1: Typical vision-based operation flowchart.

a 3D model of the scene. On mobile robots, the relative distances and bearings of the detected objects can be used in a localization module that computes the position and orientation (i.e., pose) of the robot in the global frame of reference. The 3D model of the scene and/or the pose information is used by the robot to determine the high-level behavior suitable to achieve the desired goal (e.g., navigate to a location to retrieve an object). See [14] for an instance of such a robot vision system.

In order to operate robustly in dynamic environments, a robot has to deal with unexpected changes autonomously and efficiently. For instance, Figure 2 shows that the color map trained under one illumination results in poor segmentation under a different illumination. Robots however frequently have to operate in domains with changing illumination. Section 4.1 summarizes our approach for autonomous learning of color models and adaptation to illumination changes.

A mobile robot equipped with multiple sensors can process its sensory inputs using many algorithms that may have varying levels of uncertainty and computational complexity. Hence, a key requirement for autonomous operation is the ability to tailor the sensory processing to the task at hand. Section 4.2 describes an instance of such processing management of visual input using probabilistic sequential decision processes. The overall goal of our research is to enable autonomous mobile robot operation in a wide range of applications.

2.2. Test Platforms. In this work, we use two different test platforms to evaluate the algorithms: a four-legged robot in the robot soccer scenario, and a mobile robot *playmate* in a human-robot interaction scenario.

2.2.1. Robot Soccer. The color learning and illumination invariance methods were evaluated on the SONY *ERS-7* Aibo, a four-legged robot whose primary sensor is a CMOS camera at the tip of its nose, with a limited field of view (56.9° horz., 45.2° vert.). The images are captured at 30 Hz with a resolution of 208 × 160 pixels. The robot has 20 degrees of freedom, three in each leg, three in its head, and the rest in its tail, mouth, and ears. It has wireless LAN for communication with an off-board PC or other robots. However, all processing for vision, localization, motion and strategy is performed on-board using a 576 MHz processor.

One major application domain for the Aibos is the RoboCup Legged League [15], an international research initiative where teams of four robots play a competitive game of soccer on a (4 m × 6 m) indoor field. As with other robots equipped with cameras, the vision system on the Aibo has an initial color calibration phase. The calibration includes extensive manual labeling of appropriate regions in the images captured by the robot's camera, in order to obtain the color map (as described in Section 2.1). This manual labeling is a major challenge to autonomous operation, and the color map is sensitive to illumination changes—see Figure 2. Figures 3(a) and 3(b) show images of the Aibo and the soccer environment.

2.2.2. Robot Playmate. The visual processing management experiments were conducted on a mobile robot *playmate* that collaborates with a human to jointly manipulate and converse about objects on a tabletop [16] as shown in Figure 4(b). The robot is equipped with a stereo camera (640×480 images at 30 Hz), manipulator arm, on-board processors, and other sensors. The domain, though seemingly simple, represents the state of the art in cognitive robotics [17]. The processing cycle in this domain is different from the flowchart described in Section 2.1—modules operating in parallel process the vision and speech inputs, and create goals that are achieved by other modules such as manipulation. Visual processing in this domain, however, is characterized by the same features and requirements as the robot soccer scenario.

Typical visual processing tasks in this domain require the ability to find the color, shape, identity, or category of objects in the scene to support dialogues about their properties; to see where to grasp an object; to plan an obstacle free path to do so and then move it to a new location; to understand spatial relations between objects; to recognize actions performed by humans. Each of these vision tasks is a hard problem in itself, but we are faced with the formidable challenge of building a vision system capable of performing all of them. Consider the scene in Figure 4(a)

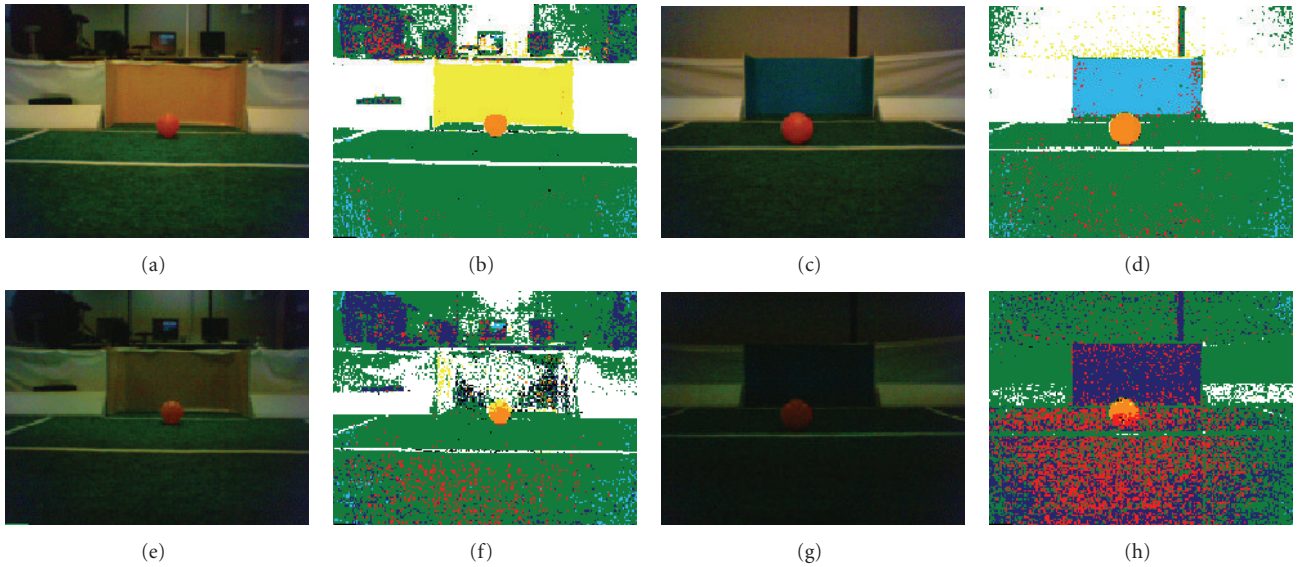


Figure 2: Illumination sensitivity: (a)–(d) color map trained under an illumination, (e)–(h) ceases to work when illumination changes.

with rectangular regions of interest (ROI) extracted from the background. The robot has to use a subset of the available visual routines to execute commands or answer queries: "is there a blue triangle in the scene?", "move the mug to the right of the circle". However, it is neither feasible nor desirable to run all the routines on each image, since the robot has to respond to dynamic changes.

3. Related Work

This paper focuses on learning, adaptation, and processing management using visual input. The topics of interest such as color learning, illumination invariance, and planning of visual processing continue to be extensively researched in the fields of computer vision, robot vision, and planning. This section therefore reviews a set of representative methods for these topics and analyzes the approaches in terms of their applicability to robots operating in dynamic environments.

3.1. Color Segmentation, Learning, and Color Constancy. Color segmentation is a well-researched field in computer vision with several good algorithms [18–20]. The mean-shift algorithm is a nonparametric technique for the analysis of complex multimodal feature spaces and the detection of arbitrarily shaped clusters [18]. The feature space is modeled as an empirical probability density function (pdf). Dense regions in the feature space correspond to the modes of the unknown pdf. Once the modes are found, the clusters can be separated based on the local structure of the feature space. Mean-shift provides good performance on tasks such as segmentation and tracking, but its quadratic complexity makes it expensive to perform on robots with computational constraints.

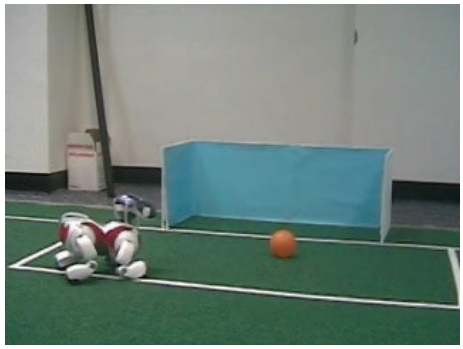
Active contours are another set of popular methods for image segmentation [20–22]. The method defines initial contours and then deforms them towards object boundaries. Manjunath et al. describe a region-based method [20] that segments images into multiple regions and integrates an edge-flow vector field-based edge function for segmenting precise boundaries. The method allows the user to specify the similarity measure based on features such as color or texture. The algorithm is not sensitive to the initial estimates and provides good segmentation results on a variety of images, but the iterative optimization is expensive to perform on robots.

Image segmentation can also be posed as a graph-partitioning problem, where each node represents a pixel in the image, and the edges connect certain pairs of neighboring pixels [19, 23]. Typically, graph-based segmentation methods find minimum cuts in the graph, where a *cut* measures the degree of dissimilarity between point sets by computing the weights of the graph edges that have to be removed to separate the two sets. Shi and Malik proposed the popular *normalized cut* algorithm, a robust global criterion that simultaneously maximizes the similarity within a cluster and the dissimilarity between clusters [19]. Normalized cuts have been used for computer vision tasks such as motion tracking [24] and 3D view reconstruction [25], but the approach is computationally expensive for robot platforms.

In the RoboCup domain, the typical approach is to create mappings (1) from the YCbCr values to the color labels [26]. Other methods include the use of decision trees [27] and axis-parallel rectangles in the color space [28]. These approaches involve the hand-labeling of images over a period of an hour or more before the color map can be generated (Section 2.1). Attempts to learn colors or make them independent to illumination changes involve the knowledge of the spectral reflectances of the objects under consideration and/or require additional transformations that are computationally expensive to perform on robots [29, 30].

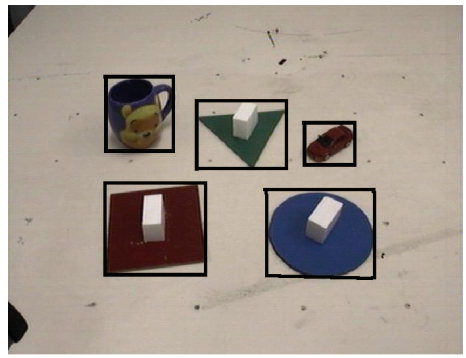
An important consideration in color learning and segmentation is the choice of color space. However, there is

(a) Robot soccer field setup



(b) Robot walking to the ball

Figure 3: Image of the robot soccer domain: the Aibo plays soccer on the robot soccer field with goals and markers.



(a) Tabletop scenario example

(b) Robot *playmate* setup

Figure 4: Image of the human-robot collaboration domain: a robot and a human jointly converse about and manipulate objects on a tabletop. Regions of interest (ROI) are bounded by rectangular boxes.

a lot of controversy on the "best" color space for different applications. In order to address this challenge, Gevers and Smeulders evaluated several color spaces to determine their suitability for recognizing multicolored objects invariant to significant changes in viewpoint, object geometry, and illumination [30]. They presented a detailed theoretical and experimental analysis of the several models. This research hence provides a good reference on the choice of color spaces.

Attempts to automatically learn the color map in the legged league have rarely been successful. Cameron and Barnes [31] present an approach that detects edges in the image and constructs closed figures to find image regions corresponding to known objects. The color information from these regions was used to build the color classifiers. Illumination changes are tracked by associating the current classifiers with the previous ones. This approach is time consuming even with the use of off-board processing. Jungel presents another approach where the color map is learned using three layers of color maps with increasing precision levels [32], with colors in each level being represented as cuboids. The colors are defined relative to a reference color (field *green* in the soccer domain) that is tracked with minor illumination changes, and all other color distributions are displaced in the color space by the same amount. However, different colors do not actually shift by the same amount with illumination changes, and hence the color map is reported to be not as accurate as the hand-labeled one. Unlike these prior approaches, our algorithm exploits domain knowledge to model color distributions and learn a color map in ≈6 minutes of robot time, resulting in performance comparable to the color map obtained after hours of human effort.

While learning can automate the color map generation, the learned map is still sensitive to illumination changes. The response obtained at a sensor can be defined as [33]

$$m_j^p = \int \left(E(\lambda)S^{\mathbf{p}}(\lambda)R_j(\lambda)\right)d\lambda, \tag{2}$$

where $E(\lambda)$ is the spectral power distribution of the illuminant, $S^{\mathbf{x}}(\lambda)$ is the surface reflectance at a scene point $\mathbf{x}$, while $R_j(\lambda)$ is the spectral response (relative) of the imaging device's jth sensor. The response of the jth sensor of the camera at pixel p, m_j^p, is the integral of the product of these three terms over the range of wavelengths. Changing the surface reflectance or the spectral power distribution of the illuminant can change the sensor response. Color constancy or illumination invariance is the ability to assign the same symbolic labels to color distributions despite illumination changes. Decades in computer vision have resulted in several methods for color constancy, most of which focus on static images and have high computational complexity.

The Retinex theory [34] is based on the assumption that white reflection induces maximal *rgb* camera responses, and uses the maximum r, g, and b responses as an estimate of

the illuminant. It was later modified to be based on global or local image color averages—the "Gray World" algorithm [35] is based on the same principle. However, the local or global image averages correlate poorly with the actual illuminant [36].

The *gamut mapping* algorithm [37] proposed by Forsyth is based on the fact that surfaces can reflect no more light than what is cast on them. The illuminant color is hence constrained by the colors observed in the image, and can be estimated using image measurements alone. The algorithm selected the most likely mapping from a set of mappings that transformed the sensor values under an unknown illuminant to the gamut of colors observed under a canonical illuminant. Finlayson proposed the *median selection* method that included a constraint on the possible color of the illuminant into the gamut mapping algorithm [38]. The more recent correlation framework [33] measures the likelihood that each of a possible set of illuminants is the scene illuminant. However, these approaches require prior knowledge of the illuminations which is not feasible in robot domains.

Brainard and Freeman use a Bayesian decision framework, which combines statistics such as gray world, subspace, and physical realizability constraints [39]. They generate a priori distributions to describe the probability of existence of certain illuminants and surfaces. A maximum local mass (MLM) estimator integrates local probabilities and uses Bayes' rule to compute the posterior distributions for surfaces and illuminants, for a given set of photosensor responses. However, significant prior knowledge of illuminants and other statistics is required, and the approach is computationally expensive. Tsin et al. present a Bayesian *maximum a posteriori* (MAP) approach for outdoor object recognition with a static surveillance camera [40]. Static overhead high-definition images collected over several days are used to learn models of reflectance and the light spectrum. A linear iterative scheme converges to the classification result on the test images. A mobile robot system, however, has to be robust to camera motions and dynamic changes.

On robots, the color constancy problem has often been avoided by using nonvisual sensors such as range finders [8]. Even when visual input is considered, the focus is on recognizing well-separated colors [3]. There has been little work on color constancy in the presence of shadows and artifacts due to rapid camera motion. Further, with few exceptions (e.g., [41, 42]), most methods do not function in real time.

Schulz and Fox estimate colors using a hierarchical Bayesian model with *Gaussian* priors and a joint posterior on position and environmental illumination [43]. Even when tested under two distinct illuminations and a small set of colors, the approach requires prior knowledge of color distributions and illuminations, in addition to being computationally expensive. Lenser and Veloso present a tree-based state description technique [41] for detecting changes in lighting on Aibo robots. A time-series of average screen illuminance is used to distinguish between illumination conditions. We however believe that the color space distributions could function as a better discriminating feature. Anzani et al. describe an attempt at illumination invariance in the RoboCup middle-size league [42], where a *Mixture of Gaussians* (MoG) is used to generate multimodal distributions for the various colors. The EM algorithm [44] is used with online adaptation of the number of mixture components, in order to adapt to minor illumination changes. However, the labeling of color classes and association with mixture components is done by human supervision, and the algorithm has been tested only over a few illuminations in the lab. In the recent DARPA grand challenges, Thrun [6] modeled colors as MoG and attempt to add additional Gaussians and modify the parameters of the existing Gaussians in response to the changes in illuminations. However, they were interested only is distinguishing safe regions on the ground from the unsafe regions and did not have to model overlapping color classes separately.

Section 4.1 summarizes our approach for modeling illuminations and overlapping colors without prior knowledge of color distributions. The learned models are used to detect and adapt to a range of illuminations. See [45] for a recent survey on color learning and illumination invariance.

3.2. Visual Processing Management. AI planning and cognitive planning architectures are well-researched fields [46–49]. The focus here is on a specific subcategory of the planning problem: the joint planning of the sensing (where to look) and information processing (what to look for) actions to achieve a desired goal.

Classical planning methods use deterministic models and require prior knowledge of state, action outcomes, and all contingencies. Many modern planning methods extend the machinery of classical planning in order to model the nondeterminism inherent in perception. Draper et al. [50] proposed C-BURIDAN, a planning scheme that incorporates a probabilistic model of the noisy sensors and effectors, while still retaining a symbolic STRIPS-like representation of action effects [51]. The plan-assessment phase treats actions as probabilistic state transitions, while the plan-refinement phase links the symbolic action effects to the symbolic subgoals of the desired goal state. Their formulation is similar to our POMDP approach as they reason about the best action to perform based on prior belief about the world and the observations obtained from action execution. However, their approach requires the action preconditions and effects to be manually specified, does not incorporate a notion of action costs, and requires a manual ordering of actions to accumulate belief from repeated execution of the same action.

In contrast to the C-BURIDAN system, Petrick and Bacchus's PKS planner [52] describes actions in a first-order language, in terms of their effect on the agent's knowledge rather than their effect on the world. The model is hence nondeterministic in the sense that the true state of the world may be determined uniquely by the actions performed, but the agent's knowledge of that state is not. For example, dropping a fragile item will break it, but if the agent does not know that the item is fragile, it must use an observational action to determine its status. PKS captures

the initial state uncertainty and constructs conditional plans based on the agent's knowledge. More recently, Brenner and Nebel [53] proposed the Continual Planning (CP) approach, based on the FF planner [54], which interleaves planning, plan execution, and monitoring. Unlike classical planning, an agent in CP postpones reasoning about unknowable or uncertain states until more information is available. Actions are allowed to assert that the preconditions for the action will be met when the agent reaches that point in the execution of the plan. If these preconditions are not met during execution, or are met earlier, replanning is triggered. CP is therefore similar to PKS in representation but works by replanning rather than constructing conditional plans. There is however *no* representation of the uncertainty in the observations and actions. In applications where observations are noisy, the optimal behavior may be to increase the confidence in the image interpretation by running the operators more than once on several images of a scene, and accumulating the evidence. This cannot be readily represented in the approaches described above.

There is a significant body of work in the image processing community on planning of visual operations [55–57]. Such approaches typically use a classical planner that takes a user-specified high-level goal and constructs a pipeline of image processing operations. The planners use deterministic models, actions represented as STRIPS-like operators with prespecified preconditions and effects, and domainspecific rules for evaluating the output of each operator. Unsatisfactory results are handled by re-planning the operator sequence or modifying the operators' parameters [57, 58].

In the field of computer vision, probabilistic sequential decision processes,, that is, Markov Decision Processes (MDPs) and Partially Observable MDPs (POMDPs), have been used for image interpretation. Darrell [59] used memory-based reinforcement learning and POMDPs to learn to foveate salient body parts in an active gesture recognition system. The action set consists of foveation actions and a special recognition action. During the learning phase, execution of the recognition action is followed by manual feedback on the target object's presence in the scene, so that each action sequence can be assigned a reward. Reinforcement learning is used to learn what foveation actions to execute, and when to execute the terminal recognition action. More recently, Li et al. [60] posed image interpretation as an MDP, using human-annotated images in an offline process to determine the reward structure by applying all possible sequences of image operators to the labeled images. Dynamic programming methods are used to determine the value function for the explored parts of the state space, which is then extrapolated to the entire state space using an ensemble learning technique. During online execution, each step consists of feature extraction and the choice of an action that maximizes the learned value functions. Approaches that require manual feedback in an initial training phase are too time-consuming to use on a robot interacting with a human. It would instead be desirable to autonomously generate the models and policies.

Sequential decision processes have also been used for planning a sequence of gaze locations (image ROIs) that are analyzed to identify the desired target. In the recent work, Vogel and de Freitas [61] have posed gaze sequence selection as a finite-horizon sequential decision process that elegantly combines bottom-up saliency, top-down target knowledge, and spatial target context. The gaze planning strategy was tested on image databases to determine the location of computer monitors. Though this approach requires a prior distribution over object locations that is difficult to compute for multiple objects in practical scenes, it clearly demonstrates the benefits of visual processing management.

There has also been considerable related work in the field of active sensing, where the goal is to decide on sensor placement and sensor information processing based on its relevance to the task at hand [62, 63]. Kreucher et al. [62] presented an active sensing approach for scheduling sensors in order to learn the number and states of a group of moving targets in a surveillance region. The joint multitarget probability density is estimated using a particle filter. A sensing action is chosen (at each time step) based on the Renyi-divergence measure, and then the probability density of the number and states of the targets is updated. However, estimating the joint probability density requires considerable prior information that is difficult to obtain in robot environments. Our visual processing management approach (Section 4.2) is also focused on computing a sequence of operations for a specific task. However, our approach uses automatic belief propagation to enable the robot to respond to dynamic changes.

Recently, there has been considerable interest in the use of submodular functions for sensor placements in spatial phenomena modeled as Gaussian processes [64, 65]. For objective functions that can be represented as submodular functions, the greedy policy provides performance that is at least 63% of optimal performance [65]. However, our approach is significantly similar to methods that aim to maximize the information gain, and such approaches cannot be represented using submodular functions [64].

Since POMDP solutions of practical-sized problems are typically intractable, several researchers have focused on imposing structure in POMDP formulations in order to make it more tractable. Pineau et al. [4, 66] propose a hierarchical POMDP approach for high-level behavior control on a nursing assistant robot, similar to the MAXQ decomposition for MDPs [67]. They impose an action hierarchy, with the top level action being a collection of simpler actions that are represented by smaller POMDPs. A hierarchical planning algorithm operates in a bottom-up manner, finding complete solutions, that is, policies for the smaller POMDPs. The execution proceeds in a top-down manner: invoking the policy at the top-level recursively traverses the hierarchy invoking a sequence of local policies until a primitive action is reached. Model parameters at all levels are defined over the same space of states, actions, and observations, but the relevant space is abstracted for each POMDP using a dynamic belief network. Hansen and Zhou [68] propose a similar Task Hierarchy (TH) for planning with POMDPs, where the policies are defined as finite-state

controllers (FSCs) and the dynamic programming policy of a subproblem is treated as an abstract action in the next higher level POMDP. The difference is that each POMDP in the hierarchy is an indefinite-horizon POMDP in order to allow FSC termination without recognition of the underlying terminal state. Similar systems have also been proposed for autonomous robot navigation [69–71]. In the actual application, however, a significant amount of data for the hierarchy and model creation has to be hand-coded.

There has been considerable work on exploring representations for hierarchical POMDPs that allow for tractable performance in practical applications. Theocharous et al. [72] represented hierarchical POMDPs as dynamic Bayesian networks (DBNs), for the specific task of using multiresolution spatial maps for indoor robot navigation. They have shown that the DBN representation can train faster (and with fewer samples) than the hierarchical POMDP or the flat POMDP. More recent work by Toussaint et al. [73] aims to learn the hierarchical representation of a POMDP based on maximum likelihood estimation, using dynamic Bayesian networks and parameter estimation based on Expectation-Maximization. However, these approaches require considerable manual supervision, or are computationally expensive to use on robots.

Similar to existing approaches, the summarized hierarchical POMDP approach defines the higher level model parameters as functions of the lower level policies. However, automatic belief propagation is achieved through a *functional* decomposition that considers images of regions in space at the higher level, and the corresponding image ROIs at the lower level. Incorporating additional operators or ROIs is hence easier.

4. Proposed Approaches

This section summarizes our work on autonomous learning, adaptation, and visual processing management. Section 4.1 describes the algorithm that enables a robot to use autonomously learned color and illumination models to detect and adapt to illumination changes. Next, Section 4.2 presents the algorithm that enables a robot to autonomously tailor its visual processing to the task at hand.

4.1. Planned Illumination-Invariant Color Learning. As described in Section 2.1, the manual calibration of the color map is time-consuming and sensitive to illumination changes. However, as with many other application domains, the robot on the soccer field knows (or can infer) a significant amount of the structure in its environment—it knows the positions and color labels of the objects of interest (e.g., goals, markers, etc.). Here, we describe an approach that enables the robot to exploit this known structure to do the following.

(i) Learn models of color distributions and illuminations, which can be refined incrementally.

(ii) Use the learned models to detect and adapt to a range of illumination changes.

The overall algorithm is summarized in Algorithm 1—specific line numbers are referenced in the text below. The robot initially has no prior information of color distributions or illumination. It has an algorithm that exploits the known structure of the environment to plan a motion sequence for learning the color map—see Algorithm 2.

The first question to address is *what to learn?* That is, we need to decide on the appropriate models for color distributions and illuminations. We use a *disjunctive* representation that models the a priori probability density function (pdf) for each color (l) *either as a 3D Gaussian or as a 3D Histogram*.

$$p(\mathbf{m} \mid l) \sim N\left(\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l\right), \quad \text{or} \quad \equiv \frac{hist_l(b_1, b_2, b_3)}{\sum hist_l}, \tag{3}$$

where (b_1, b_2, b_3) are the histogram bin indices corresponding to the color channel values $\mathbf{m} = (m_1, m_2, m_3)$. The histogram is normalized to obtain a pdf. Assuming all colors are equally likely, that is, $P(l) = 1/N$, for all $l \in [0, N-1]$, each color's a posteriori pdf is proportional to the a priori pdf. The color space is discretized and each color map cell is assigned the label of the *most likely* pdf. In a given situation, one of the two models may be more suitable for a color's pdf, and the choice is made autonomously using the bootstrap test [74]. Though other color models are feasible, the disjunctive model provides a balance between accuracy and computation.

Each illumination is represented by a color map and autonomously-collected image statistics. Based on the hypothesis that images from the same illumination have measurably similar distributions of pixels in color space, images captured by the robot are transformed into the normalized RGB space (r, g, b). Histograms in (r, g), with 64 bins in each dimension, are normalized to obtain pdfs ($rgHist_{E_{illum}}$) that form the first statistic. The distance between every pair of pdfs is computed and the distribution of distances ($D_{E_{illum}}$), modeled as a Gaussian, constitutes the second statistic. The Jensen-Shannon (JS) measure is used for computing the distance between the distributions

$$\begin{aligned} \mathrm{JS}(\mathbf{a}, \mathbf{b}) &= \frac{\mathrm{KL}(\mathbf{a}, \mathbf{m}) + \mathrm{KL}(\mathbf{b}, \mathbf{m})}{2}, \\ \mathrm{KL}(\mathbf{a}, \mathbf{b}) &= \sum_i \sum_j \left(\mathbf{a}_{i,j} \cdot \ln \frac{\mathbf{a}_{i,j}}{\mathbf{b}_{i,j}}\right), \quad \mathbf{m} = \frac{\mathbf{a} + \mathbf{b}}{2}. \end{aligned} \tag{4}$$

The JS distance is a function of the log of the pdfs $(\mathbf{a}, \mathbf{b})$ and is hence robust to peaks in the distributions, that is, large image regions of a single color.

A mobile robot typically has to operate in environments where the illumination changes unpredictably. Given the models for color distributions and illumination, the next question to address is: *how to detect and adapt?* That is, how to detect illumination changes and adapt to them. Major illumination changes, for instance, when the lamps are suddenly switched on (or off), cause large shifts in color distributions. The current color map is no longer valid and the robot is soon lost. Minor (or slow) illumination changes, for instance, the variation in natural light during the day,

Require: For each known illumination E_i, $i \in [0, M-1]$, color map Π_{E_i}, (r,g) distributions $rgHist_{E_i}$, and distribution of JS-distances D_{E_i}.
Require: Algorithm to plan motion and learn colors autonomously (Algorithm 2).
Require: Positions, shapes and color labels of the objects of interest in the robot's environment. Initial robot pose.
(1) Initialize: $M = 0$, $illum = 0$, $testTime = 0$ (no prior illumination knowledge).
(2) Plan motion and learn $\Pi_{E_{illum}}$.
(3) Generate $rgHist_{E_{illum}}$, $N(r,g)$ space distributions, and distribution of JS-distances, $D_{E_{illum}}$, using images captured at random during color learning.
(4) Save image statistics, $M = M + 1$.
(5) **while** true **do**
(6) Get new image. Segment image and detect objects.
(7) **if** minorChange $(Color)$ **then**
(8) minorUpdate $(Color)$. Get $\Pi_{\hat{E}}$ from current color distributions.
(9) Revise current illumination representation to get $rgHist_{\hat{E}}$ and $D_{\hat{E}}$, to be used for subsequent operations.
(10) **end if**
(11) **if** $currentTime - testTime \geq time_{th}$ **then**
(12) $rg_{test} = (r,g)$ distribution of current image.
(13) **for** $i = 0$ to $M - 1$ **do**
(14) $dAvg[i] = (1/N) \sum_j JSDist(rg_{test}, rgHist_{E_i}[j])$
(15) **end for**
(16) **if** Exists $(\hat{E})$ **then**
(17) $dAvg_{\hat{E}} = (1/N) \sum_j JSDist(rg_{test}, rgHist_{\hat{E}}[j])$
(18) **end if**
(19) **if** Exists $(\hat{E})$ and withinRange $(dAvg_{\hat{E}}, D_{\hat{E}})$ **then**
(20) Continue with $\Pi_{\hat{E}}$.
(21) **else if** withinRange $(dAvg[illum], D_{E_{illum}})$ **then**
(22) Continue with $\Pi_{E_{illum}}$.
(23) **else if** withinRange$(dAvg[i], D_{E_i})$, $i \neq illum$ **then**
(24) Use Π_{E_i}, $illum = i$.
(25) **else**
(26) New illumination, $illum = M$, $M = M + 1$.
(27) Learn $\Pi_{E_{illum}}$ autonomously.
(28) Learn $rgHist_{E_{illum}}$ for new illumination.
(29) Use $\Pi_{E_{illum}}$ for subsequent operations.
(30) **end if**
(31) $testTime = currentTime$.
(32) **end if**
(33) **end while**

Algorithm 1: Illumination adaptation algorithm.

cause the robot's segmentation to slowly deteriorate as the color distributions shift.

For each object detected from the color segmented image regions, the robot computes

$$\frac{numPixels_l}{totalPixels} \leq changeThreshold, \tag{5}$$

where $numPixels_l$ represents the pixels of the color label (l) of the detected object, and $totalPixels$ is the total number of pixels within the object's bounding rectangle. If the value of this ratio falls below a threshold consistently (for $\geq 60\%$ of N consecutive frames) it indicates a minor illumination change, denoted by $Detect_{minor}$ ($minorChange()$—line 7). The new illumination is denoted by $\hat{E}$. The pixels within the corresponding image region are used to build a new model (i.e., a histogram or a Gaussian) for the color distribution, which is merged with the current model for that color ($minorUpdate()$—line 8). For Gaussians, we use

Require: Ability to learn color models.
Require: Positions, shapes and color labels of the objects of interest in the robot's environment. Initial robot pose.
(1) Move between randomly selected target poses.
(2) CollectMEMData() – collect data for motion error model.
(3) CollectColLearnStats() – collect color learning statistics.
(4) NNetTrain() – Train the Neural network for the MEM, (8).
(5) UpdateFM() – Generate the statistical feasibility model, (9).
(6) GenCandidateSeq() – Generate candidate sequences, (10).
(7) EvalCandidateSeq() – Evaluate candidate sequences.
(8) SelectMotionSeq() – Select final motion sequence.
(9) Execute motion sequence and learn colors – Algorithm described in [76].

Algorithm 2: Motion sequence generation.

the measurement update of a Kalman Filter [75]

$$\begin{aligned} Gain\ \mathbf{K_l} &= \boldsymbol{\Sigma}_{l_{old}}\left(\boldsymbol{\Sigma}_{l_{old}} + \boldsymbol{\Sigma}_{l_{new}}\right)^{-1}, \\ \boldsymbol{\mu}_{l_{up}} &= \boldsymbol{\mu}_{l_{old}} + \mathbf{K_l}\left(\boldsymbol{\mu}_{l_{new}} - \boldsymbol{\mu}_{l_{old}}\right), \\ \boldsymbol{\Sigma}_{l_{up}} &= (\mathbf{I} - \mathbf{K_l})\boldsymbol{\Sigma}_{l_{old}}, \end{aligned} \tag{6}$$

where the subscripts *old*, *new*, and *up* represent the current, new, and updated model, respectively, for color l. For histograms, a weighted average is computed

$$\begin{aligned} p_{l_{old}} &= \frac{hist_{l_{old}}}{\sum hist_{l_{old}}}, \qquad p_{l_{new}} = \frac{hist_{l_{new}}}{\sum hist_{l_{new}}}, \\ p_{l_{avg}} &= w_{old}\, p_{l_{old}} + w_{new}\, p_{l_{new}}, \qquad w_{old} + w_{new} = 1, \\ p_{l_{up}} &= \frac{p_{l_{avg}}}{\sum p_{l_{avg}}}, \qquad hist_{l_{up}} = p_{l_{up}} \sum \left(hist_{l_{old}} + hist_{l_{new}}\right) \end{aligned} \tag{7}$$

for merging the normalized histograms with the existing normalized histograms to obtain the updated histograms. The weights are based on the number of samples in the corresponding histograms. The color map and the current illumination model are modified and used in subsequent operations (line 9). This adaptation scheme is called $Adapt_{minor}$.

In order to detect sudden illumination changes, the robot periodically ($time_{th} = 0.5$ seconds) generates a test image histogram in the (r, g) space (line 12). The average distance ($dAvg$) is computed between this test histogram and the set of histograms corresponding to each illumination for which a representation has been learned (lines 13–15). Illumination representations created while tracking minor illumination changes are included in this computation (lines 16–18 in Algorithm 1).

If $dAvg$ lies within the threshold range (95%) of the distance distribution corresponding to the current illumination (*withinRange()*—lines 19, 21), the robot continues to use the current color map. If $dAvg$ lies outside the range of the distance distribution of the current illumination, but within the range of the distance distribution corresponding to an illumination for which the robot has learned a model, the robot transitions to using the corresponding color and illumination models. However, if $dAvg$ lies outside the range of all known illuminations, the robot models a new illumination ($Detect_{major}$) and learns models for color distributions (lines 25–30). This adaptation scheme ($Adapt_{major}$) cannot be used with a reduced threshold to handle minor illumination changes, because it could result in a large number of color maps for changes in a few distributions. Both $Adapt_{minor}$ and $Adapt_{major}$ are hence necessary.

Given the algorithm to adapt to illumination changes, the final question to address is: *how to learn?* That is, how to learn the color map and illumination model. As summarized in Algorithm 2 we answer this question by finding a sequence of poses (x, y, θ) the robot can move through, learning one color at each pose. The goal is to simultaneously maximize color learning opportunities while minimizing localization errors—the robot may obtain more training samples by moving a larger distance, but this motion may cause larger localization errors. This goal is achieved by discretizing the robot poses into cells, and using three components: a motion error model (MEM), a statistical feasibility model (SFM), and a search routine.

The MEM predicts the error in the robot pose in response to a motion command. The inputs are the difference between the starting pose (x_i, y_i, θ_i) and target pose (x_f, y_f, θ_f), and the list of colors the robot has learned. The output is the pose error that would be incurred during this motion. The MEM is represented as a back-propagation neural network [77] with $N + 3$ inputs, three outputs and one hidden layer of 15 nodes

$$\left\{\Delta_x, \Delta_y, \Delta_\theta, c_1, c_2, \ldots, c_N\right\} \longmapsto \left\{err_x, err_y, err_\theta\right\}, \tag{8}$$

where $\{\Delta_x, \Delta_y, \Delta_\theta\}$ represent the difference in pose, and $\{c_0, c_1, \ldots, c_{N-1}\}$ are binary variables representing the target colors. If all the colors are known, all the markers can be recognized—with only some colors known, some markers are not recognizable, and the robot's localization suffers.

For each robot pose, the SFM provides the probability of learning each of the desired colors given that a subset of

the colors have been learned. A feasibility check based on the robot's joint angles and camera field of view eliminates a lot of cells—the robot can learn colors only when its camera is pointing towards a known object. The feasibility check is performed once for each object configuration. Each cell of the SFM stores a probability measure

$$\mathrm{SFM}(d, e, f, v_i) = p, \quad \forall \{d, e, f\} \in [0, K-1], \tag{9}$$

where d, e, f are cell indices of the K discrete poses and v_i, $i \in [0, M-1]$ represents all possible combinations of colors.

In the training phase, the robot moves between randomly generated target poses and executes two localization routines, one with all colors known (provides ground truth) and another with only a subset of colors known. The difference of the two pose estimates provides the training samples to build the MEM (*CollectMEMData()*, *NNetTrain()* in Algorithm 2). In parallel, the robot attempts to learn colors based on the knowledge of a subset of the colors. The Gaussian-smoothed and normalized cell counts of the successful learning attempts are used to compute the SFM (*CollectColLearnStats()*—line 3, *UpdateFM()*—line 5). The SFM has to be relearned when the object configurations change, but even with just the geometric constraints the robot is able to provide motion sequences leading to successful color learning.

Given the learned models (MEM, SFM), for any given starting pose during testing, the robot generates all candidate motion sequences (*GenCandidateSeq()*—line 6), that is, all possible paths along the discretized pose cells. The depth of the search is equal to the number of colors to be learned—we assume that the robot learns one color at each pose. If the robot is to learn N colors, the motion sequence is

$$\text{path: } \{x_i, y_i, \theta_i, \text{color}_i\} \quad \forall i \in [0, N-1]. \tag{10}$$

This formulation results in a large number of paths ($\approx 10^9$). However, only a small subset of paths ($\approx 10^4$) are evaluated completely. The MEM predicts the pose error if the robot travels from the starting pose to the first pose. The vector sum of the error and the target pose predicts the actual pose. If the desired color can be learned at this pose (high probability in SFM), the move to the next pose is evaluated. If the whole path is evaluated, the net pose error and probability of success are computed (*EvalCandidateSeq()*—line 7). The path that provides a high probability of success and a low pose error is executed (*SelectMotionSeq()*—line 8) by the robot. At each pose, the expected object location is projected on the image to extract pixels that are used to model the color distributions and learn the color map. The data for the illumination model is collected during the learning process.

4.1.1. Experimental Results. We need to evaluate the robot's ability to (a) plan a motion sequence and learn models for color distributions and illumination for different object configurations and (b) use the learned models to detect and adapt to illumination changes.

Table 1: Planning and localization accuracies in challenging configurations. Planned motion sequence always succeeds in learning colors. Localization comparable to hand-labeled color map.

Config	Plan (%)	Localization error		
		X (cm)	Y (cm)	θ (deg)
Learned	100	9.6 ± 3.7	11.1 ± 4.8	9 ± 7.7
Hand-labeled	—	6.9 ± 4.1	9.2 ± 5.3	7.1 ± 5.9

The localization accuracy is used as the performance measure. Given the colors needed for localization (*pink, yellow, blue, white, green*), the depth of the search is limited to three for ease of analysis (and without loss of generality)—the ground colors (*green*, *white*) are learned by scanning in place. The field is discretized into $(6 \times 9 \times 12)$ cells, that is, divisions of 600 mm, 600 mm, and 30° along x, y, amd θ. The back-propagation network is learned using the MATLAB Neural Network toolbox—the initial training of MEM and SFM takes ≈1 hour of autonomous robot effort.

Different object configurations were created by placing six target objects at different positions along the boundary of the field that are known to the robot. The planning capability was evaluated for 7 challenging object configurations, each with 15 different robot starting poses. In addition, the localization errors were measured as the robot moved through a sequence of poses (15 trials of 10 poses)—ground truth was obtained with a tape measure and a protractor. The results are summarized in Table 1. The robot is able to generate a valid plan over *all* the trials, and the localization accuracy is comparable to that obtained from a hand-labeled color map. Figure 5 shows some planning results—the starting position is denoted by number "0" while the direction of the arrows show the orientation. The robot smoothly trades off the ability to learn better models for color distributions based on a larger object, against the associated motion-based localization errors.

In addition to the "best" motion-plan, several of the top sequences lead to successful color learning. If the robot is unable to learn all colors during plan execution, it creates a new plan based on current knowledge. Over a set of 20 images, the average segmentation accuracy of the learned and hand-labeled color map is 94.9 ± 3.9 and 96.7 ± 4.3 respectively (no difference at 95% significance). Ground truth is provided by a human observer. The motion planning is particularly useful where object configurations change less frequently than illumination. The entire color learning process takes ≈6 minutes of robot effort instead of hours of human effort.

Next, the ability to detect and adapt to illumination changes was evaluated—here, using $Adapt_X$ implies the use of $Detect_X$ as well. First, the robot used $Adapt_{major}$ as the illumination was slowly changed (over 20 seconds) between two conditions that would not be detected as being different by $Detect_{major}$. The robot stood in place and panned its head, measuring the distance and angle to an object over the 20 seconds period. Table 2 summarizes the measurement errors averaged over four different objects and three different illuminations with ≈15 trials under each

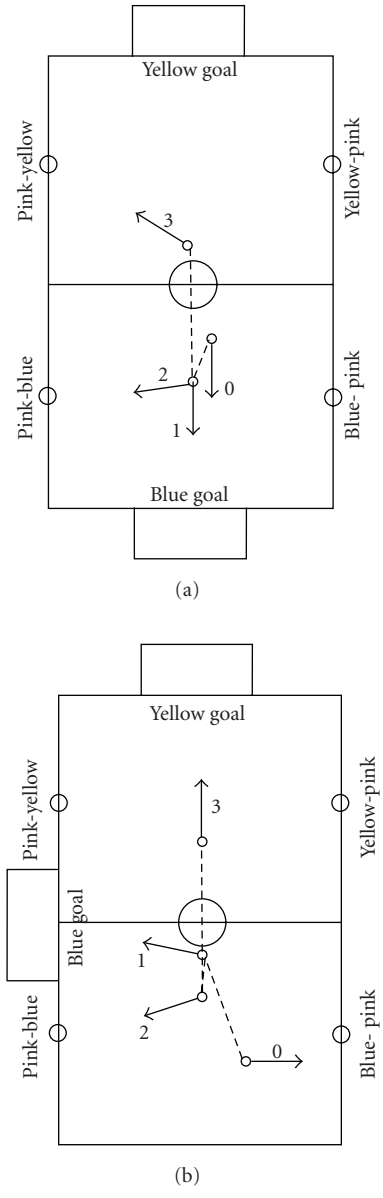


Figure 5: Sample motion plans generated by the algorithm. All plans lead to successful color learning on the robot.

Table 2: Error in distance measurements with and without $Adapt_{minor}$. Adaptation results in much smaller errors.

Illum + Alg	Dist error (mm)	Ang error (deg)
Slow + *NoAdapt*	191.31 ± 105.61	12.37 ± 2.85
Slow + $Adapt_{min}$	25.53 ± 19.14	2.11 ± 0.83

situation. Ground truth values were obtained with a tape measure and protractor. The results show that $Adapt_{minor}$ leads to segmentation accuracy (95.1 ± 4.3) and hence localization errors (≈10 cm, 12 cm, 10°) similar to those under constant illumination. In the absence of $Adapt_{minor}$ the measurement errors are significant.

Table 3: Time taken to find-and-walk-to-object.

Illum + Alg	Time (sec)	Fail
Constant + *NoAdapt*	6.18 ± 0.24	0
Slow + $Adapt_{maj}$	31.73 ± 13.88	9
Slow + $Adapt_{maj,min}$	6.24 ± 0.31	0
Sudden + $Adapt_{min}$	45.11 ± 11.13	13
Sudden + $Adapt_{maj,min}$	9.72 ± 0.51	0
Sudden + *Slow* + $Adapt_{maj,min}$	10.32 ± 0.83	0

Next, in order to show that both $Adapt_{minor}$ and $Adapt_{major}$ are essential, the time taken by the robot to *find-and-walk-to-object* is measured. The robot starts out near the center of the field with the object placed near the penalty box of the opponent's goal. Table 3 summarizes the results averaged over different illuminations, with 15 trials under each illumination. With no change in illumination, the robot can *find-and-walk-to-object* in 6.18 ± 0.24 seconds. When the illumination changes slowly, using just $Adapt_{major}$ does not help—large variance in second row. Including $Adapt_{minor}$ provides good performance (6.24 ± 0.31 seconds). Similarly, when the illumination is changed suddenly, using just $Adapt_{minor}$ does not help—the robot totally fails to perform the task most of the time, resulting in a large number of failures (fourth row, third column). With both $Adapt_{major}$ and $Adapt_{minor}$ the robot can perform the task, the additional time being used to confirm that a change in illumination did occur (9.72 ± 0.51 seconds). In these experiments, major illumination changes result in illuminations for which the robot has already learned models—$Adapt_{major}$ implies a transition to the suitable model. Finally, the illumination is changed significantly, held constant for 3 seconds, and then changed slowly over the next 5 seconds. The robot is able to *find-and-walk-to-object* in 10.32 ± 0.83 seconds *if and only if* $Adapt_{major}$ and $Adapt_{minor}$ are used. The results show that the proposed algorithm enables the operation over a range of illuminations—different intensities (≈400*Lux* to ≈1600*Lux*) and color temperatures (2300 K–4000 K) were evaluated. Additional results (videos and images) are available online: http://www.cs.utexas.edu/ AustinVilla/?p=research/autoplan_illum.

In the recent research, we have used the algorithms described above to learn models of other sensory features, and to robustly fuse information obtained from different sensory inputs on multiple robot platforms [78].

4.2. Visual Processing Management. The human-robot interaction domain described in Section 2.2.2 involves a robot equipped with multiple sensors whose inputs are processed by several algorithms with varying levels of uncertainty. Since the focus on this paper is on the visual input, we present an algorithm that autonomously tailors its visual processing to the task at hand.

Consider the example of an input image from the tabletop scenario (Section 2.2.2) that is preprocessed to yield regions of interest (ROI), that is, rectangular image regions that are different from a previously trained model of the

background—Figure 4(a) shows examples of ROIs. Consider the query: "which objects in the scene are blue?" Without loss of generality and for ease of analysis, assume that the robot has the following set of visual operators at its disposal: a *color* operator that classifies the dominant color of the ROI it is applied on, a *shape* operator that classifies the dominant shape within the ROI, a *sift* operator that uses the SIFT features [79] to detect the presence of one of the previously trained object models. The *color* operator characterizes ROIs based on color-space histograms, while the *shape* operator characterizes the dominant contour within the ROI using invariant moments. The *sift* operator characterizes target objects with local image gradients that are robust to scale, orientation, and viewpoint changes. We use the following terms interchangeably: visual processing actions, visual actions, and visual operators. Given these operators, the task is to plan a sequence of operators that can answer user queries with high confidence.

We pose the visual processing management task as an instance of probabilistic sequential decision making, and specifically as a Partially Observable Markov Decision Process (POMDP) [80]. The POMDP formulation captures the partial observability and non-determinism that characterize visual processing on robots (Section 1). The robot maintains a probability distribution over the true underlying state, called the *belief state*. Each action considers the true underlying state to be composed of the class labels (e.g., *red(R), green(G), blue(B)* for color; *circle(C), triangle(T), square(S)* for shape; *picture, mug, box* for sift), a label to denote the absence of any valid object—*empty* (ϕ), and a label to denote the presence of *multiple* classes (M). The belief state maintenance also requires an observation function that provides a probability distribution over the set of possible outcomes of each action. The set of action outcomes consists of the class labels, the label *empty* (ϕ) which implies that the match probability corresponding to the class labels is very low, and *unknown* (U) which implies that multiple classes are equally likely and the ROI may therefore contain multiple objects. U is an observation, whereas M is part of the underlying state: they are not the same since they are not perfectly correlated.

Since operators only update belief states, we include "special actions" that cause a transition to a terminal state where no further actions are applied, that is, these query-specific actions terminate processing to answer the query. The answer could report or "say" (not to be confused with language-based communication) which underlying state is most likely to be the true state, or it could simply state the presence or absence of the target object. In the description below, without loss of generality and for ease of explanation, we only consider two operators: *color* and *shape*, each of which provides three class labels. The operators are denoted with the subscripts c and s, respectively. The approach generalizes to *sift*, other vision algorithms, and more outcomes. True states and observations are distinguished by the superscripts a and o, respectively. The POMDP for a single ROI in the image can then be defined as the tuple $\langle \mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{Z}, \mathcal{O}, \mathcal{R} \rangle$.

(i) $\mathcal{S} : \mathcal{S}_c \times \mathcal{S}_s \cup term$, the set of states, is a Cartesian product of the variables describing different aspects of the underlying state. It also includes a *terminal* state ($term$). $\mathcal{S}_c : \{\phi_c^a, R_c^a, G_c^a, B_c^a, M_c^a\}$, $\mathcal{S}_s : \{\phi_s^a, C_s^a, T_s^a, S_s^a, M_s^a\}$.

(ii) $\mathcal{A} : \{color, shape, \mathcal{A}_{sp}\}$ is the set of actions. The first two entries are the visual operators. The rest are special actions that represent responses to the queries, describing the presence/absence of the target: $\mathcal{A}_{sp} = \{sFound, sNotFound\}$, or specific query responses: $\mathcal{A}_{sp} = \{sRed, sGreen, sBlue\}$, that is, actions such as "say blue". All the special actions lead to *term*.

(iii) $\mathcal{T} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow [0, 1]$ represents the state transition function. For operators such as *color* and *shape* that do not change the underlying state, it is the identity matrix. For special actions it represents a transition to *term*. For actions that change the state, the transition function can be defined suitably [81].

(iv) $\mathcal{Z} : \{\phi_c^o, R_c^o, G_c^o, B_c^o, U_c^o, \phi_s^o, C_s^o, T_s^o, S_s^o, U_s^o\}$ is the set of observations, a union of the observations for each visual action under consideration, that is, $\mathcal{Z} = \mathcal{Z}_c \cup \mathcal{Z}_s$.

(v) $\mathcal{O} : \mathcal{S} \times \mathcal{A} \times \mathcal{Z} \rightarrow [0, 1]$ is the observation function, a matrix of size $|\mathcal{S}| \times |\mathcal{Z}|$ for each action. For each visual action, it is learned offline by the robot, and it is a uniform distribution for the special actions.

(vi) $\mathcal{R} : \mathcal{S} \times \mathcal{A} \rightarrow \Re$ specifies the reward, that is, the value of taking a particular action in a particular state. It is hence a mapping from the state-action space to real numbers—a negative reward represents a *cost*. In our case

$$\begin{aligned} \forall s \in \mathcal{S}, \quad \mathcal{R}(s, shape) &= -1.25 \cdot f_s(\text{ROI} - \text{size}), \\ \mathcal{R}(s, color) &= -2.5 \cdot f_c(\text{ROI} - \text{size}), \\ \mathcal{R}(s, \text{special actions}) &= \pm 100 \cdot \alpha. \end{aligned} \tag{11}$$

The cost for visual actions depends on the relative computational complexity of the operator and the size of the ROI. For instance, the *color* operator is twice as costly as *shape*, and this is used to assign a cost factor such that the least expensive operator has a relative cost value close to 1. The dependence on ROI size is captured using a polynomial function (14)—the degree and coefficients of the function are computed experimentally as described later in this section. For special actions, a large positive (negative) reward is assigned for making a right (wrong) decision for a given query. For instance, for "what is the color of the ROI?": $\mathcal{R}(R_c^a T_s^a, \text{sRed}) = 100 \cdot \alpha$ and $\mathcal{R}(B_c^a T_s^a, \text{sGreen}) = -100 \cdot \alpha$, while for "is there a red object in the scene?": $\mathcal{R}(R_c^a T_s^a, \text{sFound}) = 100 \cdot \alpha$, and $\mathcal{R}(B_c^a T_s^a, \text{sFound}) = -100 \cdot \alpha$. The variable α trades-off computational costs against reliability. For instance, when α is large the special action is taken after executing a larger number of actions, resulting in higher reliability.

Given the belief state, that is, the probability distribution over the underlying state at time t: b_t, the belief update proceeds as

$$b_{t+1}(s') = \frac{\mathcal{O}(s', a_t, o_{t+1}) \sum_{s \in \mathcal{S}} \mathcal{T}(s, a_t, s') \cdot b_t(s)}{P(o_{t+1} \mid a_t, b_t)}, \quad (12)$$

where $\mathcal{O}(s', a_t, o_{t+1}) = P(o_{t+1} \mid s_{t+1} = s', a_t)$, $b_t(s) = P(s_t = s)$, $P(o_{t+1} \mid a_t, b_t) = \sum_{s' \in \mathcal{S}} \{P(o_{t+1} \mid s', a_t) \cdot \sum_{s \in S} P(s' \mid a_t, s)\ b_t(s)\}$ is the normalizer and $\mathcal{T}(s, a_t, s') = P(s_{t+1} = s' \mid a_t, s_t = s)$. The planning task for a single ROI involves solving this POMDP to find a policy of the form

$$\pi^* : (b) \longmapsto a, \quad (13)$$

that is, a mapping from belief states to actions that maximizes reward over a range of belief states. Plan execution corresponds to traversing a *policy tree*, repeatedly choosing the action with the highest value at the current belief state, and updating the belief state after executing that action and receiving an observation. In order to ensure that the observations are conditionally independent of each other given different images of the same scene, we take a new image of the scene if an action is to be repeated on the same ROI. This independence assumption is essential for the belief update described in (12), and though the images are not strictly independent the assumption works well in practice.

For a single ROI with m features (e.g., color, shape) each with n values (e.g., R_c^a, G_c^a, B_c^a, ϕ_c^a, and M_c^a), the POMDP has a state space of size $n^m + 1$. Actual scenes will contain several objects and hence several ROIs—for k ROIs we have $n^{mk} + 1$ states, that is, the state space grows exponentially. In addition, the (worst case) time complexity of POMDPs is exponential in the state space dimensions. Therefore, POMDP formulations of all but the very simple problems soon become intractable, even with state-of-the-art approximate solvers [82].

We ameliorate part of the inherent intractability by introducing a *hierarchical decomposition*: we model each ROI with a lower-level (LL) POMDP as described above, and use a higher-level (HL) POMDP to choose, at each step, the ROI whose policy tree is to be executed. This decomposes the overall problem into one POMDP with state space $2^k + 1$, and k POMDPs with state space $n^m + 1$. Essentially, we have achieved a *functional* decomposition by separating the problem of what information to process (i.e., which ROI to focus on) from how to process it (i.e., which operators to use). Without loss of generality, we assume that there are two ROIs in the image and define the HL-POMDP as $\langle \mathcal{S}^H, \mathcal{A}^H, \mathcal{T}^H, \mathcal{Z}^H, \mathcal{O}^H, \mathcal{R}^H \rangle$.

(i) $\mathcal{S}^H = \{R_1 \wedge \neg R_2, \neg R_1 \wedge R_2, \neg R_1 \wedge \neg R_2, R_1 \wedge R_2\} \cup term^H$ is the set of states. It represents the presence or absence of an object satisfying the query in one or more of the ROIs. It also includes a terminal state ($term^H$).

(ii) $\mathcal{A}^H = \{u_1, u_2, \mathcal{A}_{sp}^H\}$ are the actions. The sensing actions (u_i) denote the choice of executing one of the LL ROIs' policy trees. The special actions ($\mathcal{A}_{sp}^H$) are query-specific, and defined in a manner similar to that for the LL-POMDP. All the special actions lead to $term^H$.

(iii) $\mathcal{T}^H$ is the state transition function, which leads to $term^H$ for special actions and is an identity matrix for other actions.

(iv) $\mathcal{Z}^H = \{FR_1, \neg FR_1, FR_2, \neg FR_2\}$ is the set of observations. It represents the observation of finding or not-finding the target object when each ROI's policy is executed.

(v) $\mathcal{O}^H : \mathcal{S}^H \times \mathcal{A}^H \times Z^H \rightarrow [0, 1]$, the observation function of size $|\mathcal{S}^H| \times |\mathcal{Z}^H|$, is a uniform matrix for special actions. For sensing actions, it is obtained from the policy trees for the LL-POMDPs.

(vi) $\mathcal{R}^H$ is the reward specification. For each sensing action, it is the "cost" of running the visual policy solution of the corresponding LL-POMDP. For a special action, it is a large positive (negative) value if it predicts the true underlying state correctly (incorrectly): $\mathcal{R}(R_1 \wedge R_2, sR_1 \wedge R_2) = 100$, while $\mathcal{R}(R_1 \wedge \neg R_2, sR_1 \wedge R_2) = -100$.

A key challenge in such hierarchical formulations is the belief propagation between the levels in the hierarchy. In order to automate this belief propagation, the HL reward and observation functions in our hierarchy are based on the policy trees of the corresponding LL-POMDPs. More specifically, the HL observation function and reward specification are computed by traversing the corresponding LL policy trees, while propagating an initial belief and using LL observation functions that are modified based on the target query. However, these changes to the LL belief states and observation functions are used *only* for building the HL-POMDP model. Normal belief updates in the LL-POMDPs use an unmodified observation function and an appropriate initial belief, that is, for instance, uniform if nothing is known about the contents of the corresponding ROI. Complete details on the belief propagation between the LL and the HL can be found in [13].

The overall planning and execution cycle is as follows. Based on the target query, the available visual actions and the number of ROIs, the LL-POMDPs are created and solved. The policy trees of the LL-POMDPs are parsed to automatically generate the required components (e.g., observation functions, rewards) of the HL-POMDP. The HL-POMDP is then solved to obtain the HL policy. During execution, invoking the HL-Policy results in the selection and analysis of a specific ROI. The ROI is analyzed until a terminal action is reached in the LL. The control then returns to the HL, where the beliefs are updated based on the LL response and a new action is chosen, that is, a ROI is selected for further analysis. The process continues until a terminal action is executed in the HL and the input query is answered.

In practical scenes, objects may overlap due to occlusions or a change in viewpoint, resulting in multiple objects being enveloped in a single ROI. Processing such scenes would require visual operators that split ROIs into subregions based on one or more of its properties (e.g., color, shape, local gradients). Planning with such actions that change the perceived state of the system is a significant challenge in POMDP formulations. However, such actions can be included in our hierarchy by defining suitable transition and observation functions. In the case of the region-splitting actions, the only difference during execution would be that a state change in the LL could create new ROIs. In addition to creating and solving POMDP models for the new ROIs in the LL, a new POMDP will have to be created and solved in the HL. The execution cycle would otherwise remain unchanged. Though we do not provide more information here on the incorporation of actions that analyze images with overlapping objects, complete details can be found in [81, 83]. In addition, the quantitative results described below were obtained by including such actions in the experimental analysis.

In summary, we propose a two-level hierarchy in the (image) state and action space. In the LL, each ROI is assigned a POMDP and analyzed using the visual operators, while the HL-POMDP maintains the belief over the entire image and chooses (at each step) the ROI best-suited for further processing, thereby answering the input query. The process of creating and solving the POMDPs proceeds automatically because of the elegant belief propagation between the LL and HL. As a result, the proposed approach can be used to address a range of queries in the test domain.

4.2.1. Experimental Results. The experimental setup is as follows. The camera mounted on a robot captures images of a tabletop scene. Any change from the learned model of the background is identified as a salient region, and all such regions of interest (ROI) are extracted. The system has a sophisticated saliency operator for complex scenes [84], but the background subtraction suffices for the tabletop scenario—it is also computationally efficient. In an initial training phase, objects of known properties are put on the table, and the robot repeatedly applies the available operators on these objects. Statistics are collected regarding the operator outcomes and the run-times of the individual operators on specific ROIs. These statistics are used to estimate the observation functions and the rewards/costs of the visual operators. For instance, we had defined the dependence of the visual operator costs on ROI size as a polynomial

$$f(r) = a_0 + \sum_{k=1}^{N} a_k \cdot r^k, \tag{14}$$

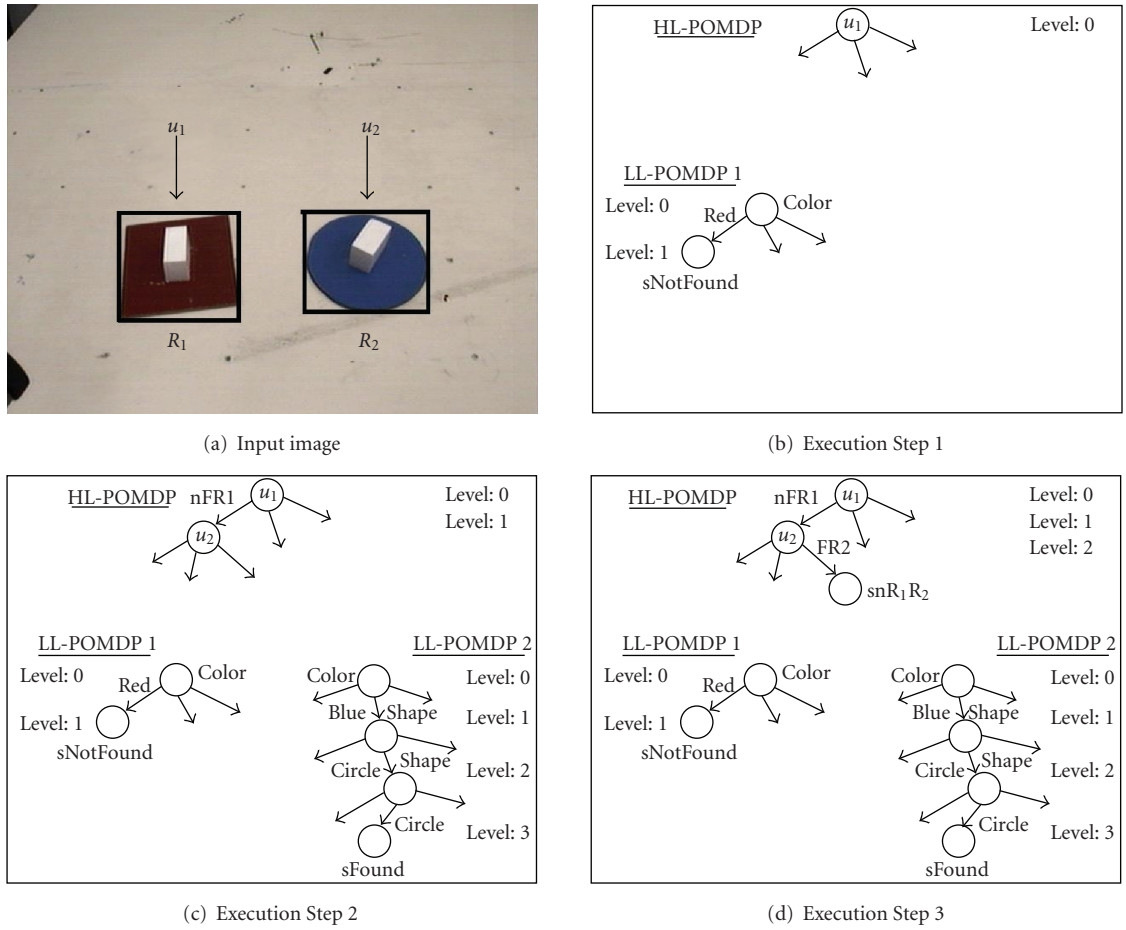
where r is the ROI-size (in pixels). The degree and coefficients of the polynomial are estimated from the collected statistics. In the experiments below, all POMDPs are created in the format of the ZMDP package [85]. The POMDPs are solved using a state of the art point-based solver [82] in the package.

We first describe the execution for the query: "where are the blue circles?" on the image shown in Figure 6(a). Since no prior information is available about either of the two ROIs the HL-POMDP first chooses to analyze the ROI R_1 because its smaller size results in lower action costs: action u_1 in Figure 6(b). The corresponding LL-POMDP runs the color operator on the ROI. Even though it is more costly, the color operator's observation function indicates a higher likelihood of success in comparison to shape, and it is hence applied first. The outcome of applying an operator is one of the possible observations (e.g., $R_c^o, G_c^o, B_c^o, U_c^o, \phi_c^o$ for *color*)—in this case the answer is R_c^o, that is, *red*. The observation is used to update the belief state. In this case, the outcome decreases the likelihood of finding a blue circle in R_1. The reward specification ($\alpha = 0.2$ in (11)) ensures a trade-off between computation and reliability, and there is no further investigation of this ROI (e.g., with a shape operator). The *best* action chosen in the next step of the LL policy for R_1 is hence a terminal action: *sNotFound*. The HL-POMDP receives the observation that the target is not found in R_1, leading to a belief update and a subsequent action selection: action u_2 in Figure 6(c). Then R_2's LL-POMDP policy tree is invoked, causing the color and shape operators to be applied in turn on the ROI. The higher noise in the shape operator causes it to be applied twice (on two independent images) before the uncertainty is reduced sufficiently. Then a terminal action (*sFound*) is chosen—the increased reliability therefore comes at the cost of execution overhead. The response from the LL-POMDP of R_2 updates the HL belief, resulting in the selection of a terminal action in the HL-POMDP: $(s\neg R_1 \wedge R_2)$, that is, a *blue circle* exists in R_2 and not R_1—Figure 6(d).

One could argue that it would be better to choose a new action in the HL at each time-step, instead of waiting for the LL-POMDP to terminate. However, the proposed approach provides the key benefit of automatic belief translation from the LL to the HL. In addition, it stops early if negative evidence is found for the target object. Finding positive evidence only increases the posterior probability of the ROI being explored—even if the HL-POMDP were to choose the next action, it would choose to process the same ROI again.

One advantage of the POMDP-based approach is that it is easy to incorporate prior knowledge in the decision-making. Consider the same scene in Figure 6(a) and the query: "where is the blue circle?", that is, the location of the single blue circle in the image is to be determined. If it is known that the *blue circle* is more likely to exist in R_2, the initial beliefs of the ROI could be modified. As a result, the cost of execution of R_2's policy would be lower (in the HL-POMDP), and R_2 would be chosen to be analyzed first leading to a faster response.

In order to evaluate the proposed approach quantitatively, the hierarchical POMDP planner (HiPPo) is compared with a modern planner that handles the non-determinism qualitatively: Continual Planning (CP) [53]. As discussed in Section 3.2, CP is a fast planner that has been applied to human-robot interaction scenarios. The planning methods were also compared against the naive approach of applying all the available operators on the ROIs. The results obtained over a set of ≈15 different queries, with ≈10 trials for each



(a) Input image

(b) Execution Step 1

(c) Execution Step 2

(d) Execution Step 3

Figure 6: Example query: "where are the blue circles?" LL-POMDP reward specification results in early termination when negative evidence is found. Belief propagation provides reliability.

such query, are summarized in Figures 7(a) and 7(b). The naive approach is denoted by "no planning" in Figure 7(b).

First, Figure 7(a) compares HiPPo against the standard POMDP approach that plans in the joint space of all the ROIs. The nonhierarchical approach soon becomes intractable, even when as few as three operators are used to analyze three or more ROIs. HiPPo, on the other hand, is reasonably efficient even as the number of ROIs in the scene increases.

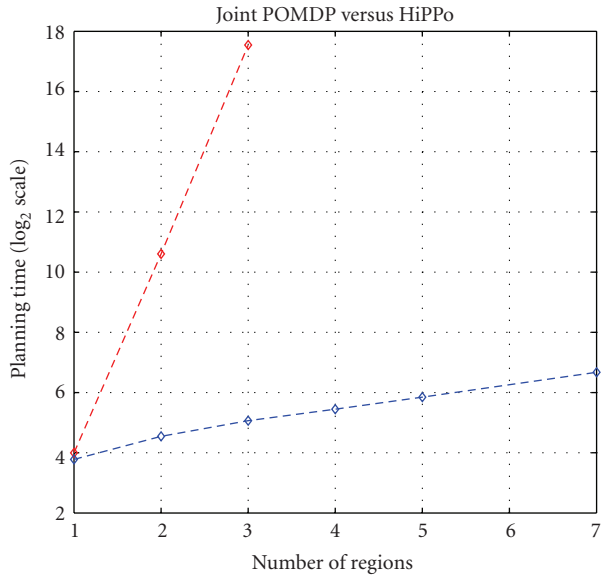
The LL-POMDPs for two ROIs typically differ only in terms of their action costs that are a function of ROI sizes. Hence, the policies computed over discretized ROI sizes were cached and reused for ROIs of similar size. This approximation makes HiPPo's planning time comparable to that of CP, and the value estimation error introduced by this approximation can be measured and used to trade-off accuracy against efficiency [81]. As observed in Figure 7(b), the total (planning + execution) time for HiPPo is only slightly larger than that of CP—HiPPo has a larger execution time because some operators are executed more than once in order to reduce the uncertainty. In addition, both planners (HiPPo and CP) are significantly faster than the naive approach.

Finally, we compared the three approaches in terms of reliability, that is, their ability to provide correct answers to queries. HiPPo provides a reliability of 90.75% that is significantly better than the reliability of CP (76.67%) or the naive approach (76.67%). CP cannot perform any better than the naive approach because it does not account for the uncertainty in operator outcomes. HiPPo, on the other hand, inherently exploits the learned models of operator uncertainties and accumulates belief to provide reliable performance.

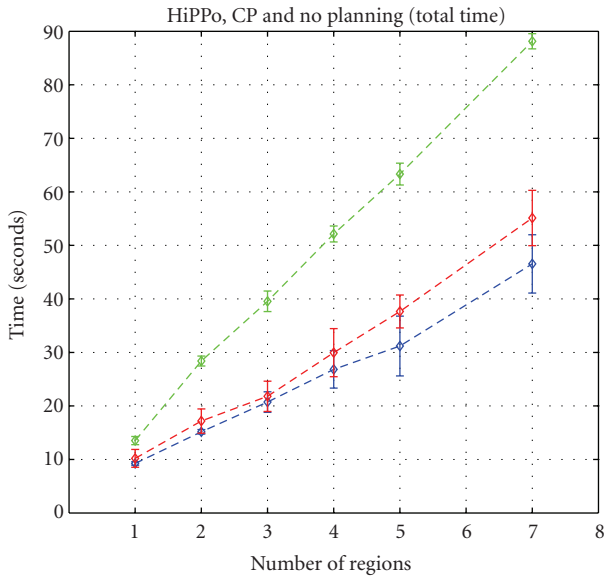
The key contribution is the hierarchical decomposition that can be modeled automatically to address a range of queries. It is easy to incorporate other operators, even those that change the state of the system. HiPPo is therefore an efficient and reliable approach towards visual processing management. Furthermore, the lessons learned in the table-top scenario can be used in other applications.

5. Conclusions and Future Work

A central goal of robotics and AI is to enable a team of robots to operate autonomously in the real world and collaborate with humans over an extended period of time. In this paper



(a) HiPPo versus joint POMDP. Joint POMDP soon becomes intractable



(b) Comparing HiPPo, CP versus No planning

FIGURE 7: Experimental results: comparing planning and execution times of HiPPo and CP against no planning. HiPPo and CP are comparable and faster than no planning.

we have described algorithms that address two key challenges to widespread deployment of mobile robots: autonomous learning and adaptation, and processing management. We have focused primarily on visual input from color cameras and summarized two key contributions: (a) a probabilistic framework where the robot autonomously learns models for color distributions and illumination, and detects and adapts to illumination changes; (b) a probabilistic sequential decision-making framework that enables the robot to autonomously tailor the visual information processing to the task under consideration.

Our bootstrap learning approach enables a robot to plan its actions in order to learn models for color distributions and illuminations. The lessons learned with this low-dimensional feature (i.e., color) can be extended to other visual features (e.g., texture, gradients) and nonvisual input (e.g., range information). Currently, the approach for planned learning requires information about the structure of the environment, that is, a map of the world. However, existing approaches in robotics and computer vision can be incorporated in this system to learn most of this structure autonomously [86, 87].

The approach described in this paper enables a mobile robot to use the learned models to detect and adapt to illumination changes. One future direction of research is to incorporate a joint model of color and illuminations. Existing research in the field of computer (and human) vision can be used to identify the parameters of this model, and the robot can estimate the values of the parameters based on data collected in its operating environment. The approach can then be extended to other visual features as well. The long-term goal would be to fully automate the learning of environmental models, and the adaptation to environmental changes.

A robot equipped with multiple sensors and multiple algorithms to process the sensory input needs a scheme to tailor the processing to the task under consideration. In this paper, we have focused on such processing management of visual input. One future direction of research is to include other operators and process more complex scenes. This increase in complexity may require a range of hierarchies in state and action spaces [66]. We are also interested in high-level scene processing, which could be defined as an additional level in the hierarchy above the existing levels. A particular region in space could be chosen for analysis with the objective of maximizing the information gain, and the existing hierarchy could then be used to process images of the chosen region in space. The key challenge would once again be the automatic belief propagation between the levels in the hierarchy.

As seen in Section 4.2, one key challenge with POMDP formulations of practical problems is the efficiency. Our hierarchical decomposition helps address part of the observed intractability. However, as the focus shifts to more complex scenarios, it may be essential to decouple the parts of the scenario that can be analyzed using nonprobabilistic methods. The POMDP-based analysis of the more uncertain components of the system would then be tractable.

Overall, we have summarized algorithms that address key challenges to the widespread deployment of mobile robots in the real world. We have shown that the robots can autonomously learn, adapt, and plan their sensory information processing. The long-term goal is to enable

robots to use a combination of learning and planning to respond autonomously and efficiently to a range of tasks, thereby collaborating with humans in a wide range of critical applications.

Acknowledgments

The author thanks collaborators from the University of Texas at Austin (Peter Stone) and University of Birmingham (UK) (Jeremy Wyatt, Richard Dearden, and Aaron Sloman). This work was supported in part by the ONR award N00014-09-1-0658.

References

[1] "Hokuyo laser," 2010, http://www.hokuyo-aut.jp/products/.

[2] "Videre design camera," 2010, http://www.videredesign.com/vision/stereo_products.htm.

[3] B. W. Minten, R. R. Murphy, J. Hyams, and M. Micire, "Low-order-complexity vision-based docking," *IEEE Transactions on Robotics and Automation*, vol. 17, no. 6, pp. 922–930, 2001.

[4] J. Pineau, M. Montemerlo, M. Pollack, N. Roy, and S. Thrun, "Towards robotic assistants in nursing homes: challenges and results," *Robotics and Autonomous Systems*, vol. 42, no. 3-4, pp. 271–281, 2003.

[5] DARPA, "The DARPA urban robot challenge," 2007, http://www.darpa.mil/grandchallenge/index.asp/.

[6] S. Thrun, "Stanley: the robot that won the DARPA grand challenge," *Journal of Field Robotics*, vol. 23, no. 9, pp. 661–692, 2006.

[7] S. Thrun, M. Beetz, M. Bennewitz, et al., "Probabilistic algorithms and the interactive museum tourguide robot minerva," *International Journal of Robotics Research*, vol. 19, no. 11, pp. 972–999, 2000.

[8] S. Thrun, D. Fox, W. Burgard, and F. Dellaert, "Robust monte carlo localization for mobile robots," *Artificial Intelligence*, vol. 128, no. 1-2, pp. 99–141, 2001.

[9] DARPA, "The DARPA grand challenge," 2005, http://www.grandchallenge.org/.

[10] S. Se, D. Lowe, and J. Little, "Vision-based mapping with backward correction," in *Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS '02)*, vol. 1, pp. 153–158, Lausanne, Switzerland, October 2002.

[11] M. Sridharan and P. Stone, "Structure-based color learning on a mobile robot under changing illumination," *Autonomous Robots*, vol. 23, no. 3, pp. 161–182, 2007.

[12] M. Sridharan and P. Stone, "Global action selection for illumination invariant color modeling," in *Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS '07)*, pp. 1671–1676, San Diego, Calif, USA, November 2007.

[13] M. Sridharan, J. Wyatt, and R. Dearden, "HiPPo: hierarchical POMDPs for planning information processing and sensing actions on a robot," in *Proceedings of the 18th International Conference on Automated Planning and Scheduling (ICAPS '08)*, pp. 346–354, Sydney, Australia, September 2008.

[14] P. Stone, M. Sridharan, D. Stronger, et al., "From pixels to multi-robot decision-making: a study in uncertainty," *Robotics and Autonomous Systems*, vol. 54, no. 11, pp. 933–943, 2006.

[15] H. Kitano, M. Asada, I. Noda, and H. Matsubara, "Robot world cup," *Robotics and Automation*, vol. 16, no. 6, p. 700, 1998.

[16] N. Hawes, A. Sloman, J. Wyatt, et al., "Towards an integrated robot with multiple cognitive functions," in *Proceedings of the 22nd National Conference on Artificial Intelligence (AAAI '07)*, vol. 2, pp. 1548–1553, Vancouver, Canada, July 2007.

[17] CoSy, "Cognitive systems for cognitive assistants," 2008, http://www.cognitivesystems.org/.

[18] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, 2002.

[19] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.

[20] B. Sumengen, B. S. Manjunath, and C. Kenney, "Image segmentation using multi-region stability and edge strength," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '03)*, vol. 3, pp. 429–432, Barcelona, Spain, September 2003.

[21] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *International Journal of Computer Vision*, vol. 22, no. 1, pp. 61–79, 1997.

[22] N. Paragios and R. Deriche, "Geodesic active regions for supervised texture segmentation," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '99)*, vol. 2, pp. 926–932, Kerkyra, Greece, September 1999.

[23] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167–181, 2004.

[24] J. Shi and J. Malik, "Motion segmentation and tracking using normalized cuts," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '98)*, pp. 1154–1160, Bombay, India, January 1998.

[25] D. Hoiem, A. Efros, and M. Hebert, "Recovering surface layout from an image," *International Journal of Computer Vision*, vol. 75, no. 1, pp. 151–172, 2007.

[26] W. Uther, S. Lenser, J. Bruce, M. Hock, and M. Veloso, "Cm-pack'01: fast legged robot walking, robust localization, and team behaviors," in *Proceedings of the 5th International RoboCup Symposium*, Seattle, Wash, USA, August 2001.

[27] S. Chen, M. Siu, T. Vogelgesang, et al., *RoboCup-2001: The Fifth RoboCup Competitions and Conferences*, Springer, Berlin, Germany, 2002.

[28] D. Cohen, Y. H. Ooi, P. Vernaza, and D. D. Lee, *RoboCup-2003: The Seventh RoboCup Competitions and Conferences*, Springer, Berlin, Germany, 2004.

[29] Y. B. Lauziere, D. Gingras, and F. P. Ferrie, "Autonomous physics-based color learning under daylight," in *Proceedings of the EUROPTO Conference on Polarization and Color Techniques in Industrial Inspection*, vol. 3826, pp. 86–100, Munich, Germany, June 1999.

[30] T. Gevers and A. W. M. Smeulders, "Color-based object recognition," *Pattern Recognition*, vol. 32, no. 3, pp. 453–464, 1999.

[31] D. Cameron and N. Barnes, "Knowledge-based autonomous dynamic color calibration," in *Proceedings of the 7th RoboCup International Symposium (RoboCup '03)*, Padua, Italy, July 2003.

[32] M. Jungel, "Using layered color precision for a self-calibrating vision system," in *Proceedings of the 8th International RoboCup Symposium (RoboCup '04)*, Lisbon, Portugal, July 2004.

[33] G. Finlayson, S. Hordley, and P. Hubel, "Color by correlation: a simple, unifying framework for color constancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1209–1221, 2001.

[34] E. H. Land, "The retinex theory of color constancy," *Scientific American*, vol. 237, pp. 108–129, 1977.

[35] G. Buchsbaum, "A spatial processor model for object colour perception," *Journal of the Franklin Institute*, vol. 310, no. 1, pp. 1–26, 1980.

[36] D. H. Brainard and B. A. Wandell, "Analysis of the retinex theory of color vision," *Journal of the Optical Society of America A*, vol. 3, no. 10, pp. 1651–1661, 1986.

[37] D. Forsyth, "A novel algorithm for color constancy," *International Journal of Computer Vision*, vol. 5, no. 1, pp. 5–35, 1990.

[38] G. Finlayson and S. Hordley, "Improving gamut mapping color constancy," *IEEE Transactions on Image Processing*, vol. 9, no. 10, pp. 1774–1783, 2000.

[39] D. H. Brainard and W. T. Freeman, "Bayesian color constancy," *Journal of the Optical Society of America A*, vol. 14, no. 7, pp. 1393–1411, 1997.

[40] Y. Tsin, R. T. Collins, V. Ramesh, and T. Kanade, "Bayesian color constancy for outdoor object recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '01)*, vol. 1, pp. I1132–I1139, Kauai, Hawaii, USA, December 2001.

[41] S. Lenser and M. Veloso, "Automatic detection and response to environmental change," in *Proceedings of the IEEE International Conference of Robotics and Automation (ICRA '03)*, vol. 1, pp. 1416–1421, Taipei, Taiwan, May 2003.

[42] F. Anzani, D. Bosisio, M. Matteucci, and D. G. Sorrenti, "On-line color calibration in non-stationary environments," in *Proceedings of the 9th International RoboCup Symposium (RoboCup '05)*, pp. 396–407, Osaka, Japan, July 2005.

[43] D. Schulz and D. Fox, "Bayesian color estimation for adaptive vision-based robot localization," in *Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS '04)*, vol. 2, pp. 1884–1889, Sendai, Japan, September 2004.

[44] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, John Wiley & Sons, New York, NY, USA, 2nd edition, 2000.

[45] M. Sridharan and P. Stone, "Color learning and illumination invariance on mobile robots: a survey," *Robotics and Autonomous Systems*, vol. 75, no. 1, pp. 1–38, 2009.

[46] M. Ghallab, D. Nau, and P. Traverso, *Automated Planning: Theory and Practice*, Morgan Kaufmann, San Francisco, Calif, USA, 2004.

[47] R. A. Brooks, "A robust layered control system for a mobile robot," *Robotics and Automation*, vol. 2, no. 1, pp. 14–23, 1986.

[48] J. E. Laird, A. Newell, and P. Rosenbloom, "SOAR: an architecture for general intelligence," *Artificial Intelligence*, vol. 33, no. 3, pp. 1–64, 1987.

[49] J. R. Anderson, D. Bothell, M. D. Byrne, S. Douglass, C. Lebiere, and Y. Qin, "An integrated theory of the mind," *Psychological Review*, vol. 111, no. 4, pp. 1036–1060, 2004.

[50] D. Draper, S. Hanks, and D. Weld, "A probabilistic model of action for least-commitment planning with information gathering," in *Proceedings of the 10th Conference on Uncertainty in Artificial Intelligence (UAI '94)*, Seattle, Wash, USA, July 1994.

[51] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, Prentice-Hall, Upper Saddle River, NJ, USA, 2nd edition, 2003.

[52] R. P. A. Patrick and F. Bacchus, "Extending the knowledge-based approach to planning with incomplete information and sensing," in *Proceedings of the 14th International Conference on Automated Planning and Scheduling (ICAPS '04)*, pp. 2–11, Whistler, Canada, June 2004.

[53] M. Brenner and B. Nebel, "Continual planning and acting in dynamic multiagent environments," *Journal of Autonomous Agents and Multi-Agent Systems*, vol. 19, no. 3, pp. 297–331, 2009.

[54] J. Hoffmann and B. Nebel, "The FF planning system: fast plan generation through heuristic search," *Journal of Artificial Intelligence Research*, vol. 14, pp. 253–302, 2001.

[55] R. Clouard, A. Elmoataz, C. Porquet, and M. Revenu, "Borg: a knowledge-based system for automatic generation of image processing programs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 2, pp. 128–144, 1999.

[56] S. Chien, F. Fisher, and T. Estlin, "Automated software module reconfiguration through the use of artificial intelligence planning techniques," *IEE Proceedings: Software*, vol. 147, no. 5, pp. 186–192, 2000.

[57] M. Thonnat and S. Moisan, "What can program supervision do for program reuse?" *IEE Proceedings: Software*, vol. 147, no. 5, pp. 179–185, 2000.

[58] S. Moisan, "Program supervision: yakl and pegase+ reference and user manual," Rapport de Recherche 5066, INRIA, Sophia Antipolis, France, December 2003.

[59] T. Darrell, "Reinforcement learning of active recognition behaviors," Tech. Rep. 1997-045, Interval Research Corp., Palo Alto, Calif, USA, 1997.

[60] L. Li, V. Bulitko, R. Greiner, and I. Levner, "Improving an adaptive image interpretation system by leveraging," in *Proceedings of the 8th Australian and New Zealand Conference on Intelligent Information Systems*, Sydney, Australia, December 2003.

[61] J. Vogel and N. de Freitas, "Target-directed attention: sequential decision-making for gaze planning," in *Proceedings of the International Conference on Robotics and Automation (ICRA '08)*, pp. 2372–2379, Pasadena, Calif, USA, May 2008.

[62] C. Kreucher, K. Kastella, and A. Hero, "Sensor management using an active sensing approach," *IEEE Transactions on Signal Processing*, vol. 85, no. 3, pp. 607–624, 2005.

[63] A. O. I. Hero, D. A. Castanon, D. Cochran, and K. Kastella, *Foundations and Applications of Sensor Management*, Springer, New York, NY, USA, 2008.

[64] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in gaussian processes: theory, efficient algorithms and empirical studies," Tech. Rep. CMU-ML-07-108, Carnegie Mellon University, 2007.

[65] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in gaussian processes: theory, efficient algorithms and empirical studies," *Journal of Machine Learning Research*, vol. 9, pp. 235–284, 2008.

[66] J. Pineau and S. Thrun, "High-level robot behavior control using POMDPs," in *Proceedings of the 8th National Conference on Artificial Intelligence (AAAI '02)*, Edmonton, Canada, July 2002.

[67] T. Dietterich, "The MAXQ method for hierarchical reinforcement learning," in *Proceedings of the 15th International Conference on Machine Learning (ICML '98)*, Madison, Wis, USA, July 1998.

[68] E. A. Hansen and R. Zhou, "Synthesis of hierarchical finite-state controllers for POMDPs," in *Proceedings of the International Conference on Automated Planning and Scheduling (ICAPS '03)*, pp. 113–122, Trento, Italy, June 2003.

[69] A. F. Foka and P. E. Trahanias, "Real-time hierarchical POMDPs for autonomous robot navigation," in *Proceedings of the IJCAI Workshop on Reasoning with Uncertainty in Robotics*, Edinburgh, Scotland, July 2005.

[70] J. M. Porta, M. T. J. Spaan, and N. Vlassis, "Robot planning in partially observable continuous domains," in *Robotics: Science and Systems*, 2005.

[71] J. Pineau and G. Gordon, "POMDP planning for robust robot control," in *Proceedings of the 12th International Symposium on Robotics Research*, San Fransisco, Calif, USA, October 2005.

[72] G. Theocharous, K. Murphy, and L. P. Kaelbling, "Representing hierarchical POMDPs as DBNs for multi-scale robot localization," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '04)*, pp. 1045–1051, New Orleans, La, USA, April 2004.

[73] M. Toussaint, L. Charlin, and P. Poupart, "Hierarchical POMDP controller optimization by likelihood maximization," in *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence (UAI '08)*, Helsinki, Finland, July 2008.

[74] B. Efron and R. J. Tibshirani, *An Introduction to Bootstrap*, Chapman and Hall, New York, NY, USA, 1993.

[75] P. S. Maybeck, *Stochastic Models, Estimation and Control*, Academic Press, New York, NY, USA, 1979.

[76] M. Sridharan and P. Stone, "Color learning on a mobile robot: towards full autonomy under changing Illumination," in *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI '07)*, Hyderabad, India, January 2007.

[77] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, NY, USA, 2008.

[78] M. Sridharan and X. Li, "Learning sensor models for autonomous information fusion on a humanoid robot," in *Proceedings of the IEEE-RAS International Conference on Humanoid Robots (ICHR '09)*, Kobe, Japan, June 2009.

[79] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[80] L. Kaelbling, M. Littman, and A. Cassandra, "Planning and acting in partially observable stochastic domains," *Artificial Intelligence*, vol. 101, no. 1-2, pp. 99–134, 1998.

[81] M. Sridharan, J. Wyatt, and R. Dearden, "E-HiPPo: extensions to hierarchical POMDP-based visual planning on a robot," in *Proceedings of the 27th Workshop of the UK Planning and Scheduling Special Interest Group (PlanSIG '08)*, Edinburgh, UK, December 2008.

[82] T. Smith and R. Simmons, "Point-based POMDP algorithms: improved analysis and implementation," in *Proceedings of the 21st Conference in Uncertainty in Artificial Intelligence (UAI '05)*, Edinburgh, UK, July 2005.

[83] M. Sridharan, J. Wyatt, and R. Dearden, "POMDP-based planning for visual processing management on a mobile robot," in *Proceedings of the 5th International Cognitive Vision Workshop (ICVW '09)*, Saint Louis, Mo, USA, October 2009.

[84] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[85] "ZMDP planning code," 2008, http://www.cs.cmu.edu/~trey/zmdp.

[86] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "MonoSLAM: real-time single camera SLAM," *Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 1052–1067, 2007.

[87] P. Felzenszwalb and D. Huttenlocher, "Efficient matching of pictorial structures," in *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR '00)*, Hilton Head, SC, USA, June 2000.

Hindawi Publishing Corporation
Advances in Artificial Intelligence
Volume 2010, Article ID 924529, 15 pages
doi:10.1155/2010/924529

Review Article

From Experimental Approaches to Computational Techniques: A Review on the Prediction of Protein-Protein Interactions

Fiona Browne,[1] Huiru Zheng,[1] Haiying Wang,[1] and Francisco Azuaje[2]

[1] *Faculty of Computing and Engineering, University of Ulster Jordanstown Campus, Shore Road, Newtownabbey, Co. Antrim BT37 0QB, UK*
[2] *Laboratory of Cardiovascular Research, Public Research Centre for Health (CRP-Santé), 120, route d'ArlonL-1150, Luxembourg*

Correspondence should be addressed to Huiru Zheng, h.zheng@ulster.ac.uk

Received 15 September 2009; Revised 13 November 2009; Accepted 6 January 2010

Academic Editor: Daniel Berrar

A crucial step towards understanding the properties of cellular systems in organisms is to map their network of protein-protein interactions (PPIs) on a proteomic-wide scale completely and as accurately as possible. Uncovering the diverse function of proteins and their interactions within the cell may improve our understanding of disease and provide a basis for the development of novel therapeutic approaches. The development of large-scale high-throughput experiments has resulted in the production of a large volume of data which has aided in the uncovering of PPIs. However, these data are often erroneous and limited in interactome coverage. Therefore, additional experimental and computational methods are required to accelerate the discovery of PPIs. This paper provides a review on the prediction of PPIs addressing key prediction principles and highlighting the common experimental and computational techniques currently employed to infer PPI networks along with relevant studies in the area.

1. Introduction

Proteins are involved in many essential processes within the cell such as metabolism, cell structure, immune response and cell signaling [1]. Although advances have been made within the realm of genome biology and bioinformatics, the function of a large proportion of sequenced proteins remains uncharacterised [2]. Uncovering the function of proteins is a complex process as one protein may perform more than one function and many proteins may have undiscovered functionality [3]. Research in [4] has suggested that the functionality of unknown proteins could be identified from studying the interaction of unknown proteins with a known protein target with a known function. Thus, the determination of protein-protein interactions (PPIs) is an important challenge currently faced in computational biology [5]. Interaction patterns among proteins can suggest novel drug targets aiding in the design of new drugs by providing a clearer picture of the biological pathways in the neighbourhoods of the potential drugs targets [6].

Large-scale high-throughput experiments have assisted in defining PPIs within the interactome (all possible PPIs in a cell). However, data generated by these experiments often contain false positives, false negatives, missing values with little overlap observed between experimentally generated datasets [3]. This may suggest that the data are erroneous, incomplete or both [3]. Previous studies have estimated that 50% of the yeast PPI map and only 10% of the human PPI network have been characterised [7].

Due to the limitations of experimental data and the need to determine PPIs, additional methods both experimental and computational are required to accelerate the discovery of PPIs. Computational methods (for example, statistical and machine learning techniques) have been applied at various stages in the inference of PPI networks, for instance, the integration of diverse heterogeneous datasets, the prediction of potential PPIs, the evaluation of predictions, and the analysis of inferred PPI networks [8–11].

The aim of this paper is to provide a review on the prediction of PPI networks focusing on the application of computational techniques to infer PPIs. The remainder of this paper is organised as follows. Section 2 describes PPI prediction tasks and principles, followed by a description on how PPIs are constructed from experimental data. Section 4 presents an overview of data sources previously employed to infer PPIs. Section 5 reviews the prediction of PPIs

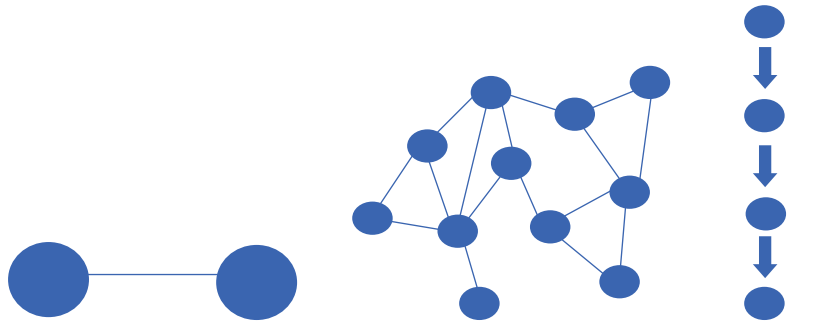


Figure 1: A graphical depiction PPI predictive tasks. (a) Direct binary interacting between two proteins; (b) proteins interact directly or interact indirectly through functional associations; and (c) Pathways represent logically linked connections for instance signal transduction. The nodes represent proteins; solid lines between the nodes represent direct interactions while dashed lines represent indirect functional associations.

using computational methods and recent studies. The paper concludes with a summary and future research.

2. Protein-Protein Interactions

Although a small percentage of proteins may operate in isolation, many proteins perform their functions by interacting with other proteins in PPI networks [9]. A protein interaction implies a specific physical contact between proteins which contributes to the formation of a biologically active protein complex. PPIs signal transduction, protein folding, cell cycle control, DNA replication and transport [10]. For instance, in signal transduction PPIs are involved in relaying signals from the cell exterior to the interior of the cell [10]. Furthermore, a protein may modify another protein through interaction. A common example of protein modification is the phosphorilation process. A kinase (a modifier protein) requires a physical contact with the target protein to add it a phosphate group. The modification of proteins can alter protein-protein interactions [9]. PPIs are involved in virtually all functions within a cell, however, a large proportion of PPIs still remain unknown [9]. This highlights the requirement to enhance our understanding of PPIs. It has been suggested that PPI patterns may aid in discovering new drug targets, and support the development of novel drugs. This is because PPI patterns illustrate biological pathways surrounding potential drugs targets [11].

2.1. Protein Interactions Prediction. The prediction of PPIs can be viewed as a binary classification problem whereby the aim is to identify pairs of proteins as either interacting or noninteracting [9, 12, 13]. There are various PPI prediction tasks including.

(1) Direct PPI prediction which involves the inference of direct physical interactions between proteins. Studies in [14, 15] have applied this predictive task to infer PPIs.

(2) Direct PPI and indirect functional association prediction whereby an interacting protein pair may not necessarily have direct physical contact but may indirectly interact through for example, complex formation. Protein scaffolding involves proteins which are important regulators in key signalling pathways. Scaffolding proteins interact with other proteins within a signalling pathway, tethering them into complexes. The study in [11] applied this principle in suggesting that proteins from the same subcellular complex may be considered "interacting" even if they do not directly physically interact with one another, but are connected through other proteins within the complex. Furthermore, the studies in [9–11, 16, 17] have employed this predictive task when inferring PPI.

(3) Pathway membership prediction whereby interactions occur in logical order (for instance, a signalling pathway). The study in [18, 19] applied this predictive task. Interactions within the pathways are often transient and may occur under specific conditions. Therefore, interactions may be difficult to measure using large-scale techniques [20].

These predictive tasks are summarised in Figure 1.

2.2. Protein-Protein Interaction Principles. PPI networks can be constructed by applying the principles of pair-wise (PW) interaction prediction or module-based (MB) interaction prediction. This review paper will focus on the prediction of PW interaction prediction as the majority of studies [9–13, 16, 21, 22] inferring PPIs apply the PW interaction prediction principle.

The aim of PW interaction prediction is to infer if two proteins are located in same protein complex [11]. The prediction of PW interaction deals with the prediction of the direct contact between two proteins. This interaction might occur between proteins appearing in the same cellular compartment by participation in the same protein complex. By contrast, the prediction of MB interactions deals with

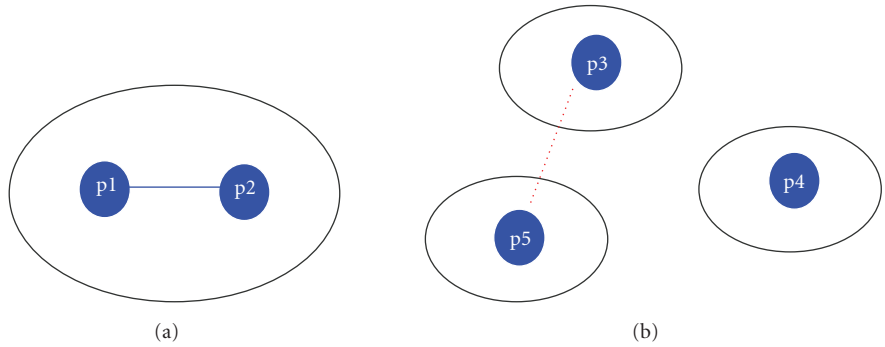


Figure 2: Graphic (a) illustrates an interaction between a pair of proteins (positive case). Graphic (b) illustrates a (negative case) noninteracting protein pair.

interactions of group of proteins, although in this case a direct contact between proteins is not required [23, 24]. Both the PW and MB prediction approaches aim to classify protein pairs or groups of proteins as either "interacting" or "noninteracting". PW and MB predictions can be used to construct a PPI network.

The concept of a positive PW interaction is graphically depicted in Figure 2(a) whereby one protein (p1) is connected (in an abstract sense) to protein (p2) for example, within the same subcellular complex. A noninteracting PW interaction is represented in Figure 2(b), whereby protein pairs in different clusters are considered to be unconnected. For instance proteins p4 and p5 are said to be noninteracting as they are in different protein complexes. Although a physical contact can be possible (as indicated by the dashed red line in (b)), an actual interaction is improbable due to these proteins belong to different protein compartments.

A graphical representation of a PPI network is illustrated in Figure 3, in which the nodes graphically represent proteins and edges represent binary interactions between proteins. This graph describes all 237 binary interactions associated with tumour suppressor proteins P53 (TP53) which has the highest degree found in the July 2009 release of HPRD database.

Limited research has been performed in the area of supervised MB PPI network prediction. The MB approach applied aims to detect whether (or not) a group of proteins (rather than a pair of proteins) belongs to the same protein complex. MB interaction prediction aims to predict various "modules" (that can vary in module size) of interacting proteins. A module can consist of a group of interacting proteins. This group may represent a protein complex. Publicly available sources, for instance, the Munich database of Interacting Proteins (MIPS) Complex Catalogue [25] contains definitions on known protein complexes and proteins within these complexes for different organisms. Figure 4 graphically illustrates the MB prediction task: (a) illustrates a group of proteins (p1, p2, p3, p4, p5, p6) found within the same complex representing a positive case and (b) the proteins p4, p8, p9 can be defined as a negative case as these proteins are found in different subcellular complexes.

Groups of genes are involved in many cellular activities. These genes behave in a coordinated fashion to perform specific biological processes [24]. Publically available high-throughput large-scale data contain a wealth of information to uncover PPI networks. The vast majority of this data is currently used for the prediction of PW interactions. However, the full potential of these data may not be fully utilised. These data could be further exploited to discover MB PPI networks [24]. Initial research suggests that modular-specific interaction predictions are an important area in predicting PPIs [24].

3. Experimental Data

Data relating to PPIs have been generated through the application of small-scale and large-scale high-throughput experimental methods. Using these data, efforts have been made to map PPIs on a proteomic-wide scale [26, 27].A review of experimental methods employed to detect PPIs including an outline of their advantages and limitations is presented in Table 1.

3.1. Small-Scale Experimental Methods. Small-scale methods focus upon specific bio-chemical or bio-physical properties of protein complexes [3]. Experimentalists often investigate several or one PPI at a time. Small-scale experiments are often applied for the detection and selection of proteins which bind to other proteins. This could be performed via affinity measurement of binding partners [3]. Small-scale experiments can be performed in vitro or in vivo. In vitro experiments are done outside of a living organism in a controlled environment and may provide valuable insights into PPIs [4]. In contrast, in vivo experiments are performed inside an organism. A selection of experimental methods is described in Table 1.

3.2. Large Scale Experimental Methods. Large-scale experiments are used to screen a vast number of proteins within the cell (i.e., across the whole proteome) [3]. Thousands of PPIs are produced which can be used to construct PPI networks. To increase the speed of discovery of PPIs, large-scale high-throughput experimental techniques have been developed to detect PPIs on a proteomic-wide scale, resulting in the production of a vast amount of interaction data [3].

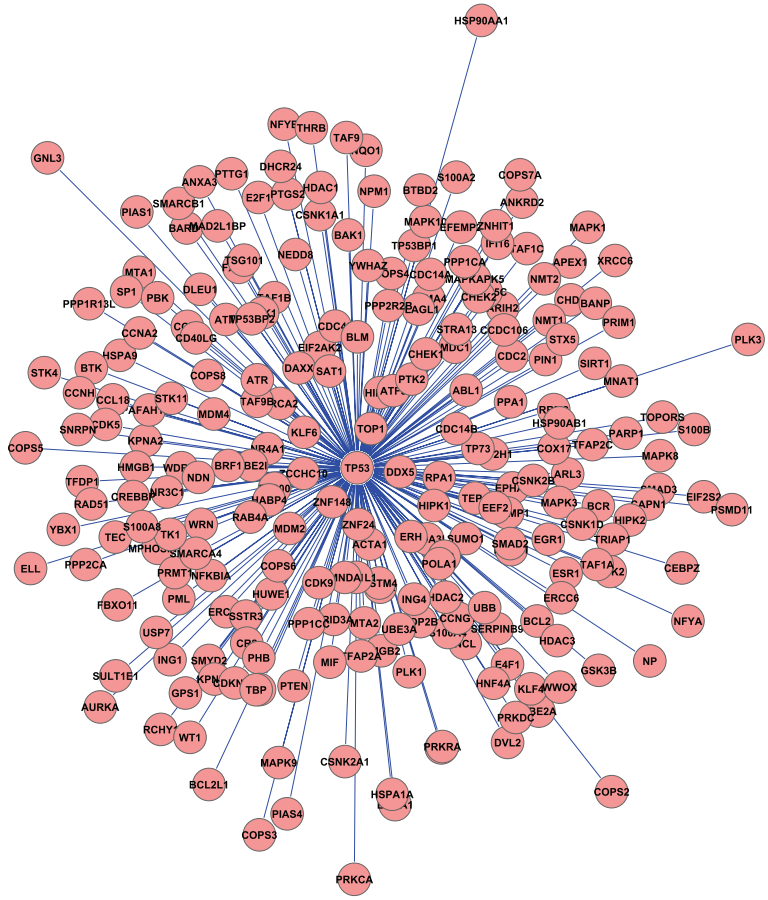


Figure 3: A graphical representation of proteins interacting with the tumour suppressor protein TP53 highlighted in yellow. All red circles represent its interaction partners found in the July 2009 release of HPRD database. This representation was performed using Cytoscape software.

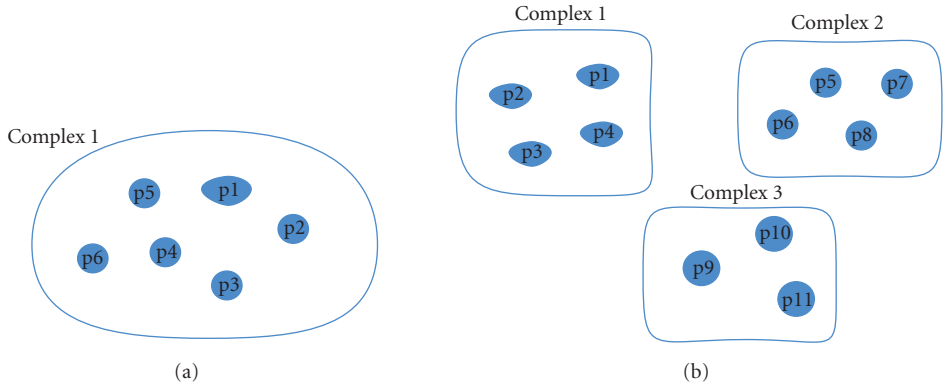


Figure 4: Graphic (a) illustrates a positive MB interaction between a group of proteins. Graphic (b) illustrates a (negative case) noninteracting group of proteins as some belong to different subcellular complexes.

A number of different experimental methods are usually required to determine, characterise and validate PPIs [3]. Common large-scale detection techniques include the Yeast Two-Hybid (Y2H) [33] and Mass Spectrometry and Tandem Affinity Purification (MS TAP) [34] which directly measure protein interactions and synthetic lethality [35], and gene coexpression [36] which indirectly provide evidence of PPIs. Descriptions of these techniques are presented in Table 2.

3.3. Construction of PPI Networks from Experimental Data. Efforts to map PPIs on a proteomic-wide large-scale have been made across different organisms including yeast [26, 27,

Table 1: A description of small-scale experimental methods applied to identify PPIs. A description and application of each method and reference to each technique are described below.

Technique	Description	Reference
Co-immunoprecipitations	To determine if two or more proteins are interacting, a purification procedure is applied to identify unknown or novel interactions	[28]
Surface Plasmon Resonance	A bait protein is attached to a gold surface where a laser light is reflected to measure small changes related to protein binding to identify unknown or novel interactions	[29]
Nuclear Magnetic Resonance (NMR)	Provides a dynamic view of PPIs when in a solution to investigate PPIs at the atomic level (i.e., smallest particle)	[30]
X-ray Crystallography	Aids in defining the structure of the interaction through crystallisation of the PPIs to investigate PPIs at an atomic level	[31]
Label Transfer	A known protein is "tagged". Interactions with this protein are obtained from detecting the tag to verify predicted or known PPIs	[32]
FRET	Proteins are labelled with fluorescence to detect interacting proteins to verify predicted or known PPIs	[32]
Far Western Blot	Proteins are applied to the blot to detect proteins of interest to verify predicted or known PPIs	[31]

33, 38, 44], fruit fly [45, 46], worm [47–49] and human [2, 7, 50, 51] through the use of experimental high-throughput technologies. Among these, yeast is perhaps one of the most investigated organisms [52]. PPI networks for yeast have been produced using various experimental techniques including Y2H, MS, and Tandem Affinity Purification (TAP) [44, 53]. Pioneering studies carried out by Schwikowski et al. [54], Ito et al. [38], Uetz et al. [33] and Gavin et al. [44] performed a comprehensive analysis of PPIs in yeast. For instance, Ito et al. [38] and Uetz et al. [33] applied the Y2H approach to infer PPI networks. Although there are limitations to the Y2H approach, it has been estimated that Y2H projects [44] have increased the amount of potential PPI data available [38].

Recent studies reported by Gavin et al. [26] and Krogan et al. [27] have utilised the experimental methods TAP and MS to construct PPI networks in yeast. Krogan et al. [27] produced a dataset consisting of 7,123 PPIs using 2,708 yeast proteins and obtained a greater coverage and accuracy in comparison to other high-throughput methods. In their study coverage was enhanced by applying rigorous computational procedures to assign confidence values to the predictions [27]. The related study in [26] produced a PPI network of the proteome averaged over all phases of the cell cycle.

The recognised significance of PPI networks has triggered huge efforts to construct PPI networks for more complex organisms. For instance, the study by Lehner and Fraser [55] developed the first draft of the human PPI map. In their study, Lehner and Fraser [55] applied the hypothesis "protein functions are usually conserved between species". Experimental data was obtained from other organisms such as yeast and integrated to produce a PPI network for human. The completed PPI network predicted interactions for one third of human genes [55]. A study by Bunescu et al. [56] produced a PPI network for human by extracting data from Medline abstracts using natural language processing and literature-mining algorithms techniques [56]. A total of 6580 interactions were identified among 3,737 human proteins and a network consisting of 31,609 interactions among 7,748 human proteins was produced through the integration of functional "omic" datasets [56].

Similar work has been performed using the organisms fruit fly and worm [45]. Formstecher et al. [57] and Giot et al. [46] both constructed a PPI network for the fruit fly uncovering 4,679 proteins and 4,780 interactions.

3.4. Limitations of Experimental Methods. The development and application of large-scale high-throughput technologies have resulted in the generation of vast amounts of data on PPI. This has contributed to the identification of PPIs [3]. However, data obtained by large-scale experimental methods are often noisy, incomplete and contradictory (i.e., weak predictive data sources) with thousands or tens of thousands interactions yet unknown [3]. Experimental methods can only identify a subset of the interactions that occur in an organism, therefore coverage (i.e., the area of the proteome covered by protein pairs) of the interactome is limited [58]. Furthermore, high-throughput studies are difficult to reproduce [3]. Methods such as the Y2H system exhibit high false positive and false negative interaction rates [3]. Traditional methods (e.g., small scale manual experiments) to infer PPIs may produce more accurate results compared to single source high-throughput methods. However, they are expensive and time consuming [4]. Furthermore different experimental conditions applied in different laboratories protocols makes it difficult to compile this information in a meaningful way. Therefore the use of a uniform method which is occurring in the large-scale approach facilitates the comparison. Due to inadequacies exhibited by both the small and large-scale experimental methods, advancements in computational methods are needed in the prediction of PPIs [8].

Table 2: An overview of large-scale experimental methods applied for the detection of PPIs on an interactome-wide scale. The first column states the name of the experiment followed by a description and a reference.

Technique	Description	Reference
Y2H	The Y2H employs a "two-plasmid" approach in yeast. The yeast protein GAL4 is a transcriptional activator consisting of two domains. The domains must be in close proximity to start the transcription process. The two plasmids are placed into a cell. A physical interaction occurs if the GAL4 binding and activation domains come together if physical interaction occurs, demonstrating that the "bait" and "target" bind [37]. This technique provides evidence of direct physical interactions between proteins.	[33, 38, 39]
MS TAP	Affinity tags are attached to a protein of interest (target), systematic precipitation of bait proteins is performed. Proteins are separated according to their mass to uncover purified protein complexes. Proteins are removed from a gel and analysed by MS techniques [40]. This technique provides evidence of direct physical interactions between proteins.	[34]
Gene coexpression	Gene expression profiles can be obtained from cell cycle experiments and the measurement of gene expression levels when the cell is under different conditions [41]. Gene expression similarity values may be calculated as the Pearson correlation co-efficient between expression levels of two proteins Protein pairs that are coexpressed are more likely to be interacting proteins [35, 42]. This technique provides indirect evidence of interactions between proteins.	[35, 42]
Synthetic lethality	This method involves the deletion or mutation of two genes which are viable alone, but cause lethality when combined in a cell under specific conditions [36, 43]. Synthetic interactions may detect PPIs between gene products, their occurrence in a pathway or participation in a function [40]. This technique provides indirect evidence of interactions between proteins.	[36, 43]

4. Data Sources

Data obtained from large-scale high-throughput experiments and "omic" information can be employed to support large-scale prediction of PPI networks [11]. However, individually these data are often limited in terms of accuracy and interactome coverage [6]. For example, estimated error rates of high-throughput experimental PPI datasets range 41–90% [6]. Studies in [10, 16, 17, 58, 59] have suggested that the integrating heterogeneous biological data using supervised machine learning methods can improve both the interactome coverage and predictions of PPIs. For example, Jansen et al. [11] integrated four features: (1) mRNA coexpression correlation, (2) MIPS functional similarity, (3) GO annotations, and (4) coessentiality using a Naïve Bayesian (NB) approach to infer PPIs in yeast. An increase in interactome coverage and predictive performance was observed when these features were integrated in comparison to the application of individual features alone [11]. Rhodes et al. [60] inferred PPIs in human by combining biological features within a probabilistic framework. These features included (1) homologous PPIs, (2) mRNA coexpression correlations, (3) functional similarity based on GO annotations, and (4) enriched domain pairs. By integrating these diverse heterogeneous features, ~40,000 human PPIs were predicted.

In this section, a brief description of a sample of data sources employed in the prediction of PPIs are presented.

mRNA Coexpression (COE). Based on the assumption that proteins which are coexpressed are more likely to interact than protein that are not-coexpressed, the COE information has been widely employed for the predictive task of inferring PPIs. For example, in yeast, the COE has been constructed from publicly-available expression data which represent the "time course of expression fluctuations during the yeast cell cycle and the Rosetta compendium" [61]. The data consists of expression profiles from 300 deletion mutants and cells which have undergone various chemical treatments. Pearson's correlation values were calculated for each protein pair in the data set.

MIPS Functional Similarity (FunCat). The FunCat data source is based on the assumption that proteins found within the same biological process are more likely to interact in comparison to proteins from different biological processes. Protein pairs are defined as interacting if they both belong to the same biological process or noninteracting if they belong to different biological processes (as defined by the Functional Catalogue). In the study published by Jansen et al. [11],

the FunCat was constructed by calculating similarity values between protein pairs annotated in the MIPS Functional Catalogue.

Coessentiality (ESS). The construction of the ESS dataset for the prediction of PPI is based on the assumption that proteins can be experimentally characterised as either essential (EE) or non-essential (NN), which may be used an indicator that the proteins are both members of the same complex. A protein can be classified as essential or non-essential, based on the viability of the cell when the gene is knocked out [11]. If two proteins exist in the same complex they are either essential or non-essential but not both.

The ESS dataset used in [11] is derived from the MIPS complex catalogue, transposon and gene deletion experiments [25].

Absolute Protein Abundance (APA). APA has been employed as a predictive feature to infer PPIs in yeast based on the hypothesis that an interacting protein pair should be present in stoichiometrically similar amounts (that is, the calculation of reactants and products in a chemical reaction) [10]. In one of the pioneering research published by Jansen and his colleagues [11], protein abundance is calculated by counting the number of proteins within a cell. APA values have been obtained from a number of experimental methods including gel-electrophoresis and mass spectrometry which have been scaled and merged by Greenbaum et al. [62].

Domain (DOM). The DOM has been employed as a predictive feature to infer PPIs in human. PPIs involve the physical interaction between domains (of proteins), therefore, PPIs could be inferred by identifying domain pairs enriched by known PPIs [63]. Hyper geometric distribution values between protein pairs were calculated in [59] to provide DOM feature values.

Phylogenetic Profiles. Pairs of non-homologous proteins that are either absent or present together in different organisms are more likely to have co-evolved [64]. Co-evolution has been observed between interacting proteins, such as chemokine and its receptors [64]. The study by Pellegrini et al. [65] examined co-occurrence or absence of genes across multiple genomes inferring functional relatedness.

Interologs. Interolog mapping involves the transfer of interaction annotation from one organism to another using comparative genomics [11]. This approached was used in the study by Yu et al. [66] to assess the degree to which interologs can be reliably transferred between species as a function of the sequence similarity between the corresponding interacting proteins.

Synthetic Lethality. This method involves the deletion or mutation of two genes which are viable alone, but cause lethality when combined in a cell under specific conditions. As the mutations are lethal, they should be synthetically generated. Synthetic interactions may detect PPIs between gene products, their occurrence in a pathway or participation in a function [40]. For instance, the application of synthetic lethality experiment discovered that the unknown function of the gene "YLL049W" belonged to the pathway dynein-dynactin [67].

4.1. Availability of Data. Various databases store information relating to PPIs (e.g., direct physical PPIs or data relating to protein complex membership) for different organisms. These data have been extracted from manually curated data or by data-mining literature. A list of popular databases containing PPIs is provided in Table 3.

4.2. Gold Standards. Gold Standards (GS) contain known interacting (positive) and noninteracting (negative) protein pair cases and can be employed to: (1) train classifiers for the predictive task of PPI inference or (2) evaluate computationally predicted PPIs. Furthermore, the quality of statistical and machine learning methods will depend upon the relevance and validity of the GSs to the prediction problem under study [11]. The study by Jansen et al. [11] suggested that a GS should be (1) generated independently from the data sources applied to infer PPI, (2) contain a sufficient number of protein pairs to provide reliable statistics, and (3) to be free of systematic bias. However, the selection of a GS for the prediction of PPIs can be problematic. For example, selecting a GS with adequate coverage of the interactome and defining what the GS specifically measures (i.e., complex membership, direct physical interactions) can be a difficult task. High quality positive GSs (GSP) are often assembled from interactions generated from small-scale manually curated experiments [2].

The construction of a negative GS (GSN) is also difficult as there are no "gold standard" noninteractions. Two methods to construct GSNs have been described in the literature: (1) studies in [8, 9, 11, 35] have suggested that high quality noninteractions can be generated by selecting pairs of proteins from different subcellular compartments, as they are more likely to be prevented from participating within biologically relevant interactions [8]; (2) other studies in [71, 72] have selected noninteracting pairs uniformly at random from a set of all proteins pairs that are not known to interact. Both of these two methods have limitations. For example, proteins selected from different cellular compartments may interact (for example proteins in the nucleus and cytoplasm) [72]. Moreover, due to the incompleteness of PPI networks, a GSN constructed by randomly selecting protein pairs may contain undiscovered true positive protein pairs, and thus may counteract the successful prediction of those [71].

GSs employed for the predictive task of PPI inference are often highly unbalanced with many more noninteracting pairs than interacting pairs. This is because the number of true biological PPIs is a rare phenomena among all possible protein pairs in the interactome [8]. For instance, yeast has ~6000 proteins resulting in ~18 million protein pairs. Estimates place the number of interacting protein pairs in yeast around 10,000–20,000 [6].

Table 3: A list of popular databases containing PPI information for organisms.

Database	Data Type	Organisms	URL	Reference
BioGRID	Experimental and manually curated data	22 organisms including: yeast, fruit fly, worm, human	http://bind.ca	[68]
Database of Interacting Proteins (DIP)	Experimental and structural data	274 organisms including: yeast, human, rat, mouse, fly and worm	http://dip.doe-mbi.ucla.edu/	[15]
Munich Database of Interacting Proteins (MIPS)	Experimental, functional predictions and manually curated	Mouse, human, yeast	http://www.helmholtz-muenchen.de/en/ibis	[25]
Saccharomyces genome database (SGD)	Experimental and manually curated	Yeast	http://www.yeastgenome.org/	
IntACT	Experimental and manually curated	Includes the organisms: yeast, human, rat, mouse, fly and worm	http://www.ebi.ac.uk/intact/site/	[69]
Human Protein Reference Database	Experimental and manually curated	Human	http://www.hprd.org/	
MINT	Experimental and manually curated	30 organisms including: yeast, human, rat, mouse, fly and worm	http://mint.bio.uniroma2.it/mint/	[70]

The web-based system GRIP (Gold Reference dataset constructor from Information on Protein complexes) outlined in the study by Browne et al. [73] provides researchers with the functionality to create reference datasets for PPI prediction in yeast. GRIP integrates the functionality for constructing reference datasets, protein complex membership matching and protein complex matching. Recent research by [10, 11] demonstrated that the generation of reference datasets are critical for the verification of computationally-inferred PPI networks. A study by [74] implemented reference datasets constructed using GRIP to demonstrate that supervised statistical and machine learning techniques can be successfully applied to PW and MB interaction prediction.

5. Computational Prediction of PPIs

The prediction of PPIs can be defined as a classification problem. For instance, a statistical or machine learning technique can be applied to the predictive task of determining whether a pair of proteins are interacting or noninteracting [9]. However, the prediction of PPIs is a complex task. For example, the datasets are highly skewed (i.e., there are more noninteracting PPIs than interacting PPIs) [17] and may be noisy and contain missing values [11]. Therefore, the selection of an appropriate classification technique is an important task. Classifiers that perform well in other problem domains may not perform as well within the realm of PPI prediction [75]. It is essential to assess available classification models for inferring PPIs [75]. This section will provide an overview of statistical and machine learning techniques and their application to PPI inference.

5.1. Statistical and Machine Learning Techniques. Computational methods (for example, statistical and machine learning techniques) have been applied at various stages in the inference of PPI networks. For instance, the integration of diverse heterogeneous datasets; the prediction of potential PPIs; the evaluation of predictions and the analysis of inferred PPI networks [8–11]. A summary of statistical and machine learning techniques including (1) K-Nearest Neighbour (KNN), (2) Naïve Bayesian (NB), (3) Support Vector Machine (SVM), (4) Artificial Neural Networks (ANN), (5) Decision Tree (DT), and (6) Random Forest (RF) are presented in Table 4. These techniques have been selected as they have previously been employed for the predictive task of inferring PPI networks.

5.2. Review of Current Studies. A number of studies have combined both direct and indirect experimental information in a supervised learning framework to predict PPIs [9, 11, 59]. These studies focus on the prediction of PPIs in yeast and human. The Yeast is an important experimental organism for the prediction of PPIs as it has been extensively characterised and the genome is fully sequenced [83]. Furthermore, yeast displays many features of higher eukaryotes (such as human). This is important as cellular processes are often conserved between eukaryote species [83]. Relatively few studies have been performed to computationally predict PPIs in human. Compared to yeast, the human interactome is considered more complex due to a larger number of proteins, post-translational modifications, splice isoforms and dynamic regulations [59]. Mapping human PPIs could provide a framework to improve understanding of protein function in complex diseases such as cancer [60]. Table 5 provides a summary of these studies.

5.2.1. PPI Prediction for Yeast. The study by von Mering et al. [3] was one of the first studies to discuss the issues

Table 4: A summary of machine and statistical learning approaches applied to the predictive task of inferring PPI. Advantages and limitations for each approach are presented along with a reference to the studies where they have been applied.

Classifier	Description	Reference
NB	Ability to integrate diverse heterogeneous data. Can handle missing data. Assumes conditional independence between datasets. Performance of NB deteriorates when dependencies between features exist.	[10, 11]
KNN	Classification method which has been considered "simple but powerful" providing competitive performance compared with other classifiers [76]. Classifier performance may deteriorate if many variables are used or if the GS is not balanced.	[17, 74]
SVM	Can handle non-linearly separable datasets. Can incorporate prior information.	[77]
RF	Can handle missing values. Can integrate diverse heterogeneous data	[78, 79]
ANN	Ability to recognise complex patterns	[80–82]

of computationally predicting PPI using experimental data. Data such as: Y2H, MS, mRNA gene-expression, gene fusion, gene neighbourhood and phylogenic profiles were employed in their study. Results obtained highlighted a low overlap between the various data sources. This suggests that experimental methods: (1) may not have reached saturation; (2) methods produce high false positives; (3) methods identify different interactions. von Merring et al. [3] suggested high-throughput experimental data could be integrated to improve the confidence of PPI predictions. The integration of diverse heterogeneous data in their study lead to a reduction in the number of false positives, however the coverage of the interactome was limited [3]. For example, only ~2,400 of a possible 80,000 protein interactions in yeast were supported by more than one method [3].

Jansen et al. [11] applied a Bayesian Network (BN) approach to predict PPIs using four features: gene coexpression, GO biological process similarity, MIPS functional similarity, essentiality. The MIPS Complex Catalogue [25] was employed as a GS. Individually, the datasets were weak predictors of PPIs. However, when the datasets were integrated via BN, accurate PPI networks were produced providing a comprehensive view of the yeast interactome [11]. Troyanskaya et al. [13] also applied a BN approach to combine diverse data sources for the inference of PPIs in yeast. The data sources employed included: gene coexpression and physical associations. The GS was constructed from information extracted from the GO [84]. The study in [14] employed a confidence measure for predictive PPIs using a Logistic Regression approach. Their study produced a high-confidence PPI network for over one third of the yeast proteome. Lin et al. [9] repeated the experiments by [11] and employed the classifiers NB, Random Forest (RF) and Logistic Regression to infer PPIs. Using only a subset of the integrated datasets with no missing values, Lin et al. [9] discovered that the MIPS and GO functional datasets were the most dominant features.

The study by Browne et al. [85] investigated the integration of functional genomic data for the prediction of PPI in yeast. A Bayesian classifier was employed to reassess the limits of genomic integration using seven genomic features ranging from coexpression to essentiality. Assessment methods such as true positive/false positive (TP/FP) rate and sensitivity were applied as comparative predictive measures to the ROC curve. A clear increase in predictive performance was observed using the measures TP/FP and sensitivity when the features were integrated.

A RF classification method was employed by Qi et al. [78] for the prediction a PPI network in yeast. The RF classifier predicted PPIs with an average sensitivity of ~80% and a specificity below 65%. Additionally, Qi et al. [78] demonstrated how selection and encoding of datasets has an impact upon the PPI predictive performance. Various classification techniques such as RF, RF integrated with KNN, NB, DT, Logistic Regression and SVM were applied. It was discovered that the RF classifier performed robustly in inferring PPIs.

Lu et al. [10] extended a study in [11] to evaluate the predictive limits of "omic" integration using a NB approach. Sixteen diverse datasets ranging from: synthetic lethality to MIPS functional similarity was integrated to predict PPIs. Compared to the previous study in [11], relatively high predictive accuracies were obtained. However, the addition of "weaker" datasets provided only marginal improvement in terms of predictive performance. This is in comparison to the integration of seven "strong" datasets. The NB classifier assumes conditional independence between datasets, Lu et al. [10] provided evidence of only marginal dependencies between the datasets employed in the study. However, as high-throughput technologies continue to emerge, datasets produced will present more potential dependencies. Therefore, the NB classifier may not be the optimal computational approach to predict PPIs. Dependencies between datasets may possibly cause the predictive accuracy obtained by NB to decrease [10].

Myers et al. [24] constructed a system entitled "bioPIXIE" to provide integration, analysis and visualisation of PPI predictions in yeast. This system used a BN approach; the PPIs predicted were validated by recovering networks for 31 known biological processes in yeast. Their study outlined critical issues when evaluating functional "omic" data. These include (1) bias and inconsistencies of GS, (2) the selection of negative GS, (3) number of proteins pairs in the GS. The GS employed in their study was constructed based on expert curation [24].

5.2.2. PPI Prediction for Human. The human proteome is considered more complex in comparison to the yeast proteome. This is due to a larger number of proteins, dynamic regulations, and post-translational modifications in human [2]. Moreover, more data sources are available for yeast in comparison to human [2]. This has resulted

Table 5: A summary of related work in inferring PPI networks. The first column presents the study, this is followed by advantages and limitations of the study.

Related Work	Advantages	Limitations	Ref
A Bayesian networks approach for predicting protein-protein interactions from genomic data	Pioneering study which applied a Bayesian approach to infer PPI in yeast by integrating diverse genomic data	Only four features were integrated. By integrating more features an improvement in interactome coverage and classification predictive performance may be achieved	[11]
A Bayesian networks approach for predicting protein-protein interactions from genomic data	Sixteen diverse features were integrated using a NB classifier to predict PPI in yeast.	The NB classifier approach was applied—this classifier assumes feature independence. Subtle dependencies between features may have an adverse affect on the NB performance. ROC curves were the only assessment method applied to measure the predictive performance of the classifier.	[10]
Information assessment on predicting protein-protein interactions	Application of RF, NB and logistic regression for the prediction of PPI in yeast. Discovered MIPS and GO annotation data were dominant features	Only used subset of data. Missing data was removed.	[9]
Probabilistic model of the human protein-protein interaction network	One of the first studies to integrate diverse "omic" data for the prediction of PPI in human.	A NB approach was employed to infer PPI. Three gene coexpression datasets will employed however only the maximum likelihood ratio per gene coexpression data source per protein pair was considered	[60]

in a limited number of studies which have computationally inferred PPIs for human.

Rhodes et al. [60] provided an integrated analysis of human PPIs using a NB approach. The data employed consisted of homology, gene coexpression, shared biological process and domain data. Information extracted from the Human Protein Reference Database (HPRD) [63] was used as the GS to evaluate PPI predictions. Experimental methods confirmed protein interactions predicted by the framework.

Xia et al. [86] integrated 27 heterogeneous data sources using a probabilistic approach to infer PPIs for human. An integrated network database was constructed and provides the functionality of prediction and visualisation of genes of interest. Scott and Barton [2] constructed a probabilistic framework to integrate diverse features including: gene coexpression, localisation information, domain-domain interactions. A total of 37,606 PPIs were predicted, 80% of which are not found in other human PPI databases.

A recent study by Qi et al. [59] addressed the limitation of missing data and feature redundancy in inferring PPIs in human. A "mixture-of-features" framework was applied to predict PPIs. They employed obtained Precision-Recall curves to evaluate the predictive performance of classifiers including: NB, SVM and RF. In their study, 18 potentially novel interacting protein pairs were identified.

Browne et al. [73] applied a fully connected BN approach to integrate diverse "omic" features for the inference of disease-specific PPI networks. The case study integrated three gene coexpression datasets relevant to human heart failure along with other datasets to reconstruct a PPI network relevant to the development of dilated cardiomyopathy. By modelling relationships between multiple datasets of the same "omic" type, an improvement in prediction performance was achieved in terms of partial AUC and the ratio of TP/FP by the fully connected BN approach in comparison to the maximum likelihood ratio and NB approaches.

The studies highlighted above for prediction of PPIs in human and yeast share commonality in the types of data sources that were employed and in some cases the predictive computational methods employed. A commonly

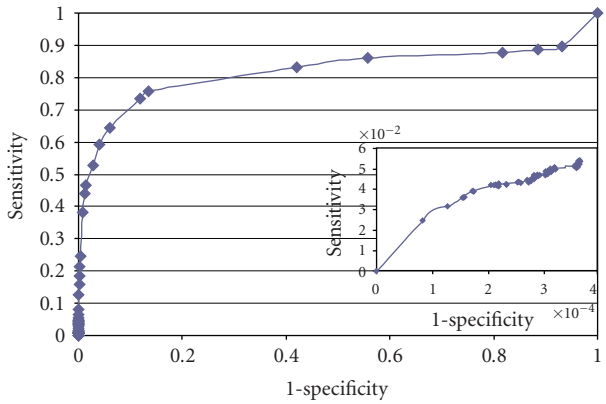


Figure 5: Plotting of a ROC curve when the integration scheme including seven features in applied. The sensitivity is plotted against the 1-specificity for different likelihood thresholds. Various likelihood thresholds have been highlighted on the ROC to illustrate the different AUC thresholds. The inset graph illustrates the AUC when the likelihood of 600 and greater is selected.

applied computational predictive approach in these studies was the Bayesian classifier. This classifier can handle diverse heterogeneous data types and missing values which is advantageous when inferring PPIs as the data is often obtained from different sources and may suffer from missing values. The studies differ in the data sources employed for the prediction of PPIs, selection of GSs and the evaluation methods employed. Therefore it is difficult to obtain a comparative view of the different computational methods in predicting PPIs. The study by Browne et al. [75] and Qi et al. [17] performed a comparative review of different computational techniques when inferring PPIs using a selection of supervised learning approaches. In this study the same data sources, GSs and evaluation methods were applied to provide a comprehensive comparison of computational approaches when inferring PPIs in yeast.

5.3. Limitations of Computational PPI Prediction. Despite the relative success of the computational methods applied to infer PPIs, no approach can accurately predict all PPIs within an interactome. A number of computational limitations outlined below need to be addressed for this to become reality. For example, computational efficiency of the classifier needs to be addressed. For instance, classifiers such as KNN have been found to be time consuming and processor intensive [17]. Statistical and machine learning methods are known to exhibit systematic bias [75]. A computational technique may produce solutions that favour a limited number of specific situations or circumstances [75]. Computational classification techniques make assumptions, such as the NB which assumes dataset independence [10]. A number of studies have applied different predictive models to predict PPIs in yeast [8, 12, 17, 21, 87, 88]. However, there is difficulty when comparing and contrasting results from these studies due to differences in the predictive models, features, GS and predictive tasks applied. For relatively simple organisms, such as yeast, more datasets are available for the prediction of PPIs compared to more complex organisms such as human [2]. As organisms increase in complexity the data obtained and the task of PPI prediction also increase in complexity [2]. Datasets obtained for organisms such as human are sparse and suffer from high rates of false positives and false negatives with little coverage of the interactome [2]. Computational docking in protein folding may be employed as a local prediction method to computationally infer PPIs. However this method has only been successful when used on a small-scale [89].

5.4. Overview of Predictive Performance Measurement Techniques. The performance of a supervised machine learning framework is evaluated in terms of predictive quality and potential significance of PPI predictions. The selection of a measurement approach is essential in determining the predictive performance of a supervised learning approach. Various studies employ different predictive quality measures making it difficult to compare classification performance. A selection of assessment methods previously applied to evaluate the predictive performance of classifiers when inferring PPIs are presented below.

5.4.1. Cross Validation (CV). To estimate the performance of a predictive model CV can be applied. In n fold CV the dataset is partitioned into segments, analysis is performed on one segment (called the training set), one segment is left out for validation (called the test set). To reduce variability CV are preformed with different partitions with the validation results averaged over the different CVs.

5.4.2. ROC Curves. ROC curves have been commonly used to illustrate classification performance when predicting PPIs [10, 75]. In ROC analysis, the accuracy by which a model can separate positive from negative instances is investigated [19]. ROC curves plot in a single graph the sensitivity against 1-specificity over a range of different thresholds. The graph consists of a set of points each computed for a different threshold. For each point, the vertical co-ordinate represents the sensitivity and the horizontal co-ordinate the 1-specificity. Therefore, the predictive quality of a classifier is assessed by measuring the sensitivity and 1-specificity. The counts of the: TP, TN, FP and FN are obtained from the CV analysis. The formula used to calculate sensitivity and specificity are detailed below:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{1}$$

$$1 - \text{Specificity} = \frac{\text{FP}}{\text{FP} + \text{TN}}. \tag{2}$$

As illustrated in Figure 5, a predictive dataset will produce a ROC curve that rises steeply to the left hand side of the graph and has a large area under the curve. The AUC is a measurement of the area under the ROC Curve. A perfect classifier will have an AUC value of 1.0. A prediction model based on random assignments of pairs of proteins to classes would give an AUC equal to 0.5.

The majority of the AUC of a ROC curve when inferring PPI in yeast may not represent biologically informative results. For example, Figure 5 illustrates a ROC curve plotted when 7 features were integrated using the NB classifier to infer PPIs in yeast [85]. Various likelihood thresholds have been highlighted to illustrate how the majority of the AUC relates to a likelihood threshold which is less than or equal to 1. Therefore, the AUC of the ROC curve is not considered biologically meaningful as a threshold greater than or equal to 600 is required to predict a positive interaction (posterior odds of an interacting protein pair in yeast). The threshold of 600 is suboptimal for the trade-off between the TP and FP rate highlighted in Figure 5, from this it can be observed that relatively little of the total AUC is represented by a threshold of 600 and above.

These results highlight the importance of selecting an adequate assessment method for the quality testing to assess the quality of a prediction model. In the study by Browne et al. [73] and Jansen et al. [11] alternative representative methods: True Positive (TP)/False Positive (FP) rate and TP/Positive (P) have been employed as alternative representative measures to assess the performance of prediction model. These are detailed below.

5.4.3. TP/FP Ratio and TP/P. The TP/FP ratio is plotted against the threshold (TH) of likelihood ratio as a measure of the probability of a real interaction. This measure has previously been employed in the study by Jansen et al. [11]:

$$\left.\frac{\mathrm{TP}}{\mathrm{FP}}\right|_{L=\mathrm{TH}} = \sum_{L=\mathrm{TH}} \frac{N_{\mathrm{pos}}(L)}{N_{\mathrm{neg}}(L)}. \tag{3}$$

The $N_{\mathrm{pos}}(L)$ and $N_{\mathrm{neg}}(L)$are the number of interacting and noninteracting protein pairs in the GS with a given likelihood ratio of L.

The TP/P ratio is applied as a measure of coverage whereby P represents the number of positives in the GS.

5.4.4. Partial ROC Curve. Rather than measuring the AUC under the entire ROC curve, it may be more informative to consider the area under a portion of the curve. This is referred to as the Partial ROC curve AUC which has previously been employed in the study by Browne et al. [73] to illustrate the number of true positives identified by the Bayesian classifier against specified likelihood cut-off rates which represent thresholds of biologically meaningful predictions.

Partial ROC curves have been applied as evaluation measures in recent studies [2, 90]. In these studies, the partial ROC plots the AUC whereby the false positive rate is low (for instance, measuring the AUC until 50 negative predictions have been reached) [90]. The partial curve applied in the study by Browne et al. [73] differs from previous studies as the area of the ROC whereby the predictions exceeding a selected threshold is measured. For yeast the threshold selected is 600 and for human 400. These thresholds are based upon the prior odds of an interacting protein in yeast and human, respectively. The partial ROC measures are referred to as ROC_{600} for yeast and ROC_{400} for human.

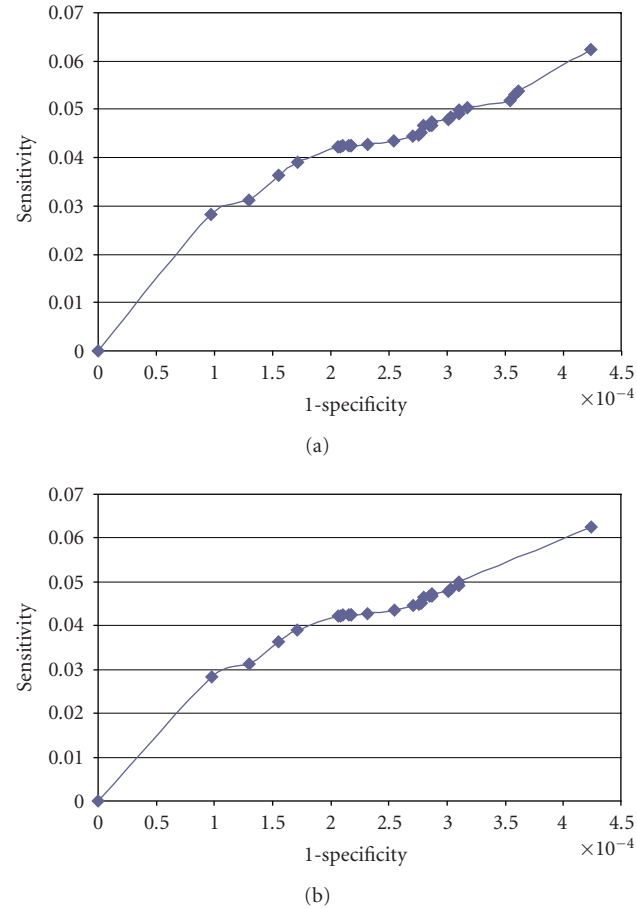


Figure 6: Graphical representation of a Partial ROC curves. (a) represents a portion of the ROC curve plotted when predicting PPIs in yeast (the threshold is greater than 600); (b) represents a portion of the ROC curve plotted when inferring PPI predictions in human where the threshold is greater than 400.

ROC_{600} and ROC_{400} measure high quality predictions. Figure 6 illustrates examples of partial ROC curves, (a) a portion of the ROC curve plotted representing PPI predictions in yeast whereby the threshold is greater than 600; (b) a portion of the ROC curve plotted representing PPI predictions in human whereby the threshold is greater than 400.

6. Conclusions and Future Trends

PPIs play an important role in many biological functions and diseases [7]. A wealth of biological data has been provided though the advent of experimental high-throughput technologies [3]. Data obtained from large-scale high-throughput experiments and "omic" information (e.g., essentiality and functional information) can be employed to support large-scale prediction of PPI networks [11]. However, individually, these data are often limited in terms of accuracy and interactome coverage [6]. For example, estimated error rates of high-throughput experimental PPI datasets range from 41–90% [6]. Studies in [10, 16, 17, 58, 59] have suggested that the integration heterogeneous

biological data using supervised machine learning methods can improve both the interactome coverage and predictions of PPIs.

PPI networks can be constructed using a number of prediction principles including PW interaction prediction and MB interaction prediction.

Statistical and machine learning techniques can be applied in the computational prediction of PPI [10, 11, 74]. These techniques are required for the integration of heterogeneous features and the inference of PPI networks. However, computational techniques may make assumptions and as of yet, there is no standard machine learning technique within the area of PPI prediction [75]. Further investigation is required to assess the predictive performance of different statistical and machine learning techniques employed to integrate diverse features for the prediction of PPIs.

AUC values from the ROC curves are commonly employed as the evaluation method to assess the predicative performance of the classifiers when inferring PPIs [10, 75]. However, this method may not be the most optimal approach to evaluate the predictive performance of classifiers when inferring PPIs. The study by Browne et al. [85] has demonstrated that the additional application of other assessment techniques such as partial AUC values from ROC curves, TP/FP rates, and sensitivity could be employed as comparative predictive measures to the ROC curve approach when evaluating the classification performance for the predictive task of PPI inference.

The computational inference of PPI networks is still a relatively new research area. Future research in inferring PPI networks may be performed in the areas including the recovery of PPIs between proteins [80, 88], identification of protein complexes [23, 91, 92], investigating network topology of PPI networks [67], defining and modelling pathways (for instance, signalling and metabolic pathways) [93].

References

[1] B. Alberts, *Essential Cell Biology: An Introduction to the Molecular Biology of the Cell*, Garland, New York, NY, USA, 1998.

[2] M. S. Scott and G. J. Barton, "Probabilistic prediction and ranking of human protein-protein interactions," *BMC Bioinformatics*, vol. 8, article 239, 2007.

[3] C. von Mering, R. Krause, B. Snel, et al., "Comparative assessment of large-scale data sets of protein-protein interactions," *Nature*, vol. 417, no. 6887, pp. 399–403, 2002.

[4] E. M. Phizicky and S. Fields, "Protein-protein interactions: methods for detection and analysis," *Microbiological Reviews*, vol. 59, no. 1, pp. 94–123, 1995.

[5] P. Bork, L. J. Jensen, C. von Mering, A. K. Ramani, I. Lee, and E. M. Marcotte, "Protein interaction networks from yeast to human," *Current Opinion in Structural Biology*, vol. 14, no. 3, pp. 292–299, 2004.

[6] J. Yu and F. Fotouhi, "Computational approaches for predicting protein-protein interactions: a survey," *Journal of Medical Systems*, vol. 30, no. 1, pp. 39–44, 2006.

[7] G. T. Hart, A. K. Ramani, and E. M. Marcotte, "How complete are current yeast and human protein-interaction networks?" *Genome Biology*, vol. 7, no. 11, article 120, 2006.

[8] R. Jansen and M. Gerstein, "Analyzing protein function on a genomic scale: the importance of gold-standard positives and negatives for network prediction," *Current Opinion in Microbiology*, vol. 7, no. 5, pp. 535–545, 2004.

[9] N. Lin, B. Wu, R. Jansen, M. Gerstein, and H. Zhao, "Information assessment on predicting protein-protein interactions," *BMC Bioinformatics*, vol. 5, article 154, 2004.

[10] L. J. Lu, Y. Xia, A. Paccanaro, H. Yu, and M. Gerstein, "Assessing the limits of genomic data integration for predicting protein networks," *Genome Research*, vol. 15, no. 7, pp. 945–953, 2005.

[11] R. Jansen, H. Yu, D. Greenbaum, et al., "A Bayesian networks approach for predicting protein-protein interactions from genomic data," *Science*, vol. 302, no. 5644, pp. 449–453, 2003.

[12] C. L. Myers and O. G. Troyanskaya, "Context-sensitive data integration and prediction of biological networks," *Bioinformatics*, vol. 23, no. 17, pp. 2322–2330, 2007.

[13] O. G. Troyanskaya, K. Dolinski, A. B. Owen, R. B. Altman, and D. Botstein, "A Bayesian framework for combining heterogeneous data sources for gene function prediction (in Saccharomyces cerevisiae)," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 14, pp. 8348–8353, 2003.

[14] J. S. Bader, A. Chaudhuri, J. M. Rothberg, and J. Chant, "Gaining confidence in high-throughput protein interaction networks," *Nature Biotechnology*, vol. 22, no. 1, pp. 78–85, 2004.

[15] I. Xenarios, L. Salwínski, X. J. Duan, P. Higney, S.-M. Kim, and D. Eisenberg, "DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions," *Nucleic Acids Research*, vol. 30, no. 1, pp. 303–305, 2002.

[16] Y. Xia, L. J. Lu, and M. Gerstein, "Integrated prediction of the helical membrane protein interactome in yeast," *Journal of Molecular Biology*, vol. 357, no. 1, pp. 339–349, 2006.

[17] Y. Qi, Z. Bar-Joseph, and J. Klein-Seetharaman, "Evaluation of different biological data and computational classification methods for use in protein interaction prediction," *Proteins: Structure, Function, and Bioinformatics*, vol. 63, no. 3, pp. 490–500, 2006.

[18] Y. Yamanishi, J.-P. Vert, and M. Kanehisa, "Protein network inference from multiple genomic data: a supervised approach," *Bioinformatics*, vol. 20, supplement 1, pp. i363–i370, 2004.

[19] I. Lee, S. V. Date, A. T. Adai, and E. M. Marcotte, "A probabilistic functional network of yeast genes," *Science*, vol. 306, no. 5701, pp. 1555–1558, 2004.

[20] A. Ghavidel, G. Cagney, and A. Emili, "A skeleton of the human protein interactome," *Cell*, vol. 122, no. 6, pp. 830–832, 2005.

[21] Z. Barutcuoglu, R. E. Schapire, and O. G. Troyanskaya, "Hierarchical multi-label prediction of gene function," *Bioinformatics*, vol. 22, no. 7, pp. 830–836, 2006.

[22] C. L. Myers, D. Robson, A. Wible, et al., "Discovery of biological networks from diverse functional genomic data," *Genome Biology*, vol. 6, no. 13, article R114, 2005.

[23] S. Asthana, O. D. King, F. D. Gibbons, and F. P. Roth, "Predicting protein complex membership using probabilistic network reliability," *Genome Research*, vol. 14, no. 6, pp. 1170–1175, 2004.

[24] C. L. Myers, D. R. Barrett, M. A. Hibbs, C. Huttenhower, and O. G. Troyanskaya, "Finding function: evaluation methods for functional genomic data," *BMC Genomics*, vol. 7, article 187, 2006.

[25] H. W. Mewes, C. Amid, R. Arnold, et al., "MIPS: analysis and annotation of proteins from whole genomes," *Nucleic Acids Research*, vol. 32, pp. D41–D44, 2004.

[26] A.-C. Gavin, P. Aloy, P. Grandi, et al., "Proteome survey reveals modularity of the yeast cell machinery," *Nature*, vol. 440, no. 7084, pp. 631–636, 2006.

[27] N. J. Krogan, G. Cagney, H. Yu, et al., "Global landscape of protein complexes in the yeast Saccharomyces cerevisiae," *Nature*, vol. 440, no. 7084, pp. 637–643, 2006.

[28] M.-H. Kuo and C. D. Allis, "In vivo cross-linking and immunoprecipitation for studying dynamic protein: DNA associations in a chromatin environment," *Methods*, vol. 19, no. 3, pp. 425–433, 1999.

[29] J. Homola, S. S. Yee, and G. Gauglitz, "Surface plasmon resonance sensors: review," *Sensors & Actuators, B*, vol. 54, no. 1, pp. 3–15, 1999.

[30] A. Moser and C. Detellier, "Nuclear magnetic resonance spectroscopy," in *Encyclopedia of Supramolecular Chemistry*, Marcel Dekker, New York, NY, USA, 2004.

[31] Y. Wu, Q. Li, and X. Z. Chen, "Detecting protein-protein interactions by far western blotting," *Nature Protocols*, vol. 2, no. 12, pp. 3278–3284, 2007.

[32] K. D. Pfleger and K. A. Eidne, "Illuminating insights into protein-protein interactions using bioluminescence resonance energy transfer (BRET)," *Nature Methods*, vol. 3, no. 3, pp. 165–174, 2006.

[33] P. Uetz, L. Glot, G. Cagney, et al., "A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae," *Nature*, vol. 403, no. 6770, pp. 623–627, 2000.

[34] G. D. Bader and C. W. V. Hogue, "Analyzing yeast protein-protein interaction data obtained from different sources," *Nature Biotechnology*, vol. 20, no. 10, pp. 991–997, 2002.

[35] R. Jansen, D. Greenbaum, and M. Gerstein, "Relating whole-genome expression data with protein-protein interactions," *Genome Research*, vol. 12, no. 1, pp. 37–46, 2002.

[36] J. M. Stuart, E. Segal, D. Koller, and S. K. Kim, "A gene-coexpression network for global discovery of conserved genetic modules," *Science*, vol. 302, no. 5643, pp. 249–255, 2003.

[37] D. T. Suzuki, A. J. F. Griffiths, and R. C. Lewontin, *An Introduction to Genetic Analysis*, WH Freeman, New York, NY, USA, 7th edition, 2000.

[38] T. Ito, T. Chiba, R. Ozawa, M. Yoshida, M. Hattori, and Y. Sakaki, "A comprehensive two-hybrid analysis to explore the yeast protein interactome," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 8, pp. 4569–4574, 2001.

[39] T. Ito, K. Tashiro, S. Muta, et al., "Toward a protein-protein interaction map of the budding yeast: a comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 3, pp. 1143–1147, 2000.

[40] B. A. Shoemaker and A. R. Panchenko, "Deciphering protein-protein interactions—part I: experimental techniques and databases," *PLoS Computational Biology*, vol. 3, no. 3, article e42, 2007.

[41] L. Zhou, X. Ma, and F. Sun, "The effects of protein interactions, gene essentiality and regulatory regions on expression variation," *BMC Systems Biology*, vol. 2, article 54, 2008.

[42] O. G. Troyanskaya, "Putting microarrays in a context: integrated analysis of diverse biological data," *Briefings in Bioinformatics*, vol. 6, no. 1, pp. 34–43, 2005.

[43] H. B. Fraser, A. E. Hirsh, D. P. Wall, and M. B. Eisen, "Coevolution of gene expression among interacting proteins," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 24, pp. 9033–9038, 2004.

[44] A.-C. Gavin, M. Bösche, R. Krause, et al., "Functional organization of the yeast proteome by systematic analysis of protein complexes," *Nature*, vol. 415, no. 6868, pp. 141–147, 2002.

[45] M. Middendorf, E. Ziv, and C. H. Wiggins, "Inferring network mechanisms: the *Drosophila melanogaster* protein interaction network," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 9, pp. 3192–3197, 2005.

[46] L. Giot, J. S. Bader, C. Brouwer, et al., "A protein interaction map of *Drosophila melanogaster*," *Science*, vol. 302, no. 5651, pp. 1727–1736, 2003.

[47] W. Zhong and P. W. Sternberg, "Genome-wide prediction of C. elegans genetic interactions," *Science*, vol. 311, no. 5766, pp. 1481–1484, 2006.

[48] S. Li, C. M. Armstrong, N. Bertin, et al., "A map of the interactome network of the metazoan C. elegans," *Science*, vol. 303, no. 5657, pp. 540–543, 2004.

[49] A. J. Walhout, R. Sordella, X. Lu, et al., "Protein interaction mapping in C. elegans using proteins involved in vulval development," *Science*, vol. 287, no. 5450, pp. 116–122, 2000.

[50] R. M. Ewing, P. Chu, F. Elisma, et al., "Large-scale mapping of human protein-protein interactions by mass spectrometry," *Molecular Systems Biology*, vol. 3, article 89, 2007.

[51] J. Xu and Y. Li, "Discovering disease-genes by topological features in human protein-protein interaction network," *Bioinformatics*, vol. 22, no. 22, pp. 2800–2805, 2006.

[52] H. Yu, P. Braun, M. A. Yildirim, et al., "High-quality binary protein interaction map of the yeast interactome network," *Science*, vol. 322, no. 5898, pp. 104–110, 2008.

[53] Y. Ho, A. Gruhler, A. Heilbut, et al., "Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry," *Nature*, vol. 415, no. 6868, pp. 180–183, 2002.

[54] B. Schwikowski, P. Uetz, and S. Fields, "A network of protein-protein interactions in yeast," *Nature Biotechnology*, vol. 18, no. 12, pp. 1257–1261, 2000.

[55] B. Lehner and A. G. Fraser, "A first-draft human protein-interaction map," *Genome Biology*, vol. 5, no. 9, article R63, 2004.

[56] R. Bunescu, R. Ge, R. J. Kate, et al., "Comparative experiments on learning information extractors for proteins and their interactions," *Artificial Intelligence in Medicine*, vol. 33, no. 2, pp. 139–155, 2005.

[57] E. Formstecher, S. Aresta, V. Collura, et al., "Protein interaction mapping: a Drosophila case study," *Genome Research*, vol. 15, no. 3, pp. 376–384, 2005.

[58] R. Jansen, N. Lan, J. Qian, and M. Gerstein, "Integration of genomic datasets to predict protein complexes in yeast," *Journal of Structural and Functional Genomics*, vol. 2, no. 2, pp. 71–81, 2002.

[59] Y. Qi, J. Klein-Seetharaman, and Z. Bar-Joseph, "A mixture of feature experts approach for protein-protein interaction prediction," *BMC Bioinformatics*, vol. 8, supplement 10, article S6, 2007.

[60] D. R. Rhodes, S. A. Tomlins, S. Varambally, et al., "Probabilistic model of the human protein-protein interaction network," *Nature Biotechnology*, vol. 23, no. 8, pp. 951–959, 2005.

[61] R. J. Cho, M. J. Campbell, E. A. Winzeler, et al., "A genome-wide transcriptional analysis of the mitotic cell cycle," *Molecular Cell*, vol. 2, no. 1, pp. 65–73, 1998.

[62] D. Greenbaum, R. Jansen, and M. Gerstein, "Analysis of mRNA expression and protein abundance data: an approach for the comparison of the enrichment of features in the cellular population of proteins and transcripts," *Bioinformatics*, vol. 18, no. 4, pp. 585–596, 2002.

[63] S. Peri, J. D. Navarro, R. Amanchy, et al., "Development of human protein reference database as an initial platform for approaching systems biology in humans," *Genome Research*, vol. 13, no. 10, pp. 2363–2371, 2003.

[64] C.-S. Goh and F. E. Cohen, "Co-evolutionary analysis reveals insights into protein-protein interactions," *Journal of Molecular Biology*, vol. 324, no. 1, pp. 177–192, 2002.

[65] M. Pellegrini, E. M. Marcotte, M. J. Thompson, D. Eisenberg, and T. O. Yeates, "Assigning protein functions by comparative genome analysis: protein phylogenetic profiles," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 8, pp. 4285–4288, 1999.

[66] H. Yu, N. M. Luscombe, H. X. Lu, et al., "Annotation transfer between genomes: protein-protein interrologs and protein-DNA regulogs," *Genome Research*, vol. 14, no. 6, pp. 1107–1118, 2004.

[67] H. N. Chua, W.-K. Sung, and L. Wong, "Exploiting indirect neighbours and topological weight to predict protein function from protein-protein interactions," *Bioinformatics*, vol. 22, no. 13, pp. 1623–1630, 2006.

[68] B.-J. Breitkreutz, C. Stark, T. Reguly, et al., "The BioGRID interaction database: 2008 update," *Nucleic Acids Research*, vol. 36, database issue, pp. D637–D640, 2008.

[69] H. Hermjakob, L. Montecchi-Palazzi, C. Lewington, et al., "IntAct: an open source molecular interaction database," *Nucleic Acids Research*, vol. 32, pp. D452–D455, 2004.

[70] A. Chatr-aryamontri, A. Ceol, L. M. Palazzi, et al., "MINT: the molecular interaction database," *Nucleic Acids Research*, vol. 35, database issue, pp. D572–D574, 2007.

[71] A. Ben-Hur and W. S. Noble, "Choosing negative examples for the prediction of protein-protein interactions," *BMC Bioinformatics*, vol. 7, supplement 1, 2006.

[72] J. Guo, X. Wu, D.-Y. Zhang, and K. Lin, "Genome-wide inference of protein interaction sites: lessons from the yeast high-quality negative protein-protein interaction dataset," *Nucleic Acids Research*, vol. 36, no. 6, pp. 2002–2011, 2008.

[73] F. Browne, H. Wang, H. Zheng, and F. Azuaje, "GRIP: a web-based system for constructing Gold Standard datasets for protein-protein interaction prediction," *Source Code for Biology and Medicine*, vol. 4, article 2, 2009.

[74] F. Browne, H. Wang, H. Zheng, and F. Azuaje, "Supervised statistical and machine learning approaches to inferring pairwise and module-based protein interaction networks," in *Proceedings of the 7th IEEE International Conference on Bioinformatics and Bioengineering (BIBE '07)*, pp. 1365–1369, 2007.

[75] F. Browne, H. Wang, H. Zheng, and F. Azuaje, "An assessment of machine and statistical learning approaches to inferring networks of protein-protein interactions," *Journal of Integrative Bioinformatics*, vol. 3, 2006.

[76] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, John Wiley & Sons, New York, NY, USA, 2nd edition, 2000.

[77] S. L. Lo, C. Z. Cai, Y. Z. Chen, and M. C. M. Chung, "Effect of training datasets on support vector machine prediction of protein-protein interactions," *Proteomics*, vol. 5, no. 4, pp. 876–884, 2005.

[78] Y. Qi, J. Klein-Seetharaman, and Z. Bar-Joseph, "Random forest similarity for protein-protein interaction prediction from multiple sources," *Pacific Symposium on Biocomputing*, pp. 531–542, 2005.

[79] X.-W. Chen and M. Liu, "Prediction of protein-protein interactions using random decision forest framework," *Bioinformatics*, vol. 21, no. 24, pp. 4394–4400, 2005.

[80] Z. Ma, C. Zhou, L. Lu, Y. Ma, P. Sun, and Y. Cui, "Predicting protein-protein interactions based on BP neural network," in *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW '07)*, pp. 3–7, 2007.

[81] E. Keedwell and A. Narayanan, "Discovering gene networks with a neural-genetic hybrid," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 2, no. 3, pp. 231–242, 2005.

[82] P. Fariselli, A. Zauli, M. Finelli, P. Martelli, and R. Casadio, "A neural network method to improve prediction of protein-protein interaction sites in heterocomplexes," in *Proceedings of the 13th IEEE Workshop on Neural Networks for Signal Processing (NNSP '03)*, pp. 33–41, September 2003.

[83] F. M. Ausubel, R. Brent, R. Kingston, et al., *Current Protocols in Molecular Biology*, vol. 3, John Wiley & Sons, New York, NY, USA, 2008.

[84] M. Ashburner, C. Ball, and J. Blake, "Gene ontology: tool for the unification of biology. The gene ontology consortium database resources of the national center for biotechnology information," *Nucleic Acids Research*, vol. 34, 2006.

[85] F. Browne, H. Wang, H. Zheng, and F. Azuaje, "Reassessing the genomic data integration limits for the prediction of protein-protein interactions in saccharomyces cerevisiae," in *Proceedings of the IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology*, pp. 28–35, 2008.

[86] K. Xia, D. Dong, and J.-D. J. Han, "IntNetDB v1.0: an integrated protein-protein interaction network database generated by a probabilistic model," *BMC Bioinformatics*, vol. 7, article 508, 2006.

[87] R. J. P. van Berlo, L. F. A. Wessels, D. de Ridder, and M. J. T. Reinders, "Protein complex prediction using an integrative bioinformatics approach," *Journal of Bioinformatics and Computational Biology*, vol. 5, no. 4, pp. 839–864, 2007.

[88] P. Prusis, S. Uhlen, R. Petrovska, M. Lapinsh, and J. E. S. Wikberg, "Prediction of indirect interactions in proteins," *BMC Bioinformatics*, vol. 7, article 167, 2006.

[89] S. Grosdidier and J. Fernández-Recio, "Identification of hot-spot residues in protein-protein interactions by computational docking," *BMC Bioinformatics*, vol. 9, article 447, 2008.

[90] S. R. Collins, P. Kemmeren, X.-C. Zhao, et al., "Toward a comprehensive atlas of the physical interactome of Saccharomyces cerevisiae," *Molecular and Cellular Proteomics*, vol. 6, no. 3, pp. 439–450, 2007.

[91] H. Zheng, H. Wang, and D. H. Glass, "Integration of genomic data for inferring protein complexes from global protein-protein interaction networks," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 38, no. 1, pp. 5–16, 2008.

[92] V. Spirin and L. A. Mirny, "Protein complexes and functional modules in molecular networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 21, pp. 12123–12128, 2003.

[93] R. Albert, "Scale-free networks in cell biology," *Journal of Cell Science*, vol. 118, no. 21, pp. 4947–4957, 2005.