Jan Romportl
Pavel Ircing
Eva Žáčková
Radek Schuster
Michal Polák (eds.)

Beyond AI: Interdisciplinary Aspects of Artificial Intelligence

Proceedings of Extended Abstracts Presented at The International Conference Beyond AI 2011
Pilsen, Czech Republic, December 8–9, 2011

Beyond AI: Interdisciplinary Aspects of Artificial Intelligence
Proceedings of Extended Abstracts

Editors:
Jan Romportl, Pavel Ircing, Eva Žáčková, Radek Schuster, Michal Polák
Typesetting: Pavel Ircing, Michal Polák
Cover design: Tom Vild
Printed: Typos, tiskařské závody, s.r.o.

Published by University of West Bohemia, Pilsen, 2011

Preface

Dear participants,

on behalf of the Organising Committee, I would like to welcome you to the international conference *Beyond AI: Interdisciplinary Aspects of Artificial Intelligence.*

The conference aims to shed light on philosophical, ethical, social, esthetical, medical and also technical issues that constitute current areas of AI. By supporting discussions of these topics we try to offer you an interesting opportunity to constitute an interdisciplinary platform for reaching new perspectives of edge effects in AI. We strongly believe that interdisciplinary dialogue between experts from engineering, natural sciences and humanities can be helpful for understanding ourselves as human beings closely related to the AI products and applications.

We are grateful to a number of institutions without whose help this conference would not have been possible. The European Social Fund in the Czech Republic, Ministry of Education, Youth and Sports, University of West Bohemia and its New Technologies Research Centre, Department of Cybernetics and Department of Philosophy are some of the most important sources. Last but not least, we would like to thank all those people that were helpful in organising the conference Beyond AI.

We hope you will spend stimulating time at the conference as well as in the city of Pilsen.

November 2011

Jan Romportl
Organising Committee Chair
BAI 2011

Organisation

Beyond AI 2011 was organized by the Department of Interdisciplinary Activities, New Technologies Research Centre, University of West Bohemia (UWB), Plzen, Czech Republic, in close cooperation with the Department of Cybernetics, Faculty of Applied Sciences, UWB, and the Department of Philosophy, Faculty of Philosophy and Arts, UWB. The conference web page is located at:

http://beyondai.zcu.cz

Programme Committee

Jiří Beran *(Charles University, Czech Republic)*
Jan Betka *(Charles University, Czech Republic)*
Nick Campbell *(Trinity College, Ireland)*
Luis Hernández Gómez *(Technical University of Madrid, Spain)*
Ivan M. Havel *(Centre for Theoretical Study, Czech Republic)*
Søren Holm *(University of Manchester, UK)*
Jozef Kelemen *(Silesian University, Czech Republic)*
Vladimír Mařík *(Czech Technical University, Czech Republic)*
Peter Mikulecký *(University of Hradec Králové, Czech Republic)*
Josef Psutka *(University of West Bohemia, Czech Republic)*
Raúl Santos de la Cámara *(Telefónica I+D, Spain)*
Yorick Wilks *(Florida Institute for Human & Machine Cognition, US)*
Enrico Zovato *(Loquendo, Italy)*

Organising Committee

Jan Romportl *(Chair)*
Pavel Ircing
Eva Žáčková
Radek Schuster
Michal Polák

Sponsoring

The conference is supported by the project "Interdisciplinary Partnership for Artificial Intelligence" (CZ.1.07/2.4.00/17.0055), OP Education for Competitiveness, funded by the European Social Fund in the Czech Republic and by the State Budget of the Czech Republic.

Table of Contents

Families of Binary Operations in Fuzzy Set Theory

János Fodor

Institute of Intelligent Engineering Systems, Óbuda University
Bécsi út 96/b, H-1034 Budapest, Hungary
fodor@uni-obuda.hu

Abstract. Our world is rather uncertain. However, this does not mean the exclusive presence of randomness. On the contrary, most concepts are "fuzzy", containing inherent linguistic uncertainty (i.e., imprecise description of concepts such as low price, young people, tall men) or informational uncertainty (caused by missing or incomplete information). In other words, fuzziness refers to nonstatistical imprecision, approximation and vagueness in information and data.

Keywords: fuzzy set theory, triangular norms and conorms

Contrary to classical set theory (which is based on the two-valued (Boolean) logic) in which an element either *is* or *is not* a member of a set, in fuzzy set theory (which is based on fuzzy logics) membership values reflect the *membership grades* of the elements in the set [21]. Set-theoretic operations on classical (also called crisp) sets (like intersection, union and complement) are naturally extended to fuzzy sets by using *interpretations* of logic connectives $\wedge$, $\vee$ and $\neg$, respectively [18]. It is assumed that the conjunction $\wedge$ is interpreted by a *triangular norm* (*t-norm* for short), the disjunction $\vee$ is interpreted by a *triangular conorm* (shortly: *t-conorm*), and the negation $\neg$ by a *strong negation*.

The aim of this talk is twofold. First, we would like to give a short overview of basic connectives used in fuzzy set theory and fuzzy logic. Second, we point out some of our contributions to the development of the field.

In the first part we summarize the basics of triangular norms and conorms (their axioms, elementary properties, prototypical examples, their classification, and their link to functional equations) [1]. We also outline important facts on fuzzy implications.

In the second part we intend to touch upon the following subjects:

– left-continuous t-norms and their role in fuzzy logics [2, 6–8, 15, 16];

- extensions of t-norms and t-conorms [3, 4, 13, 17, 19, 20];
- some functional equations for t-norms [5, 9–12].

The following list of references may help the interested reader to find more details and further information about the mentioned subjects.

References

1. Aczél, J.: *Lectures on Functional Equations and their Applications.* Academic Press, New York (1966)
2. De Baets, B., Esteva, F., Fodor, J. and Godo, L.: Systems of ordinal fuzzy logic with application to preference modelling. *Fuzzy Sets and Systems* 124, 353–359 (2001)
3. Calvo, T., De Baets, B., Fodor, J.: The functional equations of Frank and Alsina for uninorms and nullnorms. *Fuzzy Sets and Systems* 120, 385–394 (2001)
4. Dombi, J.: Basic concepts for the theory of evaluation: the aggregative operator. *European J. Oper. Res.* 10, 282–293 (1982)
5. Durante, F. and Sarkoci, P.: A note on the convex combinations of triangular norms. *Fuzzy Sets and Systems* 159, 77–80 (2008)
6. Esteva, F. and Godo, L.: Monoidal t-norm based logic: towards a logic for left-continuous t-norms. *Fuzzy Sets and Systems* 124, 271–288 (2001)
7. Fodor, J.C.: A new look at fuzzy connectives. *Fuzzy Sets and Systems* 57, 141–148 (1993)
8. Fodor, J.C.: Contrapositive symmetry of fuzzy implications. *Fuzzy Sets and Systems* 69, 141–156 (1995)
9. Fodor, J., Klement, E.P. and Mesiar, R.: Cross-migrative triangular norms. (submitted)(2011).
10. Fodor J. and Rudas, I.J.: On continuous triangular norms that are migrative. *Fuzzy Sets and Systems* 158, 1692 – 1697 (2007)
11. Fodor, J. and Rudas, I. J.: An extension of the migrative property for triangular norms. *Fuzzy Sets and Systems* 168, 70–80 (2011)
12. Fodor, J. and Rudas, I.J.: Migrative t-norms with respect to continuous ordinal sums. *Information Sciences* 181, 4860–4866 (2011)
13. Fodor, J.C., Yager, R.R., Rybalov, A.: Structure of uninorms. *Int. J. Uncertainty Fuzziness Knowledge-based Systems* 5, 411–427 (1997)
14. Grabisch, M., De Baets, B. and Fodor, J.: The quest for rings on bipolar scales. *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems* 12, 499–512 (2004)
15. Jenei, S.: Geometry of left-continuous t-norms with strong induced negations. *Belg. J. Oper. Res. Statist. Comput. Sci.* 38, 5–16 (1998)
16. Jenei, S.: Structure of left-continuous t-norms with strong induced negations. (I) Rotation construction, *J. Appl. Non-Classical Logics* 10, 83–92 (2000)

17. Klement, E.P., Mesiar, R. and Pap, E.: On the relationship of associative compensatory operators to triangular norms and conorms. *Internat. J. Uncertain. Fuzziness Knowledge-Based Systems* 4, 129–144 (1996)
18. Klement, E.P., Mesiar, R. and Pap, E.: *Triangular Norms*. Kluwer Academic Publishers, Dordrecht (2000)
19. Mas, M., Mayor, G., Torrens, J.: t-operators. *Internat. J. Uncertainty Fuzziness Knowledge-Based Systems* 7, 31–50 (1999)
20. Yager, R.R., Rybalov, A.: Uninorm aggregation operators. *Fuzzy Sets and Systems* 80, 111–120 (1996).
21. Zadeh, L.A.: Fuzzy sets. *Inform. Control* 8, 338–353 (1965)

Beyond AI is HAI

Matjaž Gams

Jozef Stefan Institute
Ljubljana, Slovenia
matjaz.gams@ijs.si

Abstract. In this 4-page short paper related to the invited lecture, general concepts of intelligence will be first analyzed and human intelligence next. Human progress will be examined. How did the intelligence transfer us in the past and where will it bring us in the future? The principle and paradox of multiple knowledge will be presented. In the last part, analyses of a specific real-life issues will be studied and an overview of projects at our department will be addressed, in particular results of an FP7 project Confidence helping elderly.

Keywords: artificial intelligence, human intelligence, multiple knowledge

1 Introduction

What is intelligence? According to Wikipedia, intelligence "has been defined in different ways, including the abilities for abstract thought, understanding, communication, reasoning, learning, planning, emotional intelligence and problem solving."

"Intelligence is most widely studied in humans, but has also been observed in ani-mals and plants."

"Numerous definitions of and hypotheses about intelligence have been proposed since before the twentieth century, with no consensus reached by scholars."

In this paper we propose a simple definition of intelligence based on the ability to learn and analyze it in different contexts, in particular in relations to the Turing test [1]. Then, we will study human intelligence and mind in an unusual way, including questions how beneficial intelligence is in relations to the human predecessor and if it is indeed a uniform property equally distributed throughout the population.

In the following section, the principle and paradox of multiple knowledge [2] will be presented, introducing a viewpoint similar to Minsky's or

Edelman's. A hypothesis is proposed that this computational property is essential for human-level artificial intelligence.

In modern times, intelligence is beneficial to humans beyond any doubt, but it is not clear if we humans indeed decide well at the level of nations or the whole civilization. Analyses in relation to the demographic issues, natural resources in particular and the social order will be presented.

Next, a couple of projects of the Department of the intelligent systems will be presented with extended presentation of the FP7 project Confidence helping elderly.

Finally, discussion will elaborate issues addressed in the paper.

2 Intelligence

Intelligence can be *human*, *natural*, *engineering* and *artificial*. Human intelligence is perhaps best known and discussed among broad population. Natural intelligence refers to all live beings, be it human, animal or of plants. Artificial intelligence [5] is about non-alive entities, typically machines and computers. Alien intelligence refers to alien beings, but this and other types of intelligence will not be discussed in this paper.

According to Wikipedia, "*Artificial intelligence* (AI) is the intelligence of machines and the branch of computer science that aims to create it. AI textbooks define the field as the study and design of intelligent agents."

An *intelligent system* is the one that learns, having additional properties like flexibility, adaptability or being user-friendly. Engineering intelligence is the one referring to the intelligence of products not being intelligent in the usual way, yet they exhibit some property not existing in the previous product similar to some of the properties of intelligent systems, be it flexibility or simplicity of use.

Definition of intelligence: An entity that learns is intelligent.

An important concept when determining intelligence is the *Turing test (TT)* where an interrogator can perform free tests on two entities, one being a human and the other a computer pretending to be human. If the interrogator cannot conclude in five minutes with sufficient certainty, which of the two entities is computer, the computer passed the test. The basic Turing test or TT is performed with technologically limited communication only through a keypad. If other sources of information are allowed, then this is the Total Turing test or TTT. For a computer to pass TTT, it must be in the form of a human-like android. Another version is TTTT - it denotes TTT performed over a long period of time, e.g. consisting of observing several generations of that species.

Since the Turing test serves only determining whether computers can imitate hu-mans, real-life testing of various species consists of tests that demand certain level of intelligence, for example, IQ tests are often used for humans. With these type of tests, for example, it is possible to compare various version of species or subspecies and individuals inside each (sub)species. For dogs, for example, lists of intelligence of species are widely published over the internet and everybody can test his/hers hairy friend.

Since even simple plants or primitive beings like the HIV virus learn, they are intelligent according to our definition. Their learning is not conscious like in humans, but through evolution. HIV virus, for example, creates virus generations till they penetrate the human defense, which happens in around a decade. Accepting this basic definition of intelligence, stemming from the definition of intelligent systems, one must conclude that all alive being, be it plants or humans are intelligent.

What about an intelligent agent? According to our definitions, as soon as an entity learns it is an intelligent system and is therefore intelligent. Similarly, all AI systems and machine learning systems in particular, are intelligent.

3 Human Intelligence

An intriguing question is why intelligence emerged at all in humans. It is evident that it did and that it corresponds to the anthropological principle, but how and why? Intelligence is not the magic solution as one might suspect. For example, compare the common vacuum cleaner Roomba IRobot that moves more or less random with the intelligent version that first wonders around measuring the room and obstacles and then intelligently vacuuming it. Most people buy the stupid, agent version, instead of the intelligent agent version. Intelligence is not an important asset at vacuum cleaners.

It might be similar with our predecessors [3]. Since the brain consumes around one third of an average human, in the case of a famine, the smarter humans died sooner unless their additional intelligence enabled them to find more food. It is the trade-off then, and indeed evolution tested several versions of our predecessors, some being stronger, some smaller, other more physically human-like. Most of these versions died out like versions of the HIV virus. One curiosity - till around 18.000 years ago or even much less, dwarf humans of 1m adult size lived on a Pacific island.

In terms of human intelligence, an interesting question is distribution between races and sex. This is a very sensitive question, leading to several hate-movements including for example racism, but being technical scientists not prone to ideologies we can examine those distributions. The results indicate that there are measurable differences in intelligence between sexes, ages, health, social status, religion, or even users of different browsers. Most of these differences are not of any major importance.

In addition to the human-type-related issues of intelligence there are also tools and technology-related issues of intelligence. For example, the Flynn effect stating that human intelligence grows in recent decades, has been empirically measured and is in contradiction with the fact that less educated have more children and as such, general average intelligence should fall. It seems that the more important factor is the tools and technology effect. Technology in essence provides better tools, just at more advanced level as simple tools. For example, our predecessor starting to use a stick to dig carrots, managed to get more food with less effort and as a consequence, his hand muscles decreased a bit. Similarly, humans using a calculator have a bit less developed the calculating part of the brain, but they can solve more. Humans using mobile phones and/or the internet use these tools to improve their performance and as a consequence probably also their intelligence. This might also explain the Flynn effect.

If the tool theory holds, we might be on a way to further advances, even significant jumps ahead [4]. Indeed, a human using a computer and the web can solve significantly more complex problems and perform much better in every-day tasks. Are we still just humans or humans with the web and intelligent services on it?

4 The Principle of Multiple Knowledge

The principle and paradox of multiple knowledge [2] exist in the weak and strong version. In the weak version, the principle states that better results will be obtained if several systems will be designed and sensibly combined together, compared to the best of those single systems.

The weak version of paradox claims that the integrated system is basically one system. In the strong version, the principle of multiple knowledge (PMK) states that human-level intelligence can be achieved iff the thinking processes are multiple, be it in humans or machines. Similarly to the weak paradox of multiple knowledge, the strong version claims that multiple processes can be executed as a single one.

The principle is presented by a combination of universal Turing machines and supported by six analyses: formal, simulation, Turing machines, empirical studies, human anthropological studies, physical theory of many worlds. For the most farfetched thesis, consider many worlds that emerge whenever a new cognitive decision is made, based on the interpretation of quantum mechanics. In a similar way, the thinking process multiplies at any decision point. However, an essential part of PMK is that the thinking processes interact with each other. This interaction and the final integration into one decision is still an open research question.

5 Analyses and Projects

This section will be presented in more detail in the lecture and another, extended paper. Demography, oil trends and the predominant social order will be analyzed.

Several department international and national projects will be presented. In particular, the results of Confidence FP7 project will be demonstrated.

6 Discussion

Beyond Artificial intelligence is Human Artificial intelligence - humans and computers integrated into a superior being. It is happening already, promising a new era in human civilization.

References

1. Moor, J.H. (ed.): The Turing Test; The Elusive Standard of Artificial Intelligence. Kluwer, Dordrecht (2003)
2. Gams, M.: Weak intelligence: Through the principle and paradox of multiple knowledge. *Advances in Computation: Theory and practice*, 6 (2001)
3. Gross, R.D.: Psychology: The Science of Mind and Behaviour. Hodder Arnold (2010)
4. Kurzweil, R.: The Age of Spiritual Machines: How We Will Live, Work, and Think in the New Age of Intelligent Machines. Penguin books, London (2000)
5. Russel, S., Norvig, P.: Artificial intelligence A Modern Approach. Prentice Hall, New Jersey (2009)

Beyond Knowledge Systems

Jozef Kelemen[1,2]

[1] Research Institute of the IT4Innovations Centre of Excellence and Institute of Computer Science, Silesian University
746 01 Opava, Czech Republic
[2] School of Management
851 04 Bratislava, Slovakia
kelemen@fpf.slu.cz

Abstract. The central subject of this contribution is to analyze the step-by-step process leading in the run of times from data-based to knowledge-based systems, and to relate knowledge systems to the structure and roles of human wisdom. In order to do this we formulate our main question: What we will see by focusing our sight beyond the knowledge systems? Is it the computational wisdom?

Keywords: AI and mind

1 Introduction

During the past few decades Artificial Intelligence (AI) and Cognitive Science as highly related disciplines led to numerous applicable results. One (perhaps the most important) among them is the whole field of knowledge-based systems. *Knowledge-based* (or shortly *knowledge*) *systems* are systems which include a *representation* of human knowledge on a part of a human expertise, and also *inferring procedures* for the use of represented knowledge in order to *solve problems* the solving of which usually require knowledge and skills of human experts.

Connected with the success and the large applicability of knowledge systems a question arises concerning the bounds of their applicability, and concerning their impact to the life of human individuals and human society in the near future. These questions are, among other reasons, evoked also by realization of the relatively important changes in the level of education of human beings in the present time in the remarkable decrease of the level of education. To be more precise, the specialists record not only the decrease in the level of professional preparedness of the young people in their particular disciplines, in their domain of expertise, but also in their general preparedness for participation in formulation and solution of more general problems, problems which require human *wisdom*.

The central subject of this contribution is to analyze the step-by-step process leading from data-based to knowledge-based systems, and to relate knowledge systems to the structure and roles of human wisdom. In order to do this we formulate our main question: What we will see by focusing our sight beyond the knowledge systems?

2 From Data to Knowledge – Role of the Context

Let us, first of all, to emphasize the important role of context in the development of the use of computers as primarily data processing devices.

The story started tens of years ago when the first computers have been constructed as suitable technical devices for *computing with numbers*, as machines for "number crunching". The input data interpreted as numbers of different formats have been "crunched" by suitable programs and the output data was interpreted again as numbers. The context has been in these cases defined by the computer programs.

The similar situation, but in the more explicit form appeared with data of more general meaning and led to the development of special systems of computer programs called *data bases*.

Putting an item into the data base means to establish the relations of it to other data, so to define a context in which the implanted datum will be interpreted by the given particular data base system. In other worlds, giving a datum and defining its context we receive a particular piece of information. The datum `65` as the string of symbols has in fact no meaning. But it may be, for instance, putted into the context of the data `John` and `Age` (having predefined meanings), and we receive the information that `John is 65 years old`. If we change the context replacing the age by weight, we receive the other information: `John weights 65 kg`. So, the role of context already in the data base systems appears to be crucial in the process of receiving information from data. The context provides the base for generating the results we are expecting from the data base systems some new data. In the case of our example the answer to the question `How old is John?` for instance.

Information may be stored in computer memories. It is the basic idea of *information processing systems*. Defining a suitable context of information at hand we are able to generate, using the computers, the new information. Continuing in our previous example form the information `John is`

`65 years old` and the information `John weights 65 kg` we can receive the new information `John is slim`. However, the construction `IF ((John is 65 years old) AND (John weights 65 kg)) THEN (John is slim)` is no more a simple information but an entity of a new category called in the AI literature usually as a *piece of knowledge*. So, the pieces of knowledge implicitly present in the information systems enable to receive answers to our questions in the form of new information.

3 The Structure of Knowledge

During the past decades the "good old fashioned artificial intelligence" in the form as presented e.g. in [8] provided several specific techniques for representing pieces of knowledge. Let us mention them in very short according to [2] (Chapter 4).

The characterization of approaches consists in derivation of the basic representational formats of any *piece of knowledge* from three basic representational attributes. The specification of pieces of knowledge relates each piece of knowledge with the attribute of its *declarability*, so with the property of any knowledge to be declared, expressed in symbol structures. It is the base for any *symbolic representation* of knowledge, for instance in computer memory structures. Another attribute of pieces of knowledge is their *associability* — the ability of knowledge pieces to be associatively interrelated with some other such pieces in larger networks of concepts in order to characterize the complexity of its real-world aspects, and the semantics of the pieces in the context of related ones. The term *associative* or *semantic nets* are usually used for denoting the resulting structures in artificial intelligence. The attribute of *procedurality* of pieces of knowledge refers to the possibility to manipulate the pieces of knowledge. Such manipulations may transfer them into new contexts or the effective use of knowledge make possible in different particular problem solving processes.

Because each piece of knowledge has each among the three of the just mentioned attributes — it is declared in certain formalism, it has associative links with other pieces of knowledge, and it has its own procedural part prescribing how to use it, the emphasis putted to each one of the above listed three attributes led during the history of artificial intelligence to development of more or less specific *knowledge representation schemes* which have positive but also negative properties.

The effort to integrate the positive sides of all the just mentioned representation schemes, as well as to integrate them into a representational scheme of some other aspects of knowledge like *uncertainties* or *default values* etc., led during the to different variations of schemes more or less similar to, but in basic principles almost identical with, the *frame representation scheme* as proposed in [4].

Knowledge systems are characterized e.g. in [7] as a term referring "...to a computer system that represents and uses knowledge to carry out tasks. This term focuses attention on the knowledge that the system carry, rather than on the question of whether or not such knowledge constitutes expertise. The term domain refers to a body of knowledge." In this aspect knowledge systems slightly differ from the traditional expert systems focused more directly to the given domain of expertise. Knowledge systems contains vast pieces of knowledge mutually interrelated in a context of a given problem domain, or in the case of expert systems in a domain of given expertise. In this contextualized form create the set of pieces of knowledge (the knowledge base of the knowledge system) a coherent body of knowledge on the given domain.

4 Towards the Computational Wisdom?

Consider now a realistic situation that we have lot of knowledge or expert systems at hand, so we are able to put into some contexts not only the knowledge pieces as in the case of building knowledge systems but also a number of different bodies of knowledge. What we will receive as the result of such a contextualization? In the case when the interconnected bodies of knowledge as the result will be on interrelated domains, we will receive a distributed or decentralized knowledge system. Approaching in such a way we will be able to organize more effective medical diagnostic knowledge systems with higher quality of solving problems in different interrelated domains of the medicine, for instance. What result we can image for ourselves supposing the contextual interconnection for larger sets of knowledge systems in the case when the systems are not interconnected through their subjects at least in the first glance? May such kind of construction of decentralized or distributed knowledge systems results in any usable new knowledge system?

The usual answer is YES, and the usual expectation is that such kind of systems will reflect at least a part of a universal human intellectual capacity — a part of the human wisdom. *Wisdom* is – according the Webster's New World Dictionary [3] — " ...the power of judging rightly and following the soundest course of action, based on knowledge, experience, understanding, etc." Might it be that the above mentioned knowledge systems may contribute to better judging, to advice soundest course of actions, etc., namely if their bases of knowledge will be interrelated. There are several authorized questions in interest of present day research — see e.g. [6] for instance — especially in connection with the so called wisdom of crowds (the wisdom of certain kind of human societies). There are also authorized expectations such as that one on the emergent nature of the wisdom supposing that wisdom will not need to be programmed into our knowledge systems, that they will emerge similarly as the consciousness and though will emerge in the case of robots [1]. But what happens in the case if we consider another setting of the question? In other words, if we focus to the individuals, and if the context of the bodies of knowledge will be relatively narrow.

However, are there some boundaries for construction of some kind of computerized systems of human-like wisdom? We suppose that yes. Yes because of the human wisdom is constructed not only on human the base of common sense knowledge, not only on the base of human expert knowledge in different fields, but also on the base of human experiences, on human *common sense* as discussed by many authors – by Minsky in [5], for instance — during the history of AI, generally, on what it means to be a *human subject*, by his or her *personal history*, and *the history of other human beings*, and their past and their *culture* in the present or in the past, on his or her *experiences to have a human body* in the given place and time etc.

These facts, the existence of the personal authentic history, and the existence of the authentic body in the real time-space form, among others, the boundary, and we will be hardly in the position (at least in the near future) to be build a bridge over that boundary with our computing technologies.

Acknowledgments

Research of the author is supported by the project IT4Innovations Centre of Excellence, reg. no. CZ.1.05/1.1.00/02.0070, and sponsored by the Research and Development for Innovations Operational Program from the Structural Funds of the European Union, and from the state budget of the Czech Republic.

References

1. Brooks, R.A.: Cambrian Intelligence: The Early History of the New AI. MIT Press, Cambridge, Massachusetts (1999)
2. Kelemen, J.: Knowledge in Context. Few Faces of the Knowledge Society. Iura Edition, Bratislava (2010)

3. Merriam-Webster: Webster's New World Dictionary, Second College Edition. Prentice Hall, New York (1980)
4. Minsky, M.: A framework for representing knowledge. In: Winston, P.H. (ed.) The Psychology of Computer Vision. McGraw-Hill, New York (1975)
5. Minsky, M.: The Society of Mind. Simon & Schuster, New York (1986)
6. Raisanen, T.: Five experiments of wisdom of crowds as a source of knowledge. In: *Proc. 6th International Workshop on Knowledge Management*, CD Edition. 154–160. Trenčín (2011)
7. Stefik, M.: Introduction to Knowledge Systems. Morgan Kaufmann, San Francisco (1995)
8. Winston, P.H.: Artificial Intelligence. Addison-Wesley, Reading, Mass. (1977)

Multi-agent Systems in Industry: Current Trends & Future Challenges

Paulo Leitao[1,2]

[1] Polytechnic Institute of Bragança
Campus Sta Apolonia, Apartado 1134, 5301-857 Bragança, Portugal
[2] Artificial Intelligence and Computer Science Laboratory
R. Campo Alegre 102, 4169-007 Porto, Portugal
pleitao@ipb.pt

Abstract. This paper introduces the multi-agent systems paradigm and presents some industrial applications of this AI approach, namely in manufacturing, handling and logistics domains. The road-blockers for the current weak adoption of this technology in industry are also discussed, and finally the current trends and several future challenges are pointed out to increase the wider dissemination and acceptance of the multi-agent technology in industry.

Keywords: distributed artificial intelligence, multi-agent systems, industrial applications

1 Introduction

The application of artificial intelligence (AI) mechanisms allows the development of intelligent machines/systems capable to solve very complex engineering problems. Multi-agent systems is one paradigm, derived from the distributed artificial intelligence and artificial life fields, that allows an alternative way to design distributed control systems based on autonomous and cooperative agents, exhibiting modularity, robustness, flexibility, adaptability and reconfigurability. This paper introduces the multi-agent systems paradigm and presents some industrial applications of this AI approach, namely in manufacturing, handling and logistics domains. The road-blockers for the current weak adoption of this technology in industry are also discussed, and finally the current trends and several future challenges are pointed out to increase the wider dissemination and acceptance of the multi-agent technology in industry.

2 Artificial Intelligence and Multi-Agent Systems

The management of complexity, currently found in systems ranging from washing machines to Airbus A380 aircrafts, requires the use of proper mechanisms and techniques. Artificial intelligence (AI), introduced by John McCarthy in 1956, is the science and engineering of making intelligent machines, especially intelligent computer programs mimicking the human though [7]. AI is becoming an essential part of the technology industry, providing solutions for several complex problems in engineering and computer science, namely:

– Game playing, e.g. machines beating human chess players.
– Optimization, e.g. optimizing logistics and production processes.
– Pattern recognition, e.g. detection of trends and patterns in medical or production diagnosis.
– Computer vision, e.g. the navigation of autonomous mobile robots and analysis of medical images.
– Speech recognition, e.g. supporting human-machine interfaces.
– Intelligent control, e.g. providing adaptive and intelligent behaviour to control processes.

When applying AI techniques, several topics should be considered, namely the perception, reasoning, knowledge, planning and learning, as well some philosophical issues about the ethics of creating artificial intelligent beings.

The multi-agent systems (MAS) [10][1] is a paradigm that takes inspiration from several disciplines, mainly from distributed artificial intelligence (DAI) and artificial life (that is related to study and model systems possessing life, i.e. capable of reproducing, surviving and adapting in hostile environments). Multi-agent systems are based on a society of distributed autonomous, cooperative entities, each one having a proper role, knowledge and skills, and a local view of the world, being its behaviour regulated by simple rules. Agent-based solutions replace the centralized, rigid and monolithic control by a distributed functioning where the interactions among individuals lead to the emergence of "intelligent" global behaviour (see Fig. 1). Note that such systems exhibit high degree of autonomy and re-configurability, without a fixed client-server structure.

MAS is aligned with the current trend to build modular, intelligent and distributed control systems, which exhibit innovative features, like the agile response to the occurrence of disturbances and the dynamic re-configuration on the fly, i.e. without the need to stop, re-program and restart the process.

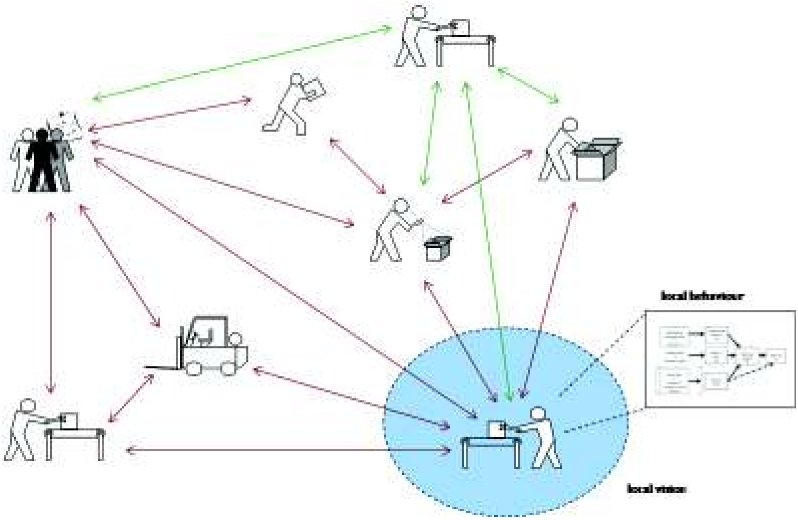


Fig. 1. MAS working in practice

3 Applications of MAS in Industry

The MAS approach is suitable to support the current requirements for modern control systems in industrial domains, providing flexibility, robustness, scalability, adaptability, re-configurability and productivity. MAS is being applied with success to a wide range of domains, namely electronic commerce, graphics (e.g., computer games and movies), transportation, logistics, robotics, manufacturing, telecommunications and energy.

As examples, it is possible to refer the application of multi-agent systems solutions in the Daimler Chrysler factory of engines in Stuttgart [9], Tankers International that operates one of the largest oil tanker pools in the world [3], Air Liquide America to optimize the distribution of medical and industrial gases [2] and US Navy ships to control the heating, ventilation and air conditioned (HVAC) systems [5]. A deep analysis of industrial applications of MAS can be found in [4] and [8].

The analysis of the surveyed industrial applications of agent-based solutions allows extracting the following conclusions:

- Relatively small adoption of agents in industry, being the implemented applications limited in terms of functionality.

– The solutions address mainly the high-level control or the pure software systems (e.g. the electronic commerce).
– Little enthusiasm from both the technology providers and the industry companies.

The reasons for this weak adoption in industry were already widely discussed in the literature by several authors, namely [4] and [6]. Briefly, the main road-blockers are the required initial investment, the need to adopt the distributed thinking, the interoperability in distributed heterogeneous systems, the missing standardization, the real-time constraints and the missing technology maturity.

4 Current Trends and Future Challenges

Lately, some promising perspectives for the adoption of the agent technology were provided by the development of multi-agent based solutions by several software developers companies, e.g. NuTech Solutions, Magenta Technology, Smart Solutions and Whitestein Technologies, and by several automation technology providers, e.g. Rockwell Automation and Schneider Electric. However, the main trend in the industrial application of multi-agent systems is to convince industry people of the benefits of using agents, e.g. by providing demonstrators running in industry that shows the maturity, flexibility and robustness of agent-based solutions. This will allow industrial companies to "believe" in the agent technology and its principles.

Additionally, several future challenges can be pointed out in industrial agents, namely:

– *Standardization*, which is pointed out by industry as a major challenge for the industrial acceptance of the agent technology, since standards may affect the development of industrial MAS solutions, namely the IEEE FIPA (Foundation for Intelligent Physical Agents), IEC 61131-3, IEC 61499, ISA 95 and semantics and ontologies standards.
– *Integration of other complementary technologies*, e.g. IEC61131-3 and IEC 61499 approaches to implement the low-level control that is not addressed by the agents, and Service Oriented Architectures (SOA) / Web services to solve the interoperability problems allowing the vertical and horizontal integration.
– *Mature engineering development methodologies, deployment and tools*, that simplifies the engineering of agent-based systems. For this purpose, simulation is a need to test the emergent behaviour before the real deployment.

– *Bio-inspired techniques*, to enhance the engineering of more robust, adaptive, reconfigurable and responsive systems. In particular, self-organization is mandatory to support re-configuration and evolution, being also important to consider other self-* properties, such as self-learning, self-adaptation, self-optimization and self-healing.

The fulfilment of these challenges leads to the development of more powerful agent-based systems that may be better accepted by industry.

5 Conclusions

As conclusions, AI provides a set of advantages to improve the performance of automatic complex systems, and the multi-agent systems, as a paradigm derived from AI, is suitable to address the current requirements imposed to industrial companies. In spite of being already adopted in several industrial domains, the multi-agent technology still has a long and difficult path to be traversed for a wider acceptance of these AI concepts in industry.

References

1. Ferber, J.: Multi-Agent Systems: An Introduction to Distributed Artificial Intelligence. Addison-Wesley Professional (1999)
2. Harper, C.N., Davis, L.: Evolutionary computation at American Air Liquide. In: Evolutionary Computation in Practice, 313–317 (2008)
3. Himoff, J., Skobelev, P., Wooldridge, M.: Magenta technology: multi-agent systems for industrial logistics. In: *AAMAS 05 Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems.* 60–66 (2005)
4. Leitão, P.: Agent-based distributed manufacturing control: A state-of-the-art survey. *Engineering Applications of Artificial Intelligence* 22(7), 979–991 (2009)
5. Maturana, F.P., Staron, R.J., Hall, K.H., Tichý, P., Šlechta, P., Mařík, V.: An intelligent agent validation architecture for distributed manufacturing organizations. In: Camarinha-Matos, L. (ed.) Emerging Solutions for Future Manufacturing Systems. 81–90 (2004)
6. Mařík, V., McFarlane, D.: Industrial adoption of agent-based technologies. *IEEE Intelligent Systems* 20(1), 27–35 (2005)
7. McCarthy, J.: What is artificial intelligence? http://www-formal.stanford.edu/jmc/whatisai/whatisai.html (2007)
8. Monostori, L., Váncza, J., Kumara, S.: Agent-based systems for manufacturing. *CIRP Annals - Manufacturing Technology* 55(2), 697 – 720 (2006)

9. Schild, K., Bussmann, S.: Self-organization in manufacturing operations. *Communications of the ACM* 50(12), 74–79 (2007)
10. Wooldridge, M.: An Introduction to Multi-Agent Systems. John Wiley & Sons, Inc., New York, NY, USA (2002)

Perspectives of Using Membrane Computing in the Control of Mobile Robots

Cătălin Buiu, Ana Brânduşa Pavel, Cristian Ioan Vasile, and Ioan Dumitrache

Politehnica University of Bucharest,
Splaiul Intependenţei, No. 313, 060042, Bucharest, Romania
{cbuiu,apavel,cvasile,idumitrache}@ics.pub.ro

Abstract. This paper presents an original approach to the control of mobile robots using a natural computing based solution, which falls beyond the traditional use of Artificial Intelligence and bio-inspired methods in robot control. The idea is to use the modeling power of P systems, which are based on the structure and functioning of living cells. Results obtained so far are presented in a synthetic way and directions for further developments are given.

Keywords: numerical P systems, enzymatic numerical P systems, membrane controllers, mobile robots, distributed and parallel systems

1 Introduction

AI-inspired methods have a long tradition in the control of mobile autonomous robots, and a number of important concepts, tools and techniques of AI have been applied successfully to various robotic applications. One can mention here planning, learning and other aspects. Recently, computational intelligence field has provided a variety of methods that are being applied to the problem of controlling mobile robots, such as fuzzy logic, neural networks, evolutionary computing.

Of more recent interest is the paradigm of natural computing which refers to that category of methods which are directly inspired by the structure and functioning of living systems, such as evolutionary computing and swarm intelligence.

Membrane computing, introduced by Gh. Paun [1], has raised a lot of interest during the last 10 years. Several results, mainly at theoretical level, have been generated and some practical applications of P systems (or membrane systems) were also presented. For a review, consult [2].

A P system (PS) is a computational model based upon the structure of an eukaryotic cell and it mimics the interaction and evolution of chemicals inside of biomembranes [2]. The architecture (or membrane structure) of a PS is a hierarchical arrangement of membranes, in which the outermost membrane is called the skin membrane and separates the system from its environment. This skin membrane is analogous to the plasma membrane of a living cell. The membranes define regions as the biomembranes define working compartments. A membrane is called elementary if it has no other membrane inside. For a sample membrane structure, refer to [3] and [2].

Next section will present the main arguments for using membrane systems in robotics applications, while section 3 will briefly enumerate the main results obtained so far, both at practical level, and at theoretical level. This paper ends with some conclusions and ideas for possible developments.

2 Membrane Computing and Robotics. An Overview of Concepts

The basic idea of membrane systems, to separate computing processes in different compartments (membranes) which are able to inter-communicate, sounds appealing to researchers in autonomous robotics. In 2009, there was proposed the first idea of using membrane systems as basic modules for cognitive architectures, in general, and for robot control architectures, in particular. The idea is to treat the control architecture as a basic membrane (skin membrane, in the terms of membrane systems), while different modules may be treated/implemented as specific membranes. For example, the execution level of a general control architecture may be implemented as a membrane containing several sub-membranes responsible for various desired behaviors of the robot.

Numerical P systems (NPS) are a class of computing models inspired by the cell structure in which numerical variables evolve inside the compartments by means of programs; a program (or rule) is composed of a production function and a repartition protocol. The variables have a given initial value and the production function is usually a polynomial. The value of the production function for the current values of the variables is distributed among variables in certain compartments according to a repartition protocol; NPS model was first introduced in [3] with possible applications in economics. Using NPS as a possible tool for modeling robot behaviors is a novel approach which will be further discussed.

In [4], several examples of robot behaviors (obstacle avoidance, wall following, following a leader) modeled with NPS, are presented. The main advantages of using NPS as a modeling tool are the numerical and the naturally parallel and distributed nature of the model. Membranes of a NPS can be distributed over a grid or over a network of microcontrollers in a robot. The computation done in each membrane region (the execution of the membranes programs) can also be done in parallel. This is very important, because given a membrane system, which is an abstract implementation of a desired behavior, it can be executed in a distributed and parallel way without having to worry about the design and implementation issues regarding parallelization and distribution. Therefore, NPS can be used as a modeling tool for parallel and distributed control systems.

The performance of a controller can be measured by the mean execution time of a cycle [4]. After designing a membrane controller, the membrane system can be simulated using a numerical P systems simulator such as SNUPS [5], [6] or SimP [7]. One benefit of using membrane controllers is that the computational performance can be increased by improving the simulator's performance (parallelization, distribution and other optimizations) and not by modifying the membrane controllers themselves. The membrane simulator can be considered as a virtual machine like Java or Python virtual machines. It can be seen as a "membrane computer" that runs "membrane programs" (in this case, the membrane controllers). The simulator is the middleware between the hardware and the membrane controllers. Therefore, by optimizing its performance, the performance of all the defined membrane controllers increases.

Design and implementation of robot controllers require deterministic computational models. In order to be deterministic, a NPS model must have only one rule per membrane and, most of the time, this restriction makes the model rigid and difficult to use. Therefore, an extension of the NPS model, Enzymatic Numerical P Systems (ENPS), was proposed in the context of modeling robot behaviors. ENPS model allows the parallel execution of more rules (programs) per membrane while keeping the deterministic behavior. ENPS use some special variables, inspired by the behavior of biological enzymes which, associated to rules (in analogy to chemical reactions), can decide whether a rule is active or not at a given computational step. A rule is active if the associated enzyme has a greater value than the minimum of the variables involved in the rule or if the rule has no associated enzyme. Details about ENPS model together with formal definition and examples can be found in [8], [9] and [10]. Theoretical results about the universality of the model are presented in [11].

By adding enzyme-like variables to the NPS models, the modeling power of NPS increases. The enzymatic mechanism controls the execution flow of a NPS with multiple rules per membrane. The possibility of selecting and executing more production functions per membrane makes NPS a more flexible modeling tool. ENPS robot controllers (as those described in [8], [9]) have a less complex structure than the NPS ones, therefore less computations must be performed, increasing the performance of the system.

3 Current Theoretical and Experimental Results

Both NPS and ENPS models could be used for modeling autonomous mobile robot behaviors [4], [8] and [9]. The numerical nature, the distributed and parallel structure and the computing power, make membrane controllers suitable candidates for the control of complex systems.

A framework for testing membrane controllers on real and simulated robots has been developed. The framework integrates xml files which store robot behaviors in a platform-independent way, a simulator for numerical P systems [7] and Webots, a professional mobile robot simulator [12]. The proposed membrane controllers have been tested both on real and simulated robots (e-puck and KheperaIII).

Membrane controllers modeled using ENPS have less complicated structures than the ones modeled with NPS. For instance, the ENPS controller for obstacle avoidance proposed in [8] is a membrane system with 9 membranes, while the NPS model for obstacle avoidance proposed in [4] has 37 membranes. Although the ENPS model for obstacle avoidance has more rules than the NPS model, not all of them are active (fewer rules are executed during a computational step in the ENPS model than in the NPS one).
A NPS structure for odometric localization has been modeled, but its structure is far way complicated than the ENPS one proposed in [9]. In this case, the NPS controller was modeled using 24 membranes, while the ENPS model has only 5 membranes. In this case, enzymes control the program flow and are used as stop conditions and synchronization mechanism. Therefore, the model of the controller is clearly simplified, easier to implement and more efficient (regarding the computational process) than the one modeled by classical NPS. The naturally parallelized membrane representation and the numerical nature of the membrane components represent advantages for both NPS and ENPS in designing and modeling robot behaviors.

Based on the theoretical and practical results, some important advantages of using ENPS to classical NPS are mentioned in [8] and [9]. The main advantages of ENPS towards NPS are that enzyme-like variables can control the program flow by deciding which rules to be executed in a computational step, they can control the synchronization between parallel computations, they can be used to filter noise from the sensors or to detect the termination of the program.

4 Conclusions

This paper presents the currently existing results in modeling robot controllers by means of P systems. Numerical P systems and its extension, enzymatic numerical P systems, have been used to achieve cognitive robot behaviors.

Taking into account the most important properties of numerical P systems: their numerical nature, parallelization, distribution and the tree-like structure of the membranes, a future aim of this research direction is to design hardware membrane controllers and prove their efficiency (for example implementing the membrane controllers on a FPGA which can be connected to the robot).

Acknowledgments

This paper is suported by the Sectorial Operational Programme Human Resources Development, financed from the European Social Fund and by the Romanian Government under the contract number SOP HRD/107/1.5/S/82514.

References

1. Păun, Gh.: Computing with membranes. *Journal of Computer and System Sciences*, vol. 61, 108–143. Elsevier (2000)
2. Păun, Gh., Rozenberg, G., Salomaa, A. (eds.): The Oxford Handbook of Membrane Computing. Oxford University Press (2010)
3. Păun, Gh., Păun, R.: Membrane computing and economics: Numerical P systems. Fundamenta Informaticae, no. 1, vol. 73, 213–227 (2006).
4. Buiu, C., Vasile, C., Arsene, O.: Development of membrane controllers for mobile robots. *Information Sciences*, doi:10.1016/j.ins.2011.10.007 (2011)
5. Arsene, O., Buiu, C., Popescu, N.: SNUPS – A simulator for numerical membrane computing. *Intern. J. of Innovative Computing, Information and Control*, vol. 7, no. 6, 3509–3522 (2011)
6. Buiu, C., Arsene, O., Cipu, C., Pătraşcu M.: A software tool for modeling and simulation of numerical P systems. *BioSystems*, vol. 103, no. 3, 442–447 (2011)
7. Pavel A.B.: Membrane controllers for cognitive robots. Master Thesis, Department of Automatic Control and System Engineering, Politehnica University of Bucharest (2011)

8. Pavel A.B., Buiu C.: Using enzymatic numerical P systems for modeling mobile robot controllers. *Natural Computing*, in press (2011)
9. Pavel A.B., Vasile, C., Buiu, C.: Robot controllers implemented with enzymatic numerical P systems. submitted (2011)
10. Pavel, A.B., Arsene, O., Buiu, C.: Enzymatic numerical P systems – a new class of membrane computing systems. *The IEEE Fifth Intern. Conf. on Bio-Inspired Computing. Theory and applications* (BIC-TA 2010), 1331–1336, Liverpool, U.K. (2010)
11. Vasile, C., Pavel, A.B., Dumitrache, I., Păun, G: On the Power of Enzymatic Numerical P Systems. submitted (2011)
12. Cyberbotics, professional mobile robot simulation
http://www.cyberbotics.com/

Implementing ENPS by Means of GPUs for AI Applications

Manuel García–Quismondo and Mario J. Pérez–Jiménez

Research Group on Natural Computing
Dpt. of Computer Science and Artificial Intelligence, University of Sevilla
Avda. Reina Mercedes s/n. 41012 Sevilla, Spain
{mgarciaquismondo,marper}@us.es

Abstract. A P system represents a distributed and parallel computing model in which basic data structures are, for instance, multisets and strings. Enzymatic Numerical P Systems are a type of P systems whose basic data structures are sets of numerical variables. Separately, GPGPU is a novel technological paradigm which focuses on the development of tools for graphic cards to solve general purpose problems. This paper proposes an ENPS simulator based on GPUs and presents general concepts about its design and some future ideas and perspectives.

Keywords: numerical P systems, enzymatic numerical P systems, membrane controllers, distributed and parallel systems

1 Introduction

Membrane computing is a bio-inspired branch of natural computing, abstracting computing models from the structure and functioning of living cells and from the organization of cells in tissues or other higher order structures [15]. This branch of natural computing studies the design and properties of membrane systems or *P systems*. P systems are distributed and parallel computing models structured in compartments known as *membranes*. Membranes have associated basic data structures such as multisets, strings or numerical variables [16]. According to the way in which membranes are structured, there are several types of P systems. For instance, there exist cell-like P systems [15], tissue P systems [14] and spiking neural P systems [11], along with other types. In P systems, membranes and their associated data structures are processed by means of rewriting rules or *programs* associated to the cells, in order to perform sequences of configurations or *computations* [15][16].

P systems have been successfully applied in a wide range of domains [4]. For instance, they have been applied in microbiological modelling in order to model phenomena such as *quorum* sensing in *Vibrio fischeri* populations [20] and ecological modelling to predict the evolution of the bearded vulture [3] population in the Catalan Pyrenees, as well as image thresholding [6]. Such a versatility makes P systems a useful tool for gaining knowledge about a vast variety of different domains, thus providing promising approaches within the range of disciplines which composes the field of study of artificial intelligence.

A special type of P systems are enzymatic numerical P systems (ENPSs) [18]. ENPSs describe a deterministic, parallel model in which the basic data structures associated to membranes are numerical values which evolve by means of *programs* associated to the membranes [16]. In order for a program to be applied, a certain amount of a specific type of variable (enzyme) may be needed [18]. This model of computation has already been successfully applied to model robot controllers for obstacle avoidance, in which a robot needs to avoid obstacles situated in a closed circuit [19].

Separately, *GPGPU* is a novel technological discipline which consists of the application of graphic cards (GPUs) in order to execute parallel, distributed algorithms [22]. The basic idea is to take advantage of the parallel architecture of GPUs, traditionally used for graphics processing, in order to execute algorithms which can be performed in parallel, thus accelerating them.

In this paper, a GPU simulator for ENPSs is proposed. The parallel architecture of ENPSs makes the simulation of their computations a suitable task to be parallelized, thus expecting an acceleration in the simulation times if compared to the ones obtained by using sequential simulators.

This paper is structured as follows. Section 2 provides a quick introduction to ENPSs as a model of computation and discusses their applications in artificial intelligence. Section 3 provides a general overview of the current state-of-the-art about the results obtained by previous GPU simulators within the field of membrane computing. Finally, section 4 presents the conclusions obtained and proposes some directions for future work.

2 Enzymatic Numerical P Systems

Numerical P systems (NPSs) are a special kind of P systems in which numerical variables evolve from initial values by means of programs. Each program is composed of a production function and a distribution protocol. Each production function is a numerical function over a set of variables. If a variable appears on at least one production function, then it is consumed and its value is set to 0. Each repartition protocol updates its values according to the result of the production function of its program and a coefficient associated to each variable [15]. A special kind of NPSs are enzymatic numerical P systems (ENPSs). Unlike NPSs, ENPSs describe a deterministic, parallel model of computation. ENPSs introduce the optional use of enzymes associated to programs. Thus, a program is applied only in the following cases: 1) The program does not have an associated enzyme. 2) The value of its associated enzyme is greater than the minimum of the values of the variables consumed by the program. All active programs in each membrane are executed in parallel. More information about ENPSs can be found in [18][19].

ENPSs have been successfully applied within the field of robotics. For instance, they have been used to model deterministic mobile robot controllers for obstacle avoidance. In this model, the speed of the two robot motors is set according to the values assigned to two variables of the system. Thus, the dynamical evolution of these variables describes the behavior of the robot through a closed circuit [19].

2.1 ENPSs and Artificial Intelligence

Mobile robot control problems, such as obstacle avoidance and odometric localizacion, can be considered as artificial intelligence problems. For instance, obstacle avoidance can be considered as a high-level planning problem [13]. In obstacle avoidance, the objective is to find a sequence of movements in a static or dynamical environment. The objective of this sequence is for robots which follow it to avoid crashing with any obstacles they might find in the environment. The input data is given as a series of sensor lectures obtained from the environment. This type of path planning problems arising from the field of robotics has already been attacked by using artificial intelligence techniques such as ant colony algorithms [9][8].

Odometric localization is a widely used method for estimation of the momentary pose of a mobile robot with respect to its starting pose [12]. This estimation is affected by several error sources, such as imprecission in the mobile robot kinematic parameters and errors in the sensor lectures [1]. Thus, odometric localization entails an optimization problem, i.e., minimizing the global error in the pose estimation. As an optimization problem, odometric localization has been previously tackled by using well-known artificial intelligence paradigms, such as genetic algorithms [10] and artificial neural networks [7]. All in all, ENPs propose a new framework which can be applied in order to solve artificial intelligence problems arising from robotics [19].

2.2 Simulation of ENPSs

ENPSs describe a parallel model. Therefore, the huge computational power required by extensive models (for instance, those necessary for massive robot swarms and robots with complex sensor networks) accounts for the need for high performance computing platforms to simulate them. Besides, their parallel structure makes them appropriate to be simulated by means of parallel architectures such as GPUs, FPGAs and computer clusters.

3 Compute Unified Device Architecture (CUDA) Parallel Programming Model

Modern GPUs can physically contain up to 240 processor cores and 30,720 threads. All these threads are executed in parallel. Thus, modern GPUs define an architecture composed of a large number of parallel processing units with a certain degree of independency from each other [22]. In order to make the most of this massively parallel architecture, it is necessary to make use of programming languages specifically designed for these devices. Two of the main standards in GPGPU are OpenCL [21] and CUDA [23]. CUDA defines a parallel programming model which is an abstraction of the specific parallel device where the program is to be executed. The CUDA programming model developed by NVIDIA allows developers to write scalable parallel programs for GPUs using a straightforward extension of the C language (named *CUDA-C*). CUDA-C is a language designed to make the most of the GPGPU approach by enabling programmers to encode parallel applications to be run on GPUs [5]. That is, programmers are able to develop code to be executed on each GPU thread at the same time. This way they can take advantage of the GPU parallel architecture in order to obtain enormous speed-up if compared to sequential versions of the same code. More information about the CUDA programming model can be found on [23].

GPGPU and CUDA have been already successfully applied in order to simulate different kinds of P systems. To the best of our knowledge, they have been applied to simulate cell-like object-based P systems [5] and spiking neural P systems [2]. Their results include data which show noticeable speed-ups in comparison to their sequential counterparts. These results demonstrate the suitability of the GPGPU approach for simulating P systems in a parallel mode.

3.1 Design of the Simulator

The objective of the proposed ENPS GPU-based simulator is to fully simulate the behaviour of enzymatic numerical P systems, performing operations in parallel whenever possible. In order to do that, it is crucial to identify the operations susceptible for parallelization and write parallel kernels for them. This way the simulator can take advantage of the underlying parallel architecture. Thus, in each computational step, the simulator performs the following operations.

First, it checks each program in a different thread. This checking selects those programs which can be applied. Then, it assigns the production function of each applied programs to a thread. If there is at least one applicable program to consume a variable, then that variable is set to 0. Once all consumed variables have been cleared, each applicable production function is computed in parallel. Then, the contribution of each applicable distribution protocol to each variable is computed in parallel. Eventually, all contributions are added to the variables in order to update their values.

This simulator will be published under open source license. It can be used for simulating complex distributed processes modeled with enzymatic numerical P systems. Therefore, several robot behaviors can be simulated in parallel (for example, a robot could avoid obstacles, follow another robot or look for a target at the same time). The synchronization between several behaviors of one robot is done by the help of the enzyme variables which can be used as stop conditions [19]. Apart from simulating several behaviors for only one robot in parallel, the simulator could be used to simulate interaction and cooperation between several robots in complex distributed robotic systems.

3.2 Simulator Performance

All parallel parts of the algorithm are executed with a degree of parallelism of at least equal to the number of programs of the simulated model. The degree of parallelism can be even greater when the repartition protocol is applied. Hence, a theoretical acceleration of at least the number of programs of the model could be reached, if compared to the runtime of sequential simulators. Specifically, the simulator was tested by using an ENPS model of obstacle avoidance [19] as an example, along with other models. These models were simulated by using SNUPS [18]. Then, the resulting runtimes were compared with the GPU simulator runtimes, in order to get an approximate speed-up. In the specific case of the obstacle avoidance model, the total number of programs is 41 [19]. Hence, an acceleration of at least 41 is to be theoretically expected in this case, if compared to sequential ENPSs simulators.

4 Conclusions

In this paper, a GPU-based simulator for enzymatic numerical P systems is proposed. Enzymatic numerical P systems describe a parallel computing model with applications in artificial intelligence. This simulator might be suitable for large scale models which can be applied within the field of robotics.

The massively parallel environment provided by the GPUs is suitable for enzymatic numerical P systems simulations. However, it would be interesting to explore the possibility of scaling-up the currenly existing robot behaviors modeled with ENPSs and simulate them by means of GPU clusters or other parallel architectures (such as FPGAs or computer clusters). These systems might be applied to model the behavior of massive robot swarms and complex sensor networks.

References

1. Antonelli, G., Chiaverini, S., Fusco, G.: A calibration method for odometry of mobile robots based on the least-square technique: Theory and experimental validation. *IEEE Transactions on Robotics*, 21, 5, 994-1004 (2005)
2. Cabarle, F., Adorna, H., Martínez-del-Amor, M.A.: A Spiking Neural P system simulator based on CUDA. *12th International Conference on Membrane Computing (CMC12)*, 23/08/2011-27/08/2011, Fontainebleau, France, 77-92 (2011)
3. Cardona, M., Colomer, M.A., Pérez-Jiménez, M.J., Sanuy, D., Margalida, A.: Modelling ecosystem using P systems: The bearded vulture, a case study. *Lecture notes in Computer Science*, 5391, 95-116 (2009)
4. Cecilia, J.M. García, J.M., Guerrero, G.D., Martínez-del-Amor, M.A, Pérez-Hurtado, I., Pérez-Jiménez, MJ.: Simulation of P systems with Active Membranes on CUDA. *Briefings in Bioinformatics*, 11, 3, 313-322 (2010)
5. Cecilia, J.M. García, J.M., Guerrero, G.D., Martínez-del-Amor, M.A, Pérez-Hurtado, I., Pérez-Jiménez, MJ.: Simulating a P system based efficient solution to SAT by using GPUs. *The Journal of Logic and Algebraic Programming*, 79, 317-325 (2010)
6. Christinal, H.A., Díaz-Pernil, D., Gutiérrez-Naranjo, M.A., Péerez-Jiménez, M.J.: Thresholding of 2D Images with Cell-like P Systems. *Romanian Journal of Information Science and Technology (ROMJIST)*, 13, 2, 131-140 (2010)
7. Conforth, M., Meng, Y.: An Artificial Neural Network Based Learning Method for Mobile Robot Localization. *Robotics Automation and Control*, InTech, Ch. 6, 2008
8. Dong, J., Liu, B., Peng, K., Yin, Y.: Robot Obstacle Avoidance based on an Improved Ant Colony Algorithm. *WRI Global Congress on Intelligent Systems (GCIS '09)*, 19/05/2009-21/05/2009, Xiamen, China, 103-106 (2009)

9. Du, R., Zhang, X., Chen, C., Guan, X.: Path Planning with Obstacle Avoidance in PEGs: Ant Colony Optimization Method. *International Conference on Cyber, Physical and Social Computing (CPSCom) 2010*, 18/12/2010-20/12/2010, Hangzhou, China, 768-773 (2010)
10. Gill, M.A.C, Zomaya, A.Y.: Genetic algorithms for robot control. *IEEE International Conference on Evolutionary Computation 1996*, 29/11/1996-01/12/1996, Perth, Australia, 462 (1996)
11. Ibarra, O., Pérez-Jimenez, M.J., Yokomori, T.: On spiking neural P systems. *Natural Computing*, 9, 2, 475-491 (2010)
12. Ivankjo, E., Komšić, I., Petrović, I.: Simple Off-Line Odometry Calibration of Differential Drive Mobile Robots. *Proceedings of 16th International Workshop on Robotics in Alpe-Adria-Danube Region - RAAD 2007*, 07/06/2007-09/06/2007, Ljubljana, Slovenija, 164-169 (2007)
13. Khatib, O.: Real-time obstacle avoidance for manipulators and mobile robots. *The International Journal of Robotics Research*, 5, 1 (1986), 90-98
14. Pan, L., Pérez-Jiménez, M.J.: Computational complexity of tissue-like P systems. *Journal of Complexity*, 26, 3, 296-315 (2010)
15. Paun, Gh, Paun, R.: Membrane Computing and Economics: Numerical P Systems. *Fundamenta Informaticae*, 73, 1-2, 213-227 (2006)
16. Paun, Gh. Computing with membranes. *Journal of Computer and System Sciences*, 61, 1, 108-143 (2000)
17. Paun, Gh.: Membrane Computing. An Introduction. Springer, XI+419 (2002)
18. Pavel, A., Arsene, O., Buiu, C.: Enzymatic Numerical P Systems - A New Class of Membrane Computing Systems. *Proceedings 2010 IEEE Fifth International Conference on Bio-inspired Computing: Theories and Applications (BIC-TA 2010)*, 23-26/09/2010, Liverpool, UK, 1331-1336 (2010)
19. Pavel, A, Buiu, C.: Using enzymatic numerical P systems for modeling mobile robot controllers. *Natural Computing*, (in press)
20. Romero, F.J., Pérez-Jiménez, M.J.: A model of the Quorum Sensing System in Vibrio Fischeri using P systems. *Artificial Life*, 14, 1, 95-109 (2008)
21. Takizawa, H, Koyama, K., Sato, K. Komatsu, K., Kobayashi, H.: CheCL: Transparent Checkpointing and Process Migration of OpenCL Applications. *International Parallel and Distributed Processing Symposium (IPDPS11)*, 16/05/2011-20/05/2011, Anchorage, USA (2011)
22. *http://www.gpgpu.org*
23. *http://www.nvidia.com/object/cuda_home_new.html*

New Way How to Create an Autonomous Creature

Pavel Nahodil and Jaroslav Vítků

Department of Cybernetics, FEE, CTU in Prague
Technicka 2, 16627, Prague 6, Czech Republic
{pavel.nahodil, vitkujar}@fel.cvut.cz

Abstract. We introduce new approaches for creating of autonomous agents. The life of agent created by us is very similar to the animal's life in the Nature, which learns autonomously from the simple tasks towards the more complex ones and is inspired in AI, Biology and Ethology. We present our proved design of artificial creature, capable of learning from experience in order to fulfil more complex tasks, which is based mainly on ethology. It integrates several types of action selection mechanisms and learning into one system. The main advantages are in its autonomy, ability to gain all information from the environment and decomposition of the decision space into the hierarchy of abstract actions, which dramatically reduces the total size of decision space. The agent learns how to exploit the environment continuously, where the learning of new abilities is driven by his physiology, autonomously created intentions, planner and neural network.

Keywords: agent, creature, behaviour, architecture, intentions, planning.

1 Introduction

One of the ultimate goals of nowadays research in the field of technology, to create an autonomous robotic system capable of flawless operation in the real environment full of disturbances and unpredicted events, still has not reached. The complexity and observability of the real environment is simply too high to be exactly described by anything less complex than the environment itself. As the given problem cannot be precisely described by the mathematics, some other approaches which are working with uncertainty and incomplete models must be used. We think that the best chance of fulfilling the given aim is in constructing something relatively simple, but capable of autonomous learning and using this new knowledge to improve itself. This paper tries to describe the possible method of solving this problem by presenting a new architecture of autonomous agent partially capable of things described above, called "An Artificial Creature Capable of Learning from Experience in Order to Fulfil More Complex Tasks" [8].

The main feature of this architecture is in its total autonomy, ability to gain all information from the surrounding environment and effective information filtering and classification. Agent can operate based only on the sensory input and by its actuator system, so the resulting architecture is almost fully independent on the concrete area and form of use. Consequently it is unimportant whether the agent is embodied in some robotic system, intelligent house, or just operates in some virtual environment. Thanks to the fact that all the designer has to specify is just the sensory layer, actuator layer and agent's needs, the architecture should be convenient especially in unknown environments, where some complex task has to be fulfilled.

Agent architecture is inspired by the layered model, combining various approaches on different layers. The life of agent is similar to a newly born animal, which explores new and unknown environment, learns from experiences and links the newly learned abilities to its needs in order to survive and increase effectiveness of its behaviour. New knowledge is learned simultaneously on various levels of abstraction using different learning approaches.

One of the main features is an alternative implementation of system similar to reactive and hierarchical planning. The system combines hierarchical reinforcement learning and planning engine into a domain independent hierarchical planner.

2 Main Parts of the Control System

The architecture is an attempt to create an agent capable of autonomous intelligent operation in nearly so complex environments as is the real world. The agent could be used in the real or virtual environments where the designer has no a priori knowledge about its fundamental regularities and is able to specify only the agent's needs and some high-level goals. The agent autonomously explores the environment and tries to "understand" the principles from simple to the complex ones in order to gain the ability to survive and provide various services to the user.

The single kind of problem representation or approach to solving the problem is almost never sufficient. More than just exploring the capabilities of one concrete kind of decision making, we have decided to use different approach for creating this agent. Each action made by living animals is consequence of superposition of many different motivations, needs, emotions, intensions etc. We simulate this by connecting several of decision and control blocks. Each block consists of some kind of well-known and widely used system, such as for example neural network or planner. We want these blocks to be interconnected in such manner, that some kind of intelligent behaviour emerges from their interaction. The resulting system also suppresses the weaknesses of particular subsystems and exploits their benefits more efficiently. Three main subsystems used in this architecture are from the bottom: artificial neural network, reinforcement learning and planning. We will describe their purpose in more details here.

The core of entire architecture is created by the block providing the hierarchical reinforcement learning. This implementation of *Reinforcement Learning* (RL) is mainly inspired by the Dissertation Thesis [1], where Kadleček presented his idea called *"Hierarchy, Abstraction, Reinforcements, Motivations Agent Architecture"* (**HARM**) [2],[3]. It is the hierarchical reinforcement learning with autonomous creation of hierarchy based on the interaction with the surrounding environment. According to the HARM, the agent has some physiological state space described by the dynamic system; this space contains several variables, where each variable represents some of the agent's needs.

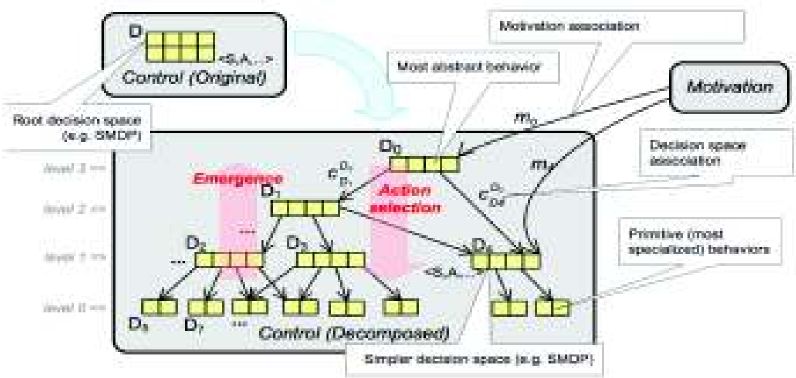


Fig. 1. Hierarchically decomposed decision space by the HARM system. The root node corresponds to the most abstract behaviors, whereas the leaf nodes correspond to the lowest behaviors-primitive actions.

At the beginning of the simulation, the hierarchy of abstract actions is empty the agent acts randomly and observes whether he caused some active change of some of his physiological variables towards the limbo area. If so, the new abstract action in the hierarchy is created and connected to the motivation. The abstract action corresponds to the new decision space used by the RL engine. Each decision space contains some subset of environment variables and agent's actions, which is defined online during the agent's life. This system dramatically reduces the total size of search space used by the RL engine. The Q-Learning algorithm is used in each decision space provides online trade-off between the exploration and knowledge exploitation. The resulting primitive action executed by the agent is then composition of decisions made by all RL engines created so far. The main benefit of this system is the fact that the HARM is capable of autonomous creation of action hierarchy based on the interaction with surrounding environment; this reduces the decision space that have to be considered.

This system was further improved for online continuous learning and augmented with the **Intentional State Space**, it has the similar purpose as the physiological state space, and the difference is that agent's intentions are created autonomously during his life. If the agent discovers that is able to actively change some environment variable (e.g. turn on the lights), the new intentional variable is created. Intentional state variables have predefined its own dynamics and motivate the agent to learn and "train" this newly discovered behaviour. Because of this part, the agent can autonomously discover new potentialities of the environment and learns how to exploit it, this new knowledge can be exploited later. This approach corresponds to learning of newly born animal by play.

The other main subsystem of the architecture is the hierarchical **Planning Engine**. It is composed of classical "flat" planner operating over the hierarchy of decision spaces. One of the most important ideas of this architecture is in the ability to represent the abstract action from the RL action hierarchy as a set of primitive actions in the *Stanford Research Institute Problem Solver* (STRIPS) language [4]. This gives the agent ability to deliberatively "think" about the actions previously learned during the interaction with the environment and to fulfil complex tasks. The advantages of hierarchical decomposition of plans are already well known, and referred for example in [5] as *Abstraction Space Hierarchy from STRIPS* (ABSTRIPS) or *Hierarchical-Task Network* (HTN) [6]. Our planner can operate over the abstract actions on arbitrary level of action abstraction. The main benefit of this solution is in the fact that the complicated outer world is pre-processed by the RL-action hierarchy, so this hierarchical planner can operate in very complex domains, while still maintaining its domain independence, which is advantage against the well-known hierarchical planners. More precisely, our planner could be referred as domain self-configurable.

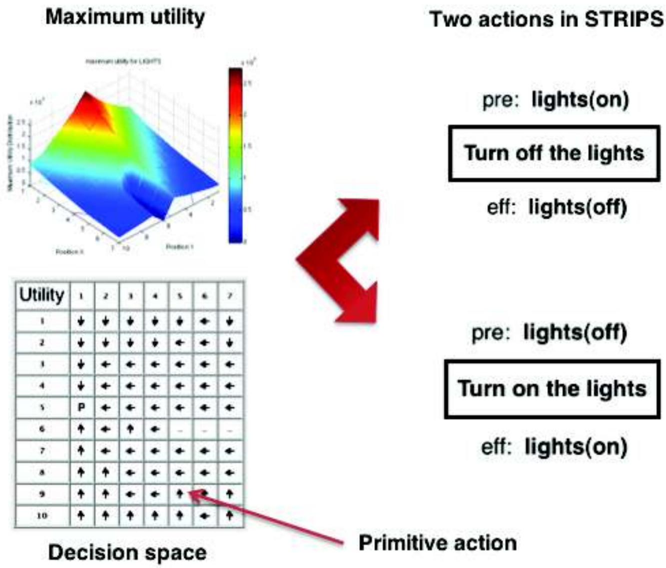


Fig. 2. Example of representing the decision space (abstract RL action) as two primitive actions in the format of STRIPS language.

The HARM action hierarchy serves as some "interface" between planning engine and outer world, this interface is created autonomously and adapts to the given problem online during the agents life.

The last part of architecture serves for learning of reflexive behaviour and gives the agent ability to react in selected situations with necessary speed of response. This subsystem is implemented by **Artificial Neural Network** and learns patterns situation-action. The *Artificial Neural Network* (ANN) learns appropriate behaviour from the HARM system. When the situation is considered as critical, the HARM system generates adequate action and the ANN learns this one pattern "actual situation"-"generated action". After some time, when the ANN shows learning error small enough, the agent can act reflexively. From this moment, the ANN can take the control over the agent and generate one primitive action in the critical situations. This system provides small reaction time where it is important and in fact increases the precision of action selection; this is caused by generalization ability of neural network.

3 Selected Experiments and Conclusion

We have conducted numerous experiments in order to test implementation of our architecture and to compare against other known approaches. The main focus was in testing the ability to effectively reduce the size of decision space (Fig.3), to act in dynamic environments (predator-prey simulation) and to test the ability to create and execute plans based on the knowledge gained autonomously.

In the Fig.3 we can see selected experiment, where the agent learned how (and when) to eat and drink. The total size of the decision space was increased by other moving agents (cows) to approximately 64×10^{12} states, the agent successfully learned that the positions of particular cows can be ignored, and thus was able to reduce the number of states to 2×100 states by identifying two important behaviours ("eat" and "drink") and creating corresponding RL decision spaces.

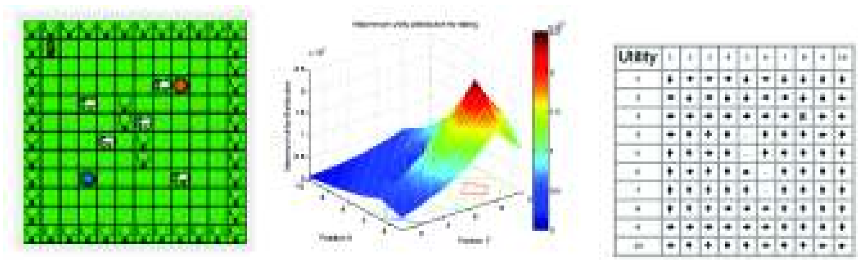


Fig. 3. Experiment testing the reduction of decision space size. The map (left) contains food (brown object) and water (blue) sources and the size of the decision space was increased by other moving agents (white cows). The goal is to learn how to survive - how to maintain the water and food levels in agent's body in bounds. During his life, the agent was able to successfully decompose his behaviour into two abstract actions, "eat" and "drink". For the "eat" behaviour, we can see the graph of maximum utility (and the table describing the best action) corresponding to the agents $\langle X, Y \rangle$ position, which means dramatic reduction of decision state space size compared to the flat RL approach.

None of the mentioned approaches could handle the all of the comparable problems alone, so the main contribution of this architecture is in combination of the advantages of particular subsystems and in their interconnection in such manner, that more complex behaviour can emerge from their mutual interaction.

We believe that the main advantage of our architecture is in combination of hierarchical reinforcement learning subsystem and the planning engine in such manner, so their collaboration provides ability of hierarchical planning while still maintaining the domain independency. This is because the fact that architecture configures itself based only on the knowledge autonomously gained from the particular domain. This means that, compared e.g. to the planner in widely used *Belief-Desire-Intention* architecture (BDI) [7], our agent does not need the predefined plan library - that is a knowledge a priory. Recently, some similar architectures, which try to combine planning and reinforcement learning (that is subset of our work) were found [9], [10], where the first one has the similar disadvantage as the BDI architecture mentioned above, and the second one uses slightly different approach. Based on our experiments concluded in order to compare it with other approaches, we believe that our architecture is superior to several widely-used principles in nowadays field of AI.

References

1. Kadleček, D.: Motivation Driven Reinforcement Learning and Automatic Creation of Behavior Hierarchies, CTU in Prague, FEE, PhD thesis, (2008)
2. Kadleček, D., Nahodil, P.: New Hybrid Architecture in Artificial Life Simulation. In: *Proc. of 6th European Conf. of Artificial Life: Advances in Artificial Life.* ECAL, Prague, LNAI Nr. 2159, vol. 1, Printed by Springer, Berlin, vol.1, 143-146 (2001)
3. Nahodil, P., Kadleček, D.: Adopting Animal Concepts in Hierarchical Reinforcement Learning and Control of Intelligent Agents, In: *Proc. of 2nd IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics, BioRob* 2008, Scottsdale, U.S.A, 122-131 (2008)
4. Fikes R., Nilsson N.: STRIPS: A New Approach to the Application of Theorem Proving to Problem Solving. *Artificial Intelligence*, 2, 189-208, (1971)
5. Sacerdoti E. D.: Planning in a Hierarchy of Abstraction Spaces. *Artificial Intelligence* 5(2), 115-135 (1974)
6. Erol, K., Nau, D., Hendler, J.: HTN Planning: Complexity and Expressivity. In *AAAI-94*, Seattle, (1994)
7. Sardina, S., de Silva, L., Padgham, L.: Hierarchical Planning in BDI Agent Programming Languages: a Formal Approach. *AAMAS '06 Proceedings of the fifth*

international joint conference on Autonomous agents and multiagent systems, 1001-1008. ACM New York, USA (2006)

8. Vítků, J.: An Artificial Creature Capable of Learning from Experience in Order to Fulfill More Complex Tasks, CTU in Prague, FEE, Diploma thesis, (2011)
9. Ryan, M., Pendrith, M.: RL-TOPs: An Architecture for Modularity and Re-Use in Reinforcement Learning, In *Proc. of the 15th International Conference on Machine Learning*, 481-487, (1998)
10. Ross, M.: Hierarchical Reinforcement Learning: A Hybrid Approach, PhD Thesis, The UNI of New South Wales, School of Computer Science and Engineering, (2002)

Medical AI – HIV/AIDS Treatment Management System

Miloslav Hajek[1] and Yashik Singh[2,3]

[1] School of Computer Science, University of KwaZulu-Natal, Westville, South Africa
mhajek@vodamail.co.za
[2] School of Information Systems and Technology, University of KwaZulu-Natal, Westville, South Africa
[3] Department of Telehealth, Nelson R. Mandela School of Medicine, University of KwaZulu-Natal, Durban, South Africa
singhy@ukzn.ac.za

Abstract. Medical AI has established itself as a robust and fruitful field in the last 30 years. Most resource poor countries face the triple burden of malaria, tuberculosis and HIV. This coupled with the problems of lack of infrastructure, scarcity of clinical staff, and complex clinical guidelines, have encouraged the application of AI in healthcare specifically on practical issues of field medical data collection, mining, and better integration with healthcare workflow. One such application is an HIV/AIDS antiretroviral therapy management system that uses AI algorithm to predict drug resistance and the progression of the disease. Another serious problem is the scarcity of personnel with sufficient AI knowledge in the medical field. A distance education has shown its potential to remedy the problem.

Keywords: medical AI, medical informatics, data mining, knowledge discovery, medical data, medical AI education, HIV antiretroviral therapy, drug resistance

1 Brief History

Medical informatics is the field of study that applies computer science techniques to medicine and one such application regards the use of AI. The first AI applications in medicine date back to 1970s when researchers tried to emulate the reasoning strategy of clinicians. Well known is an expert system MYCIN [1], which was proven to outperform a cohort of infectious disease experts. Unfortunately, MYCIN was never used in practice because of ethical issues and lack of foresight into integration of MYCIN into clinical workflow. Nevertheless, researchers from all areas of AI found medical applications interesting and challenging [2] and in 1989 a new Artificial Intelligence in Medicine (AIM) journal was launched. Participants at a panel discussion, presented at a biennial conference AIME'07, arrived at a conclusion that "AI in medicine field is robust and there is a clear evidence of progress" [3].

In the middle of 1990s a back-propagation learning algorithm was published and personal computers became readily available. Many new AI technologies, like neural networks, fuzzy systems, decision trees, Bayesian networks, and evolutionary and swarm intelligence algorithms, were not purely theoretical inventions anymore. They were built into user friendly software packages that had the capability to be adequately used on commonly available desktops. A door to wide scale applications of AI in traditionally nontechnical areas was opening [4], [5], [6], [7], [8], and [9]. The experimentation, implementation and use of AI in medicine were also positively influenced by the large amounts of data being collected by medical informaticians. This facilitated the process of using AI techniques to unveil hidden structures and relations in this data. This process came to be known as a data mining or a knowledge discovery process.

It is however unfortunate that AI in medicine is separate from medical informatics, bioinformatics, and telehealth. In order to truly take advantage of AI in medicine, there needs to be a mind shift and the field needs to be viewed as an essential component of health informatics. It must be integrated into the workflow and thus be identified as a methodology that can help to facilitate treatment of disease and solve other problems in healthcare on all levels [3], be it molecular and cell levels, organism level, health care level, or medical knowledge level.

2 Medical Data

Medical data are most rewarding and difficult to mine. Data are voluminous and heterogeneous (images, interviews, laboratory results, doctor observations and interpretations), contain unstructured free-text narrative written possibly in different languages, and many synonyms. Another reason that compounds the difficulty of mining medical data is the fact that there is no clear international standardization of the actual data itself and its representation. It is also imperative that researchers are cognizant of ethical, legal, and social constraints, like disclosure of private information and data ownership. On the other hand, a researcher can contribute to improving the quality and availability of medical care and thus save lives and/or improve their quality.

With the advent of digital pens, biosensing devices, portable data collection devices and better connectivity, there is an increase in the ability to gather new medical data but there is a much slower uptake in the advent of methods capable to deal with the resulting, gigantic data repositories [3]. It may be the case in genetics and drug development but not in field problems. Healthcare information systems frequently generate and maintain data necessary for daily operation but the data are not ready for data mining. There are two main reasons for that: the cost of an information system and scarcity of people who are aware of AI methods. Most medical information systems are not built with a data mining exercise in mind, which means that “field” knowledge is mostly lost.

Data collection and preparation for data mining is almost always the most difficult and time consuming part of data mining and knowledge discovery exercise. New intelligent data preparation techniques should be developed and built into health information systems.

3 Medical AI Technology Integration

AI researchers can be historically divided into two groups: pragmatists, for whom a system performance is more important than whether the system solves problems as human beings would, and formalists who argue that true AI requires modelling and insights into human intelligence [3]. In today’s world, people who operate effectively between both extremes are needed. Future cognitive computers may possess unlike-human intelligence but at the same time they will have to communicate with human experts and other healthcare personnel.

The evaluations of present decision support systems that failed show that is was not usually because of flawed technology but because of underestimating human issues in the design and implementation process. Should future AI technologies have their artificial thinking it would be paramount to build into them solid human communication enhancement. On the other hand, humans will have to do their bit and embrace and understand AI technologies more extensively.

4 Medical AI Education

The American Medical Informatics Association has clearly demonstrated that there is a shortage of skilled medical informaticians. They suggest that effort and resources need to be invested into education. It is believed that strong interdisciplinary education programs should be further fostered to improve the quantity and quality of researchers and practitioners and to help the dissemination of AI methods and principles in the biomedical and health care informatics community [3]. This was the driving force of launching a Postgraduate Diploma, Masters and PhD medical informatics programs at our university. The programs include an AI in medicine module. The authors' experience with those programs shows that students are either existing IT people or medical enthusiasts who are spatially distributed across sub-Saharan Africa. Thus distance education is the easiest option in delivering necessary information to build capacity in medical AI. It can hardly be expected that all graduates will become medical AI specialists but they will form a base that is needed for successful diffusion of AI techniques at all levels of medical care and research.

5 HIV/AIDS Treatment Management System

The current trend in patient healthcare is personalized medicine where treatment is individualized. Thus access and interpretation of personal patient information is vital in order to provide a sustainable and useful medical service. This is becoming more evident in the treatment of HIV/AIDS. Our research aims at developing a physician-administered AI-based decision support system tool that facilitates the management of patients on antiretroviral therapy.

HIV/AIDS is the leading cause of death in sub-Saharan Africa [10] and is one of the fastest growing epidemics in South Africa; currently there are 5.7 million confirmed cases. HIV infection can be effectively managed with antiretroviral (ARV) drugs, usually in the form of a highly active antiretroviral therapy (HAART), which consists of a regimen of three drugs from at least two of the following five drug classes: reverse transcriptase inhibitors (RTI), non-reverse transcriptase inhibitors (NRTI), protease inhibitors (PI), integrase inhibitors (II), and fusion inhibitors (FI).

Factors that influence treatment of HIV/AIDS with antiretroviral drugs include a treatment regimen prescribed by a physician, the stage of the disease (the progression of the disease), levels of drug concentration achieved, a patient's adherence to the regimen, drug resistance, and toxic effects of the drug. The drug resistance is arguably the most critical aspect of treatment and three common reasons that lead to the development of the resistance are high replication rates, selective pressure, and initial infection by resistant strains of HIV. Thus it is inevitable that drug resistance becomes a reality in most patients.

The design of a decision support system for the management of an antiretroviral therapy involves:

1. Development of an AI algorithm that analyzes HIV drug resistance data and provides interpretable information for a physician, indicating which ARVs a patient will be resistant to.
2. Using the AI algorithm to predict current and future CD_4 count (from a genomic sequence and other data).
3. Integration of the above tool with an electronic medical record such that it facilitates the storage, acquisition, and management of patient information.

The preliminary results in the development of such a management system are promising. A classification model was built to determine changes in CD_4 cell count. The changes in CD_4 (ΔCD_4) cell count to be predicted were grouped into four categories as shown in Equation 1. Three different groups of inputs where created and each was fed into the machine learning algorithm separately. These input groups were: Input 1, consisted only of genome sequence; Input 2, consisted of genome sequence and current viral load; Input 3, consisted of genome sequence, current viral load and number of weeks from the current CD_4 count to baseline CD_4 count.

$$Classification = \begin{cases} Output\ 1 & \text{if} \quad \Delta CD_4 < 0 \\ Output\ 2 & \text{if} \quad 0 \leq \Delta CD_4 \leq 50 \\ Output\ 3 & \text{if} \quad 51 \leq \Delta CD_4 \leq 100 \\ Output\ 4 & \text{if} \quad \Delta CD_4 > 100 \end{cases} \tag{1}$$

The model was built using a support vector machine and linear, quadratic (polynomial with degree two) and radial base function (RBF) kernels were used. The radial base function kernel with the parameters $cost = 3$ and $\gamma = 0.2$, and the polynomial kernel with the parameters $cost = 10$, constant polynomial coefficient of 1 and $\gamma = 1$ were determined by a coarse grid search.

The accuracy of the machine learning models is shown in Table 1. Results indicate that for the RBF, and linear and quadratic SVMs there are no differences between the Input 1 and Input 2 models, but there are differences between Input 1 and Input 3 models as well as the Input 2 and Input 3 models. The Input 3 model is more accurate than the other two models. This result was expected due to the fact that the longer a patient is on an effective ARV therapy the more the immune system reconstitutes, resulting in a higher CD_4 count. Thus, the time component is a valuable predictor. There is no difference between the quadratic and linear SVM algorithms as shown in Table 1, while the RBF model outperforms both the quadratic and linear SVMs with Input 3. It was unlikely that the data would be linearly separable hence the poorer performance of the linear SVM models. The superior performance of the RBF kernel is due to its localized and finite responses across the entire range of predictors.

Table 1. Accuracies of the different models

Input space model	RBF %	Quadratic %	Linear %
Input 1	66	66	66
Input 2	68	68	66
Input 3	83	72	71

6 Conclusion

Medical AI is a matured field that has established itself in medical research. Its full practical potential can be unleashed by seamless integration of AI applications into medical care workflow. Current applications of AI methods focus mainly on answering well-posed questions and are an important part of decision support systems. New AI methods that focus on helping the user to uncover new knowledge (discovery support systems) are yet to be developed. HIV/AIDS personalized treatment management system is an important example of using AI technologies in medical care. It is thus evident that integrating AI into medical workflow is essential for better health care delivery.

References

1. Shortliffe, E.H.: Computer-based Medical Consultations: MYCIN. Elsevier, New York (1976)
2. Szolovits, P. (Ed.): Artificial Intelligence in Medicine. Westview Press, Boulder, Colorado (1982)
3. Patel, V.L., Shortliffe, E.H., Stefanelli, M., Szolovits, P., Berthold, M.R., Bellazzi, R., Abu-Hanna, A.: The Coming of Age of Artificial Intelligence in Medicine, *AIM* 46, 1–96 (2009)

4. Dybowski, R.: Neural Computation in Medicine: Perspectives and Prospects. In: *ANNIMAB-1 Conference (Artificial Neural Networks in Medicine and Biology)*, 26–36, Springer (2000)
5. Abbod, M.F., von Keyserlingk, D.G., Linkens, D.A., Mahfouf, M.: Survey of Utilisation of Fuzzy Technology in Medicine and Healthcare. *Fuzzy Sets and Systems* 120, 331–349 (2001)
6. Bayesian Networks in Biomedicine and Health-care (Editorial). *AIM* 30, 201–214 (2004)
7. Pena-Reyes, C.A., Sipper, M.: Evolutionary Computation in Medicine: An Overview. *AIM 19*, 1–23 (2000)
8. Marghny, M.H., El-Semman, I.E.: Extracting Logical Classification Rules with Gene Expression Programming: Microarray Case Study. *AIML* 05 Conference, 11–16, Cairo (2005)
9. Marinakis, Y., Marinaki, M., Dounias, G.: Particle Swarm Optimization for Pap-smear Diagnosis. *Expert Systems with Applications* 35, 1645–1656 (2008)
10. Campbell, Y.N., Sbongile, M., Zweni, S.: Supporting People with AIDS and their Carers in Rural South Africa: Possibilities and Challenges. *Health and Place* 14, 507–518 (2008)

Patient Experience in e-Health ECA Applications

María C. Rodríguez-Gancedo[1], Javier Caminero[1], Beatriz López-Mencía[2], and Álvaro Hernández-Trapote[2]

[1] Telefónica I+D,
Distrito C-Oeste 1, Ronda de la Comunicación s/n, 28050 Madrid Spain
{mcrg,fjcg}@tid.es
[2] GAPS, Signal Systems and Radiocommunications Department, UPM,
28040 Avda Complutense 30, Madrid, Spain
{beatriz,alvaro}@gaps.ssr.upm.es

Abstract. This article is intended to open a discussion about the inclusion of Embodied Conversational Agents (ECAs) in e-Health applications. In order to that we identify several interaction and privacy aspects which could be relevant in case an ECA is used in long-term services dealing with sensitive user information. According this discussion we put forward the design of a questionnaire which permits to evaluate, in a functional prototype which monitors the treatment of chronic patients, the acceptance and privacy issues of e-Health services including ECA technology.

Keywords: e-Health services, embodied conversational agents, privacy, acceptance

1 Introduction

An Embodied Conversational Agent (ECA) – also known as Virtual Human [1], Companion [2] or Intelligent Virtual Agent [3] amongst other terms – is an animated human-like avatar capable, to some degree, of engaging in conversation with real users. This involves the ability to understand and generate speech, body gestures and facial expressions [4]. This technology, associated to the fields of Human Computer Interaction (HCI) and Artificial Intelligence (AI) has become very popular during the last few years due to its expressiveness abilities, empathic behaviour and social capabilities, which make them interesting in e-Health applications contexts. However, user acceptance of a new technology or application does not depend exclusively on technical functionality. In the case of e-Health applications there are mainly two aspects which are especially relevant: firstly, privacy and security perceptions which are motivated by the exchange of personal medical data; secondly, interaction issues about how to design a long-term interaction in order to get the goals of the application looking after the patient experience. In this paper, we analyse the impact of ECA technology in the context of applications focused in Health Services. For this we propose several open questions about how this technology could affect the user's privacy and the acceptance of new IT solutions to this area. As follow up work we plan to create a questionnaire in order to get insights for an application which controls the treatment of chronic patients.

2 How an ECA Could Be Useful in e-Health Applications

ECA technology intends to bring a more pleasant interaction to users, increasing their confidence, empathy and motivation during the interaction [5, 6], as well as trust in the system [7]. Based on these strengths, ECA technology may be useful in Health domains where long-term and trustworthy interactions are needed [7]. Remaining centred in the Health domain, we identify two points in which ECA technology have been considered useful:

1. *Doctor assistant.* Using the ECAs to give assistance and support to health workers, assisting in staff learning and training tasks on specific subjects. One example is the work carried out in [8], in which an ECA with "humane" and affective capabilities is used as a "virtual patient". In this application, doctors can learn how to give negative news to their real patients using the ECA as a virtual subject.
2. *Patient assistant.* From the patient's point of view, there are many applications that use the gestural and behavioural abilities of an ECA to act as a "virtual nurse". Thus, in [9], a hospital bedside patient education system for individuals with low health literacy is presented, focused on pre-discharge medication adherence and self-care counselling. In addition, there are studies in which several strategies of ECA behaviour are analysed in order to calm and comfort users [10]. Moreover, ECAs are used as personal trainers with the aim to help and motivate users to have healthy eating habits and to do physical exercise [11].

To sum up, it can be seen that ECA technology could bring trust, confidence and functional/emotional help in the interaction with e-Health systems.

Once we have identified those benefits, the next key question would be: *"How to evaluate the user acceptance of an e-Health application enriched with an ECA?"* According to the scope of the application, privacy and interaction issues should be taken into account in order to evaluate the whole "patient experience", augmenting the concept of "user experience" for traditional HCI system. It is a difficult task considering that there is not a universally accepted way for measuring privacy and acceptance objectively [12]. Therefore, in this approach we propose the creation of a questionnaire which contains different nuances about privacy and other ECA interaction issues in order to measure the impact of these beneficial effects.

3 Case Study

In order to obtain knowledge about the abovementioned issues we propose a working scenario which consists on a monitoring system of chronic patients. Through this system patients are able to control their medication at the same time doctors track the treatment. The application includes an ECA which plays the role of an intelligent assistant (adding a social and emotional component to the interaction). A mock-up of the proposed interface is shown in Figure 1.

3.1 Evaluation Proposal

Our proposal for the evaluation of the patient experience would consider the following three dimensions:

1. *Intimacy of use.* This aspect is two-fold: firstly, intimacy with regards to the people surrounding the patient when she uses the service and secondly, intimacy about who is granted access to patient's information (i.e. what doctors in particular). Regarding the former factor a question would be asked to patients about how they use the service at home (alone, someone helps them, etc.). This aspect is similar to "Willingness to share" studied in [13]. The latter factor could be measured in function of other parameters as the frequency of requests, with personal information, made to doctors (e.g. if requests were arranged indifferently to various doctors).
2. *Data safety and integrity.* If patients don't trust in the protection means installed against hackers then a decrease in the "Frequency of use" [13] of the service could be expected and measured.

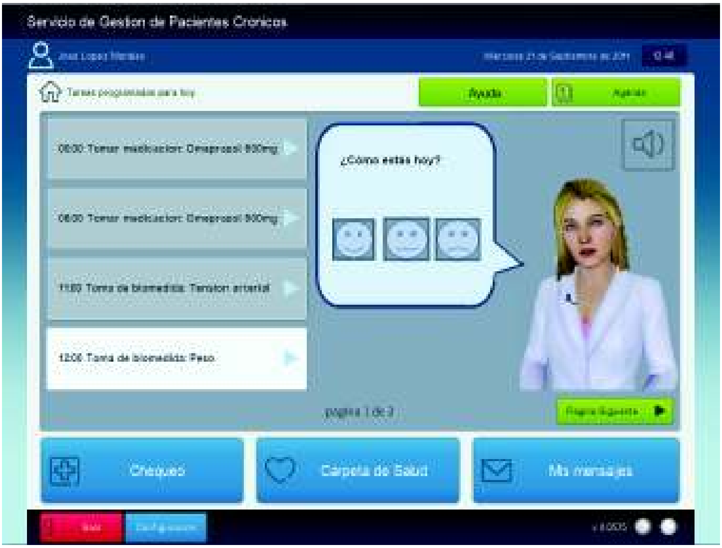


Fig. 1. Patient interface mock-up

3. Presence or behaviour of the ECA. The sole presence of the ECA, her comments or her way of interacting could affect the patients' privacy perception. Furthermore, the design of her behaviour is a key factor which has influence into the patients' perception about the usefulness of the service. We propose to measure these effects collecting users' opinions through questions in a long-term experiment, such as:
   (a) Would you want to keep the assistant?
   (b) Do you like the assistant's behaviour?
   (c) What would you change from assistant?

Additionally, this subjective information would be mixed with other objective metrics (i.e. how many questions proposed by the assistant are answered by patients).

4 Conclusions

In this work we open a discussion line in order to evaluate the patient experience for an e-Health application which incorporates an ECA. We have presented a preliminary evaluation proposal, in which we plan to take into account user's privacy concerns and other interaction information while they use a specific application, addressed to people with chronic diseases. Patients' necessity to make an intensive use of the application will afford us a very attractive long-term scenario. We plan to carry out a discussion of this proposal during the workshop "Beyond AI" and other related conferences. Thus, we expect to obtain fruitful results and feedback from other members of the specialized scientific community. In addition this work tries to shed some light in the evaluation procedures in e-Health services that incorporates virtual assistants considering the user privacy concerns.

Acknowledgments

The activities described in this paper were funded by the Spanish Ministry of Science and Technology as part of the TEC2009-14719-C02-02 project and from the European Community's Seventh Framework Program grant agreement number 258030 (FP7-ICT-2009-5).

References

1. Ruttkay, Z., Reidsma, D., and Nijholt, A.: Unexploited dimensions of virtual humans. In: *Proc. of the Int. Joint Conf. on Artificial Intelligence* (IJCAI'07) (2007)
2. Wilks, Y.: Artificial Companions as a new kind of interface to the future Internet (2006)
3. Rickel, J.: Intelligent virtual agents for education and training: Opportunities and challenges. In: *Intelligent Virtual Agents*, 15-22. Springer (2001)
4. Cassell, J.: Embodied conversational interface agents. *Communications of the ACM*, 43(4), 70-78 (2000)
5. Brave, S., Nass, C., and Hutchinson, K.: Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. *Int. J Hum.-Comput. Stud.*, 62(2), 161-178 (2005)
6. Hone, K., Akhtar, F., and Saffu, M.: Affective agents to reduce user frustration: the role of agent embodiment. *Proc. of HCI2003* (2003)
7. Bickmore, T. W.: Relational Agents: Effecting Change through Human-Computer Relationships. PhD thesis, Massachusetts Institute of Technology (2003)
8. Jung B., Ahad, A., and Weber, M.: The Affective Virtual Patient: An E-Learning Tool for Social Interaction Training within the Medical Field. *Proc. of Int. Conf. Training Education & Education (TESI)* (2005)

9. Bickmore, T., Pfeifer, L., and Jack, B.: Taking the Time to Care: Empowering Low Health Literacy Hospital Patients with Virtual Nurse Agents. *Proc. of the ACM SIGCHI Conf. on Human Factors in Computing Systems* (CHI) (2009)
10. Bickmore, T. and Schulman, D.: Practical Approaches to Comforting Users with Relational Agents. *SIGCHI Conf. on Human Factors in Computing Systems* (2007)
11. Turunen, M., Hakulinen, J., Stahl, O., Gambäck, B., Hansen, P., Gancedo, M. C. R., de la Cámara, R. S., Smith, C., Charlton, D. & Cavazza, M.: Multimodal and mobile conversational Health and Fitness Companions. *Computer Speech & Language* 25(2), 192 - 209 (2011)
12. Toledano, D.T., Fernández, R., Pozo, A. Hernández Trapote and Hernández Gómez, L.: Usability Evaluation of Multi-Modal Biometric Verification Systems. *Interacting With Computers* 18(5), 1101-1122 (2006)
13. Braunstein A., Granka L. and Staddon J.: Indirect Content Privacy Surveys: Measuring Privacy without asking about it. Simp. On Usable Privacy and Security (2011)

Voice Conservation: Towards Creating a Speech-Aid System for Total Laryngectomees

Zdeněk Hanzlíček and Jindřich Matoušek

Dept. of Cybernetics, Faculty of Applied Sciences, University of West Bohemia
Univerzitní 8, 306 14 Plzeň, Czech Republic
{zhanzlic,jmatouse}@kky.zcu.cz

Abstract. This paper describes the initial experiments on voice conservation of patients with laryngeal cancer in an advanced stage. The final aim is to create a speech-aid device which is able to "speak" with their former voices. Our initial work is focused on applicability of speech data from patients with an impaired vocal tract for the purposes of speech synthesis. Preliminary results indicate that appropriately selected synthesis method can successfully learn a new voice, even from speech data which is of a lower quality.

Keywords: speech aid, total laryngectomy, voice conservation, speech synthesis

1 Introduction

Laryngectomees are people who underwent a laryngectomy surgery. This medical intervention is performed on patients with laryngeal cancer when other types of treatments (e.g. radiation or chemotherapy) fail or are not possible. According to the extent of the carcinoma various sections of larynx are removed. In the case of total laryngectomy, the removal of the whole larynx together with the vocal folds is performed. Then, respiratory and digestive tracts are separated from each other. The laryngectomee breathes through a stoma – an opening in the trachea.

A significant consequence of this surgery is the inability to produce speech in the common manner. However, there are several alternatives for producing speech sounds:

– **tracheo-esophageal speech** – a special voice prosthesis need to be surgically placed between trachea and esophagus, it contains a one-way valve that allows the air to flow from lungs into the oral cavity. Tracheostoma has to be plugged during the speech production.

– **esophageal speech** – the air is swallowed into the esophagus and then it is pushed back into the oral cavity for articulation. This method is very exacting due to the low capacity of esophagus.
– **electrolaryngeal speech** – the function of vocal folds is substituted by an external device (electrolarynx) which is put to the neck where produces mechanical vibrations while the speaker articulates.

All the aforementioned kinds of speech (also called alaryngeal speech) suffers from lack of naturalness and speaker identity.

Recently, a new important task – alaryngeal speech enhancement – is solved in the field of computer speech processing [1, 2]. The objective is to improve the sound quality and recover the speaker identity. The final aim is to create a speech-aid device which allows the laryngectomee to communicate with a more natural voice.

From the practical point of view, the source alaryngeal speech should not disturb and coincide with the final enhanced speech. This could be done by rotating stages of alaryngeal and enhanced speech.

Another solution is using the NAM (non-audible murmur) microphone that is able to detect so-called body conducted speech. It is caused by vibrations of air in the vocal tract passed on the soft tissues of the head or neck. The standard electrolarynx excitation is swapped with a source of small-power vibrations. Thus the produced alaryngeal speech is nearly inaudible.

In fact, there are two basic approaches to design a speech-aid system for laryngectomees:

– Using **speech recognition**. Produced speech is first recognized by a specially designed ASR (automatic speech recognition) system. A text of the utterance is extracted this way. Then this text is synthesised by a new voice. The knowledge of the utterance content allows to add some higher speech properties, e.g. the course of the fundamental frequency. An important disadvantage of such a speech-aid system is a delayed response caused by speech recognition process. The resulting delay could be similar as in the case of human simultaneous translation.
– **Speech signal transformation** without speech recognition. Spectral characteristics of alaryngeal speech are converted and the enhanced (i.e. more human like) speech is reconstructed from those characteristics.

2 Voice Conservation

A natural requirement for any speech-aid system is to produce a voice which is close to the former voice of each laryngectomee. Most speech synthesis methods are able to produce voice with required speaker identity. However, a quite huge amount of training speech data is necessary to learn this voice. In the case of a laryngectomee, to obtain a sufficient amount of speech could be a substantial problem.

Naturally, speech recorded by the healthy voice (before the disease has broken out) would be preferred. Some people could have recordings related to their job, e.g. various speeches, performances, presentations, or also some personal recordings, e.g. family events, reading fairy tales to their children, etc. Unfortunately, the acoustic conditions of such audio data are often not optimal and it is not suitable for the purposes of speech synthesis. Moreover, most people do not have any usable recordings at all.

Another solution is to record the speech data after the diagnosis before surgery. However, in those stages of disease, the vocal tract is usually significantly damaged which causes various speech problems. The overall voice quality is poor or unstable and the speaking could be very exhausting for the patients. The voice could be also affected by the mental condition of patients because they are often significantly stressed by their diagnosis and expected surgery. The last chance is to acquire recordings from another willing person with a similar voice.

The process of obtaining and storing speech data of the patients can be called voice conservation. It is only one fundamental step in solving the complex problem of developing a speech-aid device for laryngectomees.

3 Speech Synthesis

Modern speech synthesis methods (e.g. unit selection or statistical parametric speech synthesis) need a large amount of training data (several hours of speech) to create a new system producing the desired voice in a high quality. The standard recording procedure is demanding even for a healthy professional speaker because high-quality data are required for the purposes of speech synthesis. Thus each utterance is repeated until it is perfectly pronounced. The overall recording process lasts for several weeks.

It would be nearly impossible to record such a huge amount of speech data by a patient with advanced laryngeal cancer. Fortunately, there are several alternatives with lower demands on speech data. The most promising one is using adaptation methods [3] within the statistical parametric speech synthesis (also known as HMM-based speech synthesis [4]). This synthesis method employs statistical models (hidden Markov models, HMMs) to represent the statistical acoustic features of speech. These models are trained by using speech data of the desired speaker. During synthesis, speech is generated from those trained models.

Model adaptation is a transformation of models from one voice to another. Source models are trained from speech data of a professional speaker, or data from more speakers can be used to trained so-called average models. For the model adaptation significantly less speech data is needed. Moreover, a great advantage of the adaptation is the lower sensitivity to the quality of the source speech data.

4 Preliminary Experiments

For our first experiments we apply an experience from our previous work in the domain of statistical parametric speech synthesis [5, 6].

In cooperation with the Motol University Hospital one voluntary female patient with laryngeal cancer diagnosis was selected and her speech recorded. Recording conditions (similar to those described in [7]) was acoustically perfect, an anechoic chamber for acoustical measuring and experiments was used.

The recording process had to be adjusted to the specific condition of the patient. Several concessions had to be done, otherwise it would be unfeasible to record required amount of speech during one session. Consequently, the recorded utterances contains various stumbles, unexpected pauses and voice failures. About 500 utterances (approx. 1 hour of speech) were recorded during this session.

In our first experiments we used about 30 minutes of speech to adapt models trained from 5 hours of speech from a professional female speaker. The speech produced by using adapted models was definitely identified as the female patient. Considering the utilised data, the quality was also acceptable.

5 Conclusion

Our first experiments are promising. Voice conservation of patients with laryngeal cancer diagnosis, even when their speech is of lower quality, opens the possibility to create a speech-aid device producing the former personal voice. Although the way to develop such a device is still long and a lot of research work has to be done, the current results can be already practically utilised. Laryngectomees can run the speech synthesizer with their own voices on their computers. This could be helpful in the post-operative stage when the possibilities of inter-personal communication are very limited which could be frustrating.

References

1. Doi, H., Nakamura, K., Toda, T., Saruwatari, H. and Shikano, K.: An Evaluation of Alaryngeal Speech Enhancement Methods nased on Voice Conversion Techniques. In: Proceedings of ICASSP 2011, pp. 5136–5139 (2011)
2. Nakamura, K., Toda, T., Saruwatari, H. and Shikano, K.: Speaking Aid System for Total Laryngectomees using Voice Conversion of Body Transmitted Artificial Speech. Proceedings of Interspeech 2006, pp. 1395-1398 (2006)
3. Yamagishi, J., Kobayashi, T., Nakano, Y., Ogata, K. and Isogai, J.: Analysis of Speaker Adaptation Algorithms for HMM-Based Speech Synthesis and a Constrained SMAPLR Adaptation Algorithm. In: IEEE Transactions on Audio, Speech, and Language Processing, vol. 17, pp. 66–83 (2009)
4. Zen, H., Tokuda, K. and Black, A. W.: Review: Statistical parametric speech synthesis. *Speech Communication*, vol. 51, 1039–1064 (2009)
5. Hanzlíček, Z.: Czech HMM-Based Speech Synthesis. In: Sojka, P., Horák, A., Kopeček, I., Pala, K. (eds.) *Text, Speech and Dialogue, LNCS (LNAI)*, vol. 6231, 291–298, Springer Berlin/Heidelberg (2010)
6. Hanzlíček, Z.: Czech HMM-Based Speech Synthesis: Experiments with Model Adaptation. In: Habernal, I. and Matoušek, V. (eds.) *Text, Speech and Dialogue, LNCS (LNAI)*, vol. 6836, 107–114, Springer Berlin/Heidelberg (2011)
7. Matoušek, J., Romportl, J.: Recording and Annotation of Speech Corpus for Czech Unit Selection Speech Synthesis. In: Matoušek, V. and Mautner, P. (eds.) *Text, Speech and Dialogue, LNCS (LNAI)*, vol. 4629, 326–333, Springer Berlin/Heidelberg (2007)

The Fourth Dimension of AmI: Co-existence

Peter Mikulecký, Petr Tučník

Faculty of Informatics and Management
University of Hradec Králové
50003 Hradec Králové
Czech Republic
{peter.mikulecky,petr.tucnik}@uhk.cz

Abstract. There are three essential dimensions of the area of Ambient Intelligence (AmI) that can be naturally considered as being essential: technological, social, and political. However, if we are viewing an intelligent environment as a collection (if not a community) of intelligent artefact, capable of mutual communication and performing activities based on a kind of mutual coordination, then another - fourth - dimension could be useful for further contemplations about ethical issues or privacy questions. We shall call this fourth dimension as co-existential one. The paper is focused on presenting our opinions and open questions in this direction.

Keywords: ambient intelligence, smart artefacts, ethics, privacy, social acceptance

1 Introduction

When speaking about Ambient Intelligence (AmI), usually we take into consideration its three main dimensions that are considered as being essential:

- technological,
- social,
- political.

In our opinion this is not enough as there are certainly other aspects of AmI area that are not covered by the three dimensions. This is the main motivation for our paper, where we intend to discuss a kind of the fourth dimension that we call "*co-existential dimension*". If we consider the problem of introducing Ambient Intelligence in an environment, we have to take into consideration that what we are creating in such a case is an environment based on mutually communicating intelligent artifacts. These intelligent artifacts create in a sense a community which is supposed to communicate with humans throughout this environment and accomplish a number of activities aiming to support relevant human activities. In other words, in such an environment we should take into account a number of aspects of mutual co-existence of human beings with these intelligent artifacts, or more precisely, with a community of these artifacts.

In what follows, we wish to discuss some ethical and privacy issues of Ambient Intelligence, related to this fourth dimension of AmI.

2 Three Dimensions of AmI

There are three main dimensions to the problem of ambient intelligence, usually considered as being essential (cf. [4]):

- technological,
- social,
- political.

Within the *technological dimension* it is necessary to study and develop technical devices, information, knowledge and communication technologies which will make the implementation of the vision of ambient intelligence possible. Key technologies might, among others, include also knowledge management, artificial intelligence, user interfaces, communication and network services, as well as solution to the problems of security and protection of data and information.

The *social dimension* focuses on studying the influences of social, economic and geopolitical trends on the quality of everyday life and the acceptance of using solutions employing information and communication technology (ICT) for solving problems accelerated by the above mentioned trends. Areas of problems of more global nature include for example ageing of society, multicultural society, lifelong education, the problem of consumer society, globalization, etc.

The *political dimension* of the problem of ambient intelligence has its starting point in the resolution adopted at the Lisbon congress of the EU in 2000, on the basis of which the European Commission resolved to secure Europe's leading role in the field of generic and applied technologies for creation of knowledge society, and thus increase Europe's ability to compete successfully, and enable all European citizens to take advantage of the merits of knowledge society. To this effect, the new technologies must not be the cause for excluding some groups of citizens from society, but must ensure universal and equal approach to its - both digital and therefore also knowledge - sources.

The concept of AmI is strongly motivated also by economic aspects - probably economic motivation is the most significant incentive in this area. A discussion about real time or now-economy has been presented by Bohn and his colleagues [2], where more and more entities in the economic process, such as goods, factories, and vehicles, are being enhanced with comprehensive methods of monitoring and information extraction.

3 The Fourth Dimension

Undoubtedly there is also another important dimension to the problem of ambient intelligence. Let us call it the *co-existential dimension*. This co-existential dimension should be focused on the problem arising from the relatively simple fact that various information devices integrated into people's everyday life represented with their intelligent interfaces capable to communicate with people, can be understood as relatively independent entities with certain degree of intelligence. Their intelligence varies, of course, from rather simple level of one-purpose machines to relatively intelligent and complex systems (e.g., an intelligent building, or an intelligent vehicle). These intelligent entities co-operate one with another, and all of them from time to time have to co-operate with humans.

Considering humans as another entity with various degree of intelligence, we are able to study the co-existence of various intelligent entities in real world. This will lead to an investigation of a number of different interesting aspects of such a co-existence, and also to a number of potentially important consequences for human lives.

When taking into account such artificial entities with a certain degree of intelligence and with a mechanism for the initiation of its activity, where the activity is oriented on certain benefit (or service) to human beings, we are able to investigate the following basic problems related to them:

- various types or levels of such artificial entities,
- their mutual relationships as well as their relationships with humans,
- their communities (virtual as well as non-virtual),
- their co-existence, collaboration and possible common interests,
- their co-existence and collaboration with humans,
- antagonism of their and human interests,
- ethical aspects of the previous problems, etc.

All of these are interesting sources of a plethora of serious scientific questions. Some of them are discussed recently, however, a number of them remain unanswered. We hope for a further discussion in this direction.

The first impression from the AmI idea is, that humans are surrounded by an environment, in which there are microprocessors embedded in any type of objects - in furniture, kitchen machines (refrigerator, coffee maker, etc.), other machines (e.g., washing machine, etc.), clothing, toys, and so on. Of course it is depending on the type of the particular environment, there are clear differences between an environment in a hospital when compared with a luxurious private house, or in comparison with a university environment.

It is straightforward that when speaking on intelligent artificial entities able to communicate mutually, we could certainly expect some relatively intelligent behavior of such a community. We can speak about the emergent behavior of such a community that can be modeled by a multi-agent system, serving to some purpose considered to be beneficial for humans. However, the emergent behavior of such an artificial community can be potentially dangerous - if the possible goal of the community differs from the human interests, or if the community is simply unable to serve to the human being goals from various (maybe also technical) reasons. We certainly have to take into account such questions, like:

– How to tune all the emergent behavior of the particular environment to be able to serve the particular human being goals?
– What to do if the emergent behavior of the environment is not in accord with the human aims, or even if it is contradictory to the intentions of the particular human?
– How the privacy of a particular human will be respected in an intelligent environment?
– Is the particular information about the concerned human safe from being exploited by another person?

Of course, these are just a few of possible questions which could arise in relation to the first attempts to introduce the AmI idea into the life. Some of other issues certainly will be mentioned in the future.

4 Conclusion

As Bohn and his colleagues pointed out [2], the fundamental paradigm of ambient intelligence, namely the notion of disappearing computer (computers disappear from the user's consciousness and recede into the background), is sometimes seen as an attempt to have technology infiltrate everyday life unnoticed by the general public in order to circumvent any possible social resistance [1]. However, the social acceptance of ambient intelligence will depend on various issues, sometimes almost philosophical ones. The most important issue seems to be our changing relationship with our environment.

Our approach based on rather wide employment of ambient intelligence technology [3] opens also a number of related ethical and privacy questions which must be solved simultaneously with introducing of the technology. We have to analyze the most important from a big variety of such questions, and bring solutions important for further development of this serious and highly applicable area.

Acknowledgments

The research has been partially supported by the Czech Scientific Foundation project No. P403/10/1310 SMEW.

References

1. Araya, A.: Questioning ubiquitous computing. In: *Proc. of the 1995 ACM 23rd Ann. Conf. on Computer Science*. ACM Press, New York (1995)
2. Bohn, J., et al.: Social, economic and ethical implications of ambient intelligence and ubiquitous computing. In: *Ambient Intelligence*, 8–29. Springer-Verlag (2005)
3. Mikulecký, P.: Remarks on ubiquitous intelligent supportive spaces. In: *Recent Advances in Applied Mathematics and Computational and Information Sciences: Proc. of the 15th American Conference on Applied Mathematics and Proc. of the International Conference on Comp. and Information Sciences*, Vols. I and II, WSEAS Press, Athens, 523–528 (2009)
4. Skordas, T., Metakides, G.: Major challenges in ambient intelligence. *Studies in Informatics and Control* 12(2), 85–89 (2003)

Analogy, Aesthetics, and Affect: What HCI Designers Can Learn from AI

William W. York and Hamid R. Ekbia

Indiana University
1320 E. 10th St., Bloomington, IN, 47405-3907, United States
{wwyork,hekbia}@indiana.edu

Abstract. Recent interest in the aesthetic and affective aspects of design within Human-Computer Interaction (HCI) has generated more questions than answers. In particular, there is a dearth of understanding of how these aspects of design are operationalized, how they relate to use, and how they relate to each other. We propose a third aspect of design – analogies – as a useful way to think about these questions. As close relatives of metaphors, analogies have been a topic of discussion in HCI but are rarely invoked in relation to affect and aesthetics. Building on a view of analogy making as an aesthetically guided process of slippage and seeing as, we argue that analogies are central to design and use, providing a link between the aesthetic and affective properties of artifacts. In doing so, we draw on examples from AI research as well as different areas of creative activity.

Keywords: human-computer interaction, analogy, aesthetics, design, humor, affect

1 Introduction

The central problems questions of aesthetics have challenged philosophers for centuries. Hume's [1] influential argument about the "standard of taste," for instance, is frustratingly circular, locating the standard in the judgment of "ideal critics," while at the same time identifying those critics via the quality of their judgments (cf. [2], [3], [4]). Equally vexing issues include the relationship between aesthetic and non-aesthetic qualities [5] and the question of whether there are general criteria for assessing aesthetic value [6]. Research in AI may not be able to resolve these questions as they have been formulated by philosophers, but it provides an opportunity to revisit these issues in a generative manner – that is, through the creation of models that can reveal some of the underlying processes behind aesthetic judgment. Analogy-making, we argue, is one such process.

Analogies have been the topic of study in many AI theories and models. Among these, one group of models has explored analogies from an aesthetic perspective. Developed by the Fluid Analogies Research Group (FARG), these models typically operate in highly constrained microdomains rather than real-world domains [7]. The Letter Spirit model, for instance, deals with the design of alphabetic fonts, a subject that raises a number of elusive questions about visual style, coherence, and sameness [8]. Another model, Copycat [9], deals with a domain whose relation to aesthetics is less immediately apparent: letter-string analogy puzzles. For example, if **abc** changes to **abd**, how would you change **ijjkkk** in "the same way"? There is no single "correct" answer, yet certain ones (e.g., **ijjkkkk** and **ijjlll**) are certainly preferable to others (e.g., **ijjddd** and **ijjkkl**). These preferences, in turn, are largely based on aesthetic factors (e.g., depth, elegance, ugliness, etc.). Copycat's successor, Metacat [10], explores these evaluative aspects of analogy-making in more depth. While far from solving the problem of how aesthetic judgment relates to analogy-making, the program is modestly successful at comparing and evaluating answers along three dimensions: *uniformity*, *succinctness*, and *abstractness*. Compared to more traditional research on analogy (e.g., [11], [12]), these projects suggest a more inclusive and open-ended portrayal of analogy-making, its role in everyday thought, and its relation to aesthetic sensibilities.

Because of its focus on microdomains, FARG research is at odds with mainstream AI, which focuses on real-world application domains. Nonetheless, these models embody the understanding-by-building ethos that initially drove AI research and brought it into contact with cognitive psychology decades ago. In particular, they highlight the importance of *conceptual slippage* in creative analogy-making. Conceptual slippage is the process whereby one concept can slip to (or be replaced by) another item in its "halo" of related concepts, given sufficient contextual pressure. In the aforementioned Copycat example, the answer **ijjkkkk** is based in part on seeing the string **ijjkkk** in terms of ascending group lengths – that is, seeing **i-jj-kkk** as 1-2-3. Whereas the change from **abc** to **abd** in the puzzle's prompt is based on the concept of *alphabetic* successorship, the change from **ijjkkk** to **ijjkkkk** is based on the parallel concept of *numeric* successorship. There is a depth to this answer that is lacking in the otherwise decent **ijjlll**, which does not incorporate the *alphabetic–numeric* slippage.

The slippage process, we would like to argue, is also central in situations and activities that involve interaction with artifacts. Our argument has four key components: (i) Slippage is prevalant not only in cognition, but also in perception and action; (ii) Slippage is a central underlying process in both error-making and creativity, and creative professions (visual artists, filmmakers, comedians, etc.) make prominent use of slippage and analogies in their work; (iii) There is a close relationship between slippage, analogy-making, and the affordances of artifacts; and (iv) Designs of artifacts (e.g., in HCI) can be evaluated on the basis of how sensibly and judiciously they incorporate slippage. Here we would like to present these arguments by examining the processes of slippage in some of these areas.

2 Perceptual Slippage and Visual Analogy

A central tenet of FARG research is that analogy-making can itself be viewed as a perceptual process – a kind of metaphorical "seeing as." But the analogy-as-perception idea applies equally well when "perception" is used, in the more traditional sense, to refer to specific sensory modalities such as vision or hearing.

Visual analogies come in many forms. Some are based on likeness in terms of relatively basic attributes (size, shape, color), while others are based on more abstract or gestalt relationships. For instance, a French coffee press and a piston are configured in similar ways but have starkly different functions. To liken one to the other (as a friend recently did, jokingly, in describing the operation of a French press in terms of thermodynamics) is to highlight their configurations at the expense of their underlying functions.

One interesting form of visual analogy is the visual pun, which Arthur Koestler (using the term "optical pun") described as follows: "The sleeper producing a Freudian dream, in which a broomstick resembles a phallus, has made an optical pun: he has connected a single visual form with two different functional contexts" ([13], p. 182); likewise for "the caricaturist who equates a nose with a cucumber, the discoverer who sees a molecule as a snake, the poet who compares a lip to a coral" (ibid.). Visual puns and analogies surface frequently in surrealist art, among other domains. Salvadore Dali's *Mae West*, for example, makes use of visual analogies – between a pair of windows and a pair of eyes, and between a set of curtains and a woman's hair – to form the basis for a more elaborate "blend" (cf. [14]) involving the wall of a room and a human face.

3 Action Slips and Creativity

Slippage is prevalent not only in perception but in action as well. Studies of error-making in everyday activity offer numerous examples of such errors: subjects flicking cigarette ashes into their coffee cups rather than their ashtrays; mistaking a can of rice pudding for cat food; or pouring tea into the milk container ([15], p. 95–96).

Just as Freud [16] saw slips of the tongue as portals into the "psychopathology of everyday life," more recent researchers such as [15] and [17] have used action slips as windows into the mechanisms underlying skilled behaviors, ranging from the mundane (making tea) to the highly specialized (flying an airplane). This research has had immediate practical applications, from improving product design to preventing disastrous errors (e.g., airplane crashes, nuclear power meltdowns). And just as slips of the tongue often reveal close ties to the processes involved in creative thought, action slips can be also striking in their seeming ingenuity.

For example, one subject in [17] reported accidentally throwing a dirty T-shirt into the toilet bowl – rather than the intended laundry hamper, which was in another room. Another subject reported stopping his car and then proceeding to unbuckle his *wristwatch* rather than his seatbelt. The perceptual slippages involved in these errors – *watchband* for *seatbelt* and *toilet bowl* for *laundry hamper* – have much in common with the kinds of *conceptual* slippages that play a central role in the kind of creative analogy-making modeled in Copycat and Metacat. The *seatbelt–watchband* slip, for example, is actually supported by a clear and coherent set of mappings: from *waist* to *wrist*, from *seatbelt buckle* to *watch buckle*, and from *seatbelt webbing* to *watchband*. While the action was unintended, there is an aptness to this "analogy" that renders it worthy of appreciation in the same way that a good joke does.

These examples illustrate the fine line that separates inadvertent slips from creative slippages. This close relationship is illustrated by the prevalence of action slips in comedies ranging from *The Three Stooges* to *The Naked Gun*. Comedy writers play on our tendencies to commit as well as to notice (and laugh at) such errors, but these writers are being creative in their own right when they concoct these scenes. Setting these fictional and real-life examples next to one another, we can see that the difference between unintended slips and purposive, creative slippages is often simply a matter of context (e.g., fictional vs. real life) or point of view.

4 Affordances and Slippage: Artifacts in Use

The use of artifacts also involves frequent slippages, sometimes as creative work-arounds by users who seek to get a job done by an artifact not intended for that job, and other times by misperception of the artifact's affordances. The notion of *affordances* was originally put forth by Gibson [18], who explained, "The affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill" (p. 127). This notion was later appropriated by HCI practitioners, who expanded it to the realm of artifacts: "Affordances provide strong clues to the operations of things. Plates are for pushing. Knobs are for turning. Slots are for inserting things into" ([19], p. 9).

Despite these clues, affordances can lead us astray. Action slips often result from misperceived (or carelessly perceived) affordances. On the one hand, a given object or material can afford unintended actions (think back to the tuba that Leslie Nielsen's bumbling *Naked Gun* character regurgitated into); on the other hand, two different objects can afford the same action, which can also pave the way for unwitting slips as well as purposeful slippages. A comical example of this occurs in Mel Brooks' *High Anxiety* (1977), in which Brooks' character is harmlessly "stabbed" in the shower by a disgruntled bellhop using a rolled-up newspaper as a weapon. This is a parody of the famous shower scene in Alfred Hitchcock's *Psycho* (1960), in which the main character is fatally stabbed in a motel shower. Semantically, *knife* and *newspaper* are rather distant concepts; it is only because a rolled-up newspaper *affords* stabbing (albeit not very *effective* stabbing) that this aspect of the scene makes sense.

While we can appreciate slips and slippages of various sorts in visual art, on the movie screen, and even in studies of human error-making, they are not always desirable. In HCI, good design makes effective use of analogy and slippage, incorporating useful conceptual and perceptual mappings without leading the user in the direction of unintended or harmful slippages. Likewise, good design facilitates fulfilling, user-friendly experiences, which is to say that it also has an affective component [20]. Take, for example, the computer desktop metaphor, which is so familiar by now as to hardly seem metaphorical at all. The surprising complexity of this metaphor is illustrated in [14], though the authors point out that certain aspects of it initially proved disturbing and counter-intuitive to users – most notably the practice of dragging the floppy disk icon to the trashcan in order to eject the physical floppy disk from the disk drive (a practice that was, of course, eventually dropped).

The goal of our research, building on the work of others described in this article, is to deepen our understanding of the relationship between the three

“A”s in our title, as well as to develop heuristics for the use of analogies in the process of affective and aesthetic design in HCI.

References

1. Hume, D.: Essays Moral, Political, and Literary, Vol. 1. Longmans, Green, and Co., London (1875)
2. Marshall, D.: Arguing by analogy: Hume's standard of taste. *Eighteenth-Century Studies* 28(3), 323–343 (1995)
3. Shelley, J.: Hume's Double Standard of Taste. *Journal of Aesthetics and Art Criticism* 52(4), 437–445 (1994)
4. Herwitz, D.: Aesthetics: Key Concepts in Philosophy. Continuum, London (2008)
5. Sibley, F: Aesthetic Concepts. Philosophical Review, 68, 421–450 (1959)
6. Beardsley, M. Aesthetics. Hackett, Indianapolis (1958)
7. Hofstadter, D. R.: Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought. Basic Books, New York (1995)
8. Rehling, J.: Letter Spirit (Part Two): Modeling Creativity in a Visual Domain. Doctoral dissertation, Indiana University, Bloomington (2001)
9. Mitchell, M.: Analogy-Making as Perception. MIT Press, Cambridge, Mass. (1993)
10. Marshall, J.: Metacat: A Self-Watching Cognitive Architecture for Analogy-Making and High-Level Perception. Doctoral dissertation, Indiana University, Bloomington (1999)
11. Gentner, D.: Structure-mapping: A theoretical framework for analogy. *Cognitive Science* 7(2), 155–170 (1983)
12. Holyoak, K. J., Thagard, P.: Analogical mapping by constraint satisfaction. *Cognitive Science* 13, 295–355 (1989)
13. Koestler, A.: The Act of Creation. Hutchinson, London (1964)
14. Fauconnier, G., Turner, M.: The Way We Think. Basic Books, New York, 2002.
15. Reason, J., Mycielska, K.: Absent-Minded? The Psychology of Mental Lapses and Everyday Errors. Prentice-Hall, Englewood Cliffs, N.J. (1982)
16. Freud, S. The Psychopathology of Everyday Life. Norton, New York (1960)
17. Norman, D.: Categorization of Action Slips. *Psychological Review* 88(1), 1–15 (1981)
18. Gibson, J.: The Theory of Affordances. In: Shaw, R., Bransford, J. (eds.) Perceiving, Acting, and Knowing. Lawrence Erlbaum, Hillsdale, N.J. (1977)
19. Norman, D.: The Psychology of Everyday Things. Basic Books, New York (1988)
20. Norman, D.: Emotional Design. Basic Books, New York (2005)

Nonlinear Trends in Modern Artificial Intelligence: A New Perspective

Elena N. Benderskaya

Faculty of Computer Science, St. Petersburg State Polytechnical University
Politechnicheskaya 21, 194021, St.Petersburg, Russia
helen.bend@gmail.com

Abstract. The paper faces the challenge to generalize existing trends and approaches in the field of artificial intelligence. The most perspective research directions are revealed through multidisciplinary analysis. It was shown that dynamic models with unstable dynamics are logical continuation of approaches sequentially becoming more complex. Conclusion is that in the nearest future nonlinear dynamics and chaos will become the most demanded apparatus to understand and model cognition processes.

Keywords: multidisciplinary analysis, distributed AI, bio-inspired algorithms, nonlinear dynamics, chaotic neural networks, self-organization

1 Introduction

This paper analyses trends of modern approaches (including methods partially) in artificial intelligence (AI) and change (transformation) of the multidisciplinarity role and its level necessary in order to attain a qualitative leap in AI area. It is still a matter of future research to create true human intelligence although a lot of promising results in some local AI areas were published [5, 11, 19]. We aim to generalize up to date accumulated experience in intelligent systems development. To reach this aim it is necessary to stay above separated and fragmented multidisciplinary studies and to find the common approach which will let to aggregate them in the way to obtain synergy effect. We focus on the essence of AI paradigms and discuss on promising directions in developing of true human intelligence.

Two main approaches within AI field, namely Symbolic AI (associated with experts systems and knowledge bases, etc.) and Connectionist AI (associated with neural networks (NN) and implicit knowledge learning, etc.) no more compete but supplement each other [8, 14], e.g. in the form of neuro-fuzzy methods when both learning processes and explicit knowledge statements are combined.

However, despite this mutual enrichment AI theory is still far from its ambitious goal that seemed to be so quick and easy to reach creation of genuine intelligent systems [1, 3, 4, 15] capable to learn, set goals and solve problems without external assistance, find new solutions that were not foreseen at the development stage.

Therefore it is natural to look on the brain, its texture and principles of functioning, because exactly brain makes it possible to solve intellectual tasks.

2 Brain Science and Limitations of Basic AI Approaches

Latest promising results in the field of genetic engineering together with wide accumulation of experimental data on brain dynamics obtained by means of modern techniques in noninvasive supervision over brain functioning [9, 10, 20] extended greatly the knowledge on cortex activities. High performance computing makes possible to model and in detail examine memory and behavioral processes that are accomplished by cells. A lot of research activities deal with estimation of chemical compound concentrations and electrical potential levels that indicate different states of neural systems [5, 9, 10]. It is worth mentioning that neural equations that describe neuron functioning were deduced in the late 1940's. It would seem that having knowledge on the structure and properties of basic construction element it is possible to extend this knowledge to create similar artificial intelligent system. However time passed but things are nearly right where they started. Only recently there were obtained results on part of rat brain modeling and on cultivation of biological neurons and networks on substrates (biological chips). It's time to integrate all practical AI knowledge and knowledge of relative scientific fields by usage of system analysis.

But wide knowledge about system micro-level functioning seldom leads to understanding a system as a whole entity, because it is necessary to take into account synergetic effects. This way helps to get and understand total system picture and interactions inside the system [4, 6, 15, 18].

So the dilemma sounds like this: at what extent we should fix our efforts on biological processes imitation and at the same time what level of abstraction from origin prototype is bearable?

There is a huge amount of isolated single-purpose AI models and methods [11, 19] that are effective in narrowly defined problems. But it is hard to consider them separately as basic ideas for general theory. Thus there are a lot of methods, great theoretical contributions and empirical materials deeply worked through but general picture happens to be rather messy.

It is obvious that idea to combine all existing artificial intelligence models into the unified super one wouldn't give a positive effect. And here arises more particular research question — what trend is more perspective from the point of development general AI theory.

One of the main issues that challenge Symbolic AI researchers consists in impossibility to formalize the representation of all variants of situations a system can meet during functioning [14, 15]. After making some assumptions a set of problems of a certain class (dependent on the limitations) are successfully solved but universality appears to be unattainable - the system is intelligent at the extent that is predetermined by input data comprehensiveness and predesigned scenarios. But main task is to get generation of new knowledge inside the system even if input data are not complete or contradictory.

Connectionist AI researches have faced the same restriction on representation of adequate size and quality of training samples. Neural networks are good at operating with implicit data, generalizing through learning process but approximation results on feed-forward networks depend drastically [4, 7] on the quality of training sample. In real world comprehensive data in most cases is unavailable.

Briefly speaking classical NN represent parallel implementation of corresponding pattern recognition methods by algebraic or probabilistic computations and thus inherit their advantages and most of disadvantages.

What can be used as integrator of different levels of knowledge about very complex system about brain? One of possible answers "synergy" science about systems self-organization [6, 18]. And use of synergy for development of AI theory is rather natural if we take into account that brain is dynamic system. Because of this there is growth of interest to bio-inspired and dynamic approaches today.

3 Novel Bio-inspired Methods and Dynamic Neural Networks

In spite of the fact that idea of artificial intelligence is already bio-inspired (like in the nature) the approximation to biological prototype can differ in the process of AI development. Classical NN models are comprised of formal neurons. Now more and more attention is given to bio-inspired NN with neurons similar to nerve cells. Recent investigations speak for the sputter out phase in NN models design as the solutions quality improvements are insufficient in comparison to huge efforts spent in this direction. Most likely detailed reproduction of the processes occurring in separate nerve cells without understanding the principals of mutual interactions will result with analogous effect.

Obtained results supplement our knowledge on multiform behavior of biologic systems but simple reproduction of neuron ensembles characteristics for partial regimes is hardly to be generalized. It is underlined that improvement of one part of parameters negatively influence on other previously well-tuned. Bottom-up de-sign starting from the lowest level most likely will not result with emergence of system with new qualities. Synergy effects occur mainly when self-organization principals underlie the system [6, 21].

There are two classes of NN: static (nearly all possible results here are reached already) and dynamic (interest to them is growing today). The evolution curve of dynamic NN can be separated by stages:

1. First stage – appearing of Hopfield's and then Haken's NN – attractors NN. Attractors in these dynamic NN are simple enough – fix-point attractors [6, 7].
2. Second stage – dynamic NN with more complex cycle attractors [22].
3. Current stage – dynamic NN with unstable dynamics – most complex chaotic attractors. Their dynamic is characterized by set of trajectories scaled in the phase space to a location with infinite number of switching states [2, 13].

Separate subclass of current stage dynamic NN are NN of reservoir computing [12, 13]. The origins of such systems can be found starting from second stage - they had random structure capable to generate promising complex dynamics.

When speaking about reservoir computing one of the main difficulties is the absence of guarantee that system will produce the required dynamics because there exists a great uncertainty about reservoir parameters (especially actual for Maass bio-inspired model). Empiric (or intuitive) parameter assignment is done on the basis of preliminary experiments series. There is strong proof that Turing machine can be realized on reservoir NN [13] but wide practical application of this approach is still the matter of future investigations.

To develop network with complex dynamics basic transfer functions in recurrent NN are replaced with chaotic maps [2, 16]. Application of functions that provide deterministic chaos corresponds with the current trend to combine linkage complexity concept with gradual complication of processing units. The most applicable is one-dimension logistic map that allows to control degree of generated chaos by means of one parameter.

The development of unified approach to AI requires to examine not isolated system but system in the environment where this system will be used. This idea is widely discussed in [17] where role of chaos intelligent agent development is considered. Most likely that such holistic approach will help to formalize at last such complex notion as context.

Just chaotic systems, whose dynamic strongly depends from external conditions, can help to represent context and as well all variety of possible situations (of really mobile and often preliminary undefined environment) in which tasks should be solved. Some promising results are obtained already [2, 6, 13, 17, 18, 21, 22].

4 Conclusions

Development of intellectual systems specified to solve certain class of problems should be obviously accomplished by means of approved methods or their hybrid modifications. Detailed reproduction of separate neuron cell ensembles in attempt to investigate some cortex domain is rational in neurophysiology models but not in AI most probably. Unified approach to development of artificial intelligent systems with quality commensurable to natural neural systems to our opinion should be based on distributed ensembles of chaotic (coupled) maps. This direction of research is attractive as it combines ideas of agent theory, NN theory, nonlinear dynamics, synchronization theory, formal logics. Consequently ample opportunities to apply complex synergetic effects to deal with uncertainty not only in technical but also in biological, economic, geopolitical systems foster an idea that in the nearest future nonlinear dynamics and chaos will become the most demanded apparatus to understand and model cognition processes.

References

1. Baum, S.D., Goertzel, B., and Goertzel, T.: How long until human-level AI? Results from an expert assessment. *Technological Forecasting and Social Change* 78, 185–195 (2011)
2. Benderskaya, E.N., and Zhukova S.V.: Fragmentary synchronization in chaotic neural network and data mining. *LNAI* 5572, 319–326 (2009)
3. Bobrow, D.G., and Brady, M. : Artificial Intelligence 40 years later. *Artificial Intelligence* 103, 1–4 (1998)
4. Cristianini, N.: Are we still there? *Neural Networks* 23, 466–470 (2009)

5. Goertzel, B., Ruiting, L., Itamar A., Hugo, G., and Shuo, C.: A world survey of artificial brain projects, PartII: Biologically inspired cognitive architectures. *Neurocomputing* 74, 30–49 (2010)
6. Haken, H.: Synergetic Computers and Cognition: A Top-Down Approach to Neural Nets. SSS: Springer (2010)
7. Haykin, S.: Neural Networks: A Comprehensive Foundation. Prentice Hall (1998)
8. Jang, J.R., and Sun, C. and Mizutani, E.: Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence. Prentice Hall (1997)
9. Kazanovich, Y.B.: Nonlinear dynamics modeling and information processing in the brain. *Optical Memory and Neural Networks* 16(3), 111–124 (2007)
10. Korn, H., and Faure, Ph.: Is there chaos in the brain? I. Concepts of nonlinear dynamics and methods of investigation. *Life Sciences* 324, 773–793 (2001)
11. Luger, G. F.: Artificial Intelligence: Structures and Strategies for Complex Problem Solving. Addison Wesley (2008)
12. Lukoševičius, M., and Jaeger, H.: Reservoir computing approaches to recurrent neural network training. *Computer Science Review* 3(3), 127–149 (2009)
13. Maass,W., Natschlaeger, T., and Markram, H. : Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Computation* 14(11), 2531–2560 (2002)
14. Mira, J.M.: Symbols versus connections: 50 years of artificial intelligence. *Neurocomputing* 71, 671–680 (2008)
15. Oliveira F.: Limitations of learning in automata-based systems. *European Journal of Operational Research* 203, 684–691 (2009)
16. Pikovsky, A., and Rosenblum, M., and Kurths, J.: Synchronization: A Universal Concept in Nonlinear Sciences. CNSS: Cambridge University Press (2003)
17. Potapov, A.V., and Ali, M.K.: Nonlinear dynamics and chaos in information processing neural networks. *Differential Equations and Dynamical Systems* 9(3-4), 259–319 (2009)
18. Prigogine, I.: The End of Certainty. Free Press (1997)
19. Russell, S. and Norvig, P.: Artificial Intelligence: A Modern Approach. Prentice Hall (2002)
20. Velazquez, J. : Brain, behaviour and mathematics: Are we using the right approaches? *Physica D* 212, 161–182 (2005)
21. Wolfram, S. A: A New Kind of Science. Wolfram Media (2002)
22. Zak, M.: Quantum-inspired resonance for associative memory. *Chaos, Solitons and Fractals* 41, 2306–2312 (2009)

Embodied Agent or Master of Puppets. Human in Relation with His Avatar

Mateusz Woźniak

Institute of Psychology, Jagiellonian University
Cracow, Poland
mgwozniak@gmail.com

Abstract. Neuronal and psychological processing of avatars is a relatively new but fast growing topic of interest. Most of the research focus on avatars meant as virtual agents controlled by humans, but the same explanations should also apply to other types, such as human-controlled robots. The most recent neuroscientific results prove that processing avatar information tends to increase activation in brain areas responsible for processing information about body enhancements and embellishments. The author presents results of his behavioral studies and argues that the effect of treating an avatar as a body enhancement can be mediated by cognitive content such as attitudes, convictions and beliefs, causing different prospective behavior.

Keywords: virtual reality, avatar, identity, embodied cognition, self concept

1 Introduction

People frequently engage in virtual interactions with artificial intelligence. For much of the time it seems to be a relatively direct communication. For example, when speaking with chatterbots like Alicebot [1] or Jabberwacky [7] people seem to meet them "in person", though this meeting is mediated by a computer communication service. Similarly, when interacting with the so-called Embodied Conversational Agents we can not only speak to but also interact face-to-face with virtual "people" [3] (for example with GRETA [6]). But it does not necessarily have to look like this. Sometimes we interact with other agents through an avatar. Like a professional puppeteer, we can use various objects to act on our behalf. Then we interact face-to-face, but with the face of our avatar instead of our real. Within this text, the emphasis will be put on computer avatars but the conclusions can be easily and effectively extrapolated onto other analogous situations.

2 Treating Artificial as Real

Recent research have provided a great amount of evidence showing that humans tend to intuitively and naturally engage in interactions with artificial agents as if they were real (*Media Equation Theory*: [9]). "Human interactions with computers, television and new media are basically as social and natural as interactions in normal life" (ibid. p. 15); moreover, this is the case even though people engaging in such interactions explicitly state that they do not consider their interaction partners to be living or intelligent beings. The consequence is that people tend to apply many human traits to machines or virtual agents, in a process frequently referred to as anthropomorphism [11], which subsequently results in a behavior different from that which would be expected from a strictly task-oriented, depersonalized approach. These insights are further strengthened by neuroimaging studies. Gazzola et al. [5] have conducted an fMRI study in which they were able to observe activation in the mirror neuron system in participants watching the movements of a robot when the dynamics of those movements were close to biological motion. Research indicates that the mirror system is probably responsible for, among others, such traits as empathy towards or imitation of observed object.

Among virtual agents, we can distinguish a special set of characters — avatars, which are virtual agents controlled by us. Interactions with other people or subjects within digital (or artificial) realms seem to correspond to our everyday experiences, but in the case of avatars controlled by us it is a little harder to find an analogy in the real world. It is not usual to "jump into" somebody else, into his body, identity and personality. Despite that, new technologies allow such bizarre situations; moreover, people experiencing them seem to feel surprisingly comfortable. After all, computer entertainment, especially computer role-playing games, make it possible. Scientific investigation concerning the topic of avatars has begun only recently, but some initial results are already available. For example Yee and Bailenson [12] have introduced the term "Proteus effect". They have conducted an experiment in which participants entered a highly immersive virtual environment using virtual helmets. They then had some time to acquaint themselves with how their digital bodies look like and afterward they were given an experimental task to perform. They had to stand at a comfortable distance from another avatar and tell him as much about themselves as they felt comfortable with. As a result, participants controlling attractive avatars were maintaining on average smaller distances and revealed more information about themselves than their counterpart controlling "unattractive" avatars. According to Daryl Bem's *Self-perception Theory* [2] they were modifying their behavior in order to fit their (even though artificial and temporary) look. They were behaving as if they were clothed in another person's body.

The latest inquiries into the relation with virtual bodies focus on the neural underpinnings of this process. In a recent article Ganesh et al. [4] have analyzed fMRI data regarding brain responses from two groups: intensive World of Warcraft gamers and nongamers. The first group (the gamers) was analyzed concerning their responses to WoW avatars, while the second group (the nongamers) was analyzed concerning their responses to known and liked cartoon characters. The result was that WoW players responded with greater activation in the left angular gyrus of the inferior parietal lobe when rating their avatars than when rating themselves and close others. Furthermore, the magnitude of this avatar-referential activity was positively related to their level of body plasticity measured by a questionnaire estimating the ease to incorporate and self-identify with body enhancements (for example prostheses). It suggests that, at least to some extent, avatars are perceived as extensions of the body.

3 Research

Continuing this line of research, the author has conducted an experiment regarding behavior of people controlling an avatar depending on the character of relation with it. The participants were told that they are going to play a computer game. A computer game modification was used in order to create a virtual variation on a classic obedience-to-authority experiment, originally conducted by Stanley Milgram in the 1960s [8]. The participants' task was to, using a virtual avatar, teach another digital person (the learner) a set of pairs of words and then to test his knowledge of those pairs. For each incorrect answer the learner gave he received a virtual electric shock starting at 30 volts and gradually increasing by 30 volts up to the maximum of 450 volts. The victim expressed signs of discomfort or pain with floating text, but, being virtual (of which the participants were fully aware), no real harm was done to him. Participants could also easily stop the experiment by contacting a virtual experimenter. The scene was seen from a third-person perspective from the isometric projection. The participants were divided into two experimental conditions. When meeting the learner for the first time, they introduced themselves. The first group introduced themselves with their real names, while the second one with a fixed number instead of a name: "Nice to meet you. I am 101115". The manipulation was intended to increase or decrease the level of identification with controlled agent. In this approach sharing the same name with a digital character was supposed to induce a cognitive interpretation that an avatar is a part of me and therefore I am responsible for his actions. On the other hand, using a number instead of name was meant to reduce this effect in favour of an interpretation that an avatar is something distinct from myself and therefore I am more psychologically distanced from the victim and less directly responsible for his suffering. The results have shown that:

First: Participants tended to sympathize with the virtual victim and anthropomorphize him as evidenced by the fact that among 64 participants only 22 inflicted the strongest electric shock and 42 refused to continue the experiment. Moreover, the qualitative data collected with a survey supports this claim. For example, a person who resigned at 210 volts has written that the reason of his resignation was: "The learner had health problems or at least he was saying so, the risk was too big". Other person explained: "Because I couldn't inflict such strong pain to another human, eventhough I knew he is not a real person". These results stand in line with those acquired by Mel Slater and his group [10] who created the first virtual replication of Milgram's experiment and with claims of media equation theory [9].

Second: People who introduced their avatar with their real names were more eager to quit or terminate the experiment earlier then the second group, measured by mean maximum inflicted electric shock. This means that possibly simply controlling a virtual agent does not necessarily define the character of a relation with it. These results suggest that different behavior can reflect differences in neuro-cognitive processing of the avatar. People who were more objectively identified with their avatar by sharing the same name seemed to be more "embodied" into it (or the opposite), while the second group was closer to the situation of instrumentally controlling an independent puppet.

Results described above show that people tend to automatically get engaged in virtual social situations. Moreover their behavior is mediated by a number of other factors. These first results suggest that the interpretation of relation with an avatar could be one of those. People seem to act differently when acting as "themselves" and as "controllers of somebody/something else". In current research a controversial and methodologically difficult method was used and therefore the results require more experimental data to be fully supported.

4 Conclusion

Recent research in the fields of psychology, neuroscience and media studies highlight the problem of relation between the user (gamer) and the avatar being under his or her control. It combines the latest discoveries concerning the body image and self concept with a more sociological approach. Overview of the results of the latest experiments suggests that user-controlled avatars are, at least in some cases, represented as extensions of physical body. On the other hand, the degree of such incorporation might be mediated by content of higher cognitive processes, especially interpretation of the relation with an avatar, as suggested by author's research. Further investigation, especially within the neuroscientific domain, should be conducted in order to support or discard this insight.

References

1. Alicebot: http://alicebot.blogspot.com/
2. Bem, D.: Self perception theory. In: Berkovitz, L. (ed.) Advances in experimental social psychology (Vol. 6). Academic Press, New York (1972)
3. Bickmore, T., Cassell, J.: Social dialogue with embodied conversational agents. In: van Kuppevelt, J., Dybkjaer, L., Bernsen, N. (eds.) Natural, Intelligent and Effective Interaction with Multimodal Dialogue Systems. Kluwer Academic, New York (2004)
4. Ganesh, S., van Schie, H., de Lange, F., Thompson, E., Wigboldus, D.: How the human brain goes virtual: Distinct cortical regions of the person-processing network are involved in self-identification with virtual agents. *Cereb Cortex* (2011)
5. Gazzola, V., Rizzolatti, G., Wicker, B., Keysers, C.: The anthropomorphic brain: the mirror neuron system responds to human and robotic actions. *NeuroImage* 35(4), 1674–1684 (2007)
6. GRETA: http://perso.telecom-paristech.fr/˜pelachau/Greta/
7. Jabberwacky: http://www.jabberwacky.com/
8. Milgram, S.: Obedience to Authority: an Experimental View. HarperCollins (1974)
9. Reeves, B., Nass, C.: The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. Cambridge University Press, Cambridge (1996)
10. Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., Pistrang, N., Sanchez-Vives, M.V.: A virtual reprise of the Stanley Milgram obedience experiments. PLoS ONE 1 (2006)
11. Waytz, A., Epley, N., Cacioppo, J.T.: Social cognition unbound: Insights into anthropomorphism and dehumanization. *Current Directions in Psychological Science* 19(1), 58–62 (2010)
12. Yee, N., Bailenson, J.: The Proteus effect: The effect of transformed self-representation on behavior. *Human Communication Research* 33(3), 271–290 (2007)

What Is Moral Agency of Artificial Agents?

Eva Prokešová

Institute of Philosophy, Slovak Academy of Sciences
Bratislava, Slovak Republic
evaprokesova@yahoo.co.uk

Abstract. The aim of the paper is to argue in favor of the view on moral agency of artificial agents. According to that, the status of actual moral agents cannot be given to artificial agents neither today, nor in the future. The main message of the paper is, that moral agency is something specific only for human agents and the status of actual moral agents cannot be given to any other beings, including artificial ones. Argumentation is supported by few briefly outlined discussions about important topics such as: motivation vs. action; original vs. derivative purposes; in/ability to feel shame, guilt, respect as culturally determined; agents' mutual recognition; understanding of motivation and behavior of another agent and the like.

Keywords: moral agency, human agent, artificial agent, moral motivation, original and derivative purposes

1 Introduction

There are three different views on moral agency of artificial agents in recent discussion. The first is that artificial agents are moral agents, whereas people are not. Supporters of this view argue that contrary to human agents, artificial agents follow pure rational and logic consideration and therefore act morally. Their actions are not affected by feelings and emotions, which influence behavior of human agents. According to the second view, artificial agents in recent state of knowledge and technology are not moral agents, but they can become ones thanks to expected technological progress, by which their abilities and competencies, such as cognitive abilities or autonomy, will be expanded. As stated by the third view, artificial agents are on principle not able to become moral agents neither today, nor in the future. There are various reasons, why various philosophers refuse to give moral status to artificial agents.

The aim of my paper is to argue in favor of the third view. I will claim, that moral agency is something specific only for human agents and the status of actual moral agents cannot be given to any other beings.

2 What Kind of Agents Are Artificial Agents?

Philosophical starting point of my view is ethical concept of strong evaluation created by Charles Taylor. Although Taylor's theory is concerned only with the nature and characteristics of human agency, much can be learned about the moral status of artificial agents exactly by comparing with moral statuses of ones considered to be fully developed moral agents.
Some authors dealing with the moral status of artificial agents [1] claim, that if some of their crucial abilities develop, it will be possible to give them full moral status. But which are these abilities and to what extent they have to be fulfilled to enable giving full moral status to artificial agents? Absence of feelings, emotions, beliefs and convictions as well as restricted autonomy, interaction, insufficient freedom - these all are reasons, why it is possible to object to morality of artificial agents. Although, it is necessary to realize, that thanks to technological progress machines will gain many of the abilities mentioned above, therefore I do not consider these objections relevant.
My paper is focused on what I consider the fundamental feature of human agency. By comparison I will try to prove, that differences between human and artificial agents are so primal and crucial, that it is not possible to rid of them or to overcome them by technological progress.

- Common feature of concepts mentioned above is, that they are concerned with the way that artificial agents act and what consequences their actions have. Moral status is given to them usually on the basis of the fact, that consequences of their actions have impact on patient and that it can be evaluated as harmful or useful.
  I will start by claiming, that in moral life of a person, motivation is more important than the action itself [5],[8],[9]. Human agent acts on the basis of his or hers own aims, preferences, values and long-term commitments, and lead his/hers life to something, that he or she consider worthy. For this reason I will try to point out the difference between the quality of motivation of the human and motivation of the machine. In case of human agents it is possible to speak about original purposes, whereas in case of artificial agents it is worthy to speak only about derivative purposes. In my opinion the fact, that machines cannot have original purposes, is the most crucial in giving moral status to them.
- Artificial agents do not have and under no circumstances can have moral emotions. I do not exclude, that in future machines will be equipped with sensors and will be able to recognize heat, pressure, touch etc. It is even possible, that they will be able to interpret these perceptions as hurtful or

pleasurable. But we gain moral emotions in absolutely different way; we do not have special sense for it. They are complexity of inner experience and social context. Emotions such as shame or guilt are specifically human and are always connected with certain culture [8],[9]. Inability to feel shame, guilt, respect, regard excludes the possibility of artificial agents becoming actual moral agents.

– In my opinion another crucial fact is that machines in moral action can behave only as agents, never as patients. They neither feel consequences of action, nor have the inner sense for what human agents consider worthy, suitable or useful. As a result, they do not have full moral experience but only very limited one. Therefore no machine can properly understand the significant dimension of its action.
– Mutual recognition is important for creating actual moral agent [5], [10], that artificial agents are not able to do so. The concept of recognition requires a certain state of understanding of meaningfulness of motivation and behavior of another agent.
– The result of inability of artificial agents to create original purposes and have moral emotions is impossibility to be responsible for its actions, in the moral as well as in the legal level. If artificial agent is not a source of its own aims, then it is considered only as a mean for achieving objectives of human agents. Practical problems arise: If artificial agent is only a mean and is not responsible for its own actions, then is the responsible one its designer, its producer or its user? Because of problematic determination of responsibility the credit or guilt for consequences of actions moves to artificial agent as a performer of the objective. In my opinion claiming, that machine is holder of moral responsibility is the same as claiming, that the same is a hammer.

If artificial agents cannot be considered to be moral agents, questions arise: "What kind of agents are artificial agents, if they are not moral agents?", "What is their position in moral space?", "How should we act towards artificial agents?", "To what extent is it desirable to expand competencies of artificial agents?", and many others, that I will try to answer in my paper.

3 Conclusion

As fast as technology is moving forward, new and specific moral questions arise. I believe we need to deal with these questions before it might cause any irreversible damages. The goal of the paper was to provide another point of view on moral agency based on the concept of strong evaluation. I was trying to do so by comparing what is considered to be full and actual human agency and agency of modern sophisticated machines. This differentiation might be helpful in moral recognition and law acknowledging in the future.

References

1. Dennet, D.C.: When HAL Kills, Who's to Blame? Computes' Ethics. In: Stork, D.G. (eds.) HAL's Legacy: 2001's Computer as Dream and Reality. MIT Press, Cambridge (1997)
2. Ferrara, A.: Reflective Authenticity. Rethinking the Project of Modernity. Routledge, London (1998)
3. Floridi, L.: On the Morality of Artificial Agents. In: Anderson, M., Anderson S.L. (eds.) Machine Ethics. Cambridge University Press, Cambridge (2011)
4. Grau, Ch.: There Is No "I" in "Robot": Robots and Utilitarianism. In: Anderson, M., Anderson S.L. (eds.) Machine Ethics. Cambridge University Press, Cambridge (2011)
5. Johnson, D.G.: Computer Systems: Moral Entities but Not Moral Agents. In: Anderson, M., Anderson S.L. (eds.) Machine Ethics. Cambridge University Press, Cambridge (2011)
6. Laitinen, A.: Strong Evaluation without Sources: On Charles Taylor's Philosophical Anthropology and Cultural Moral Realism. University of Jyväskylä, Jyväskyla (2003)
7. McDermott, D.: What Matters to a Machine? In: Anderson, M., Anderson S.L. (eds.) Machine Ethics. Cambridge University Press, Cambridge (2011)
8. Sullins, J.P.: When Is a Robot a Moral Agent? In: Anderson, M., Anderson S.L. (eds.) Machine Ethics. Cambridge University Press, Cambridge (2011)
9. Taylor, Ch.: Human Agency and Language: Philosophical Papers, vol.1. Cambridge University Press, Cambridge (1985)
10. Taylor, Ch.: Sources of the Self. Cambridge University Press, Cambridge (1989)
11. Taylor, Ch.: The Ethics of Authenticity. Harvard University Press, Cambridge (1992)
12. Taylor, Ch.: Philosophical Arguments. Harvard University Press, Cambridge (1995)
13. Wallach, W., Allen, C.: Moral Machines: Teaching Robots Right from Wrong. Oxford University Press, New York (2009)

New Emergence as Supervenience Relieved of Problems

Eliška Květová

Faculty of Philosophy and Arts, University of West Bohemia
Sedláčkova 38, 306 14 Plzeň, Czech Republic
eliska.kvetova@gmail.com

Abstract. Supervenience and emergence are remarkable notions of cognitive science, which notably influenced especially philosophy of mind in the twentieth century. Issue of supervenient or emergent relationships is complicated, but it is possible to observe tendency to prefer emergence at the expense of notion supervenience from nineties. This paper aims to answer question why it is so. These two notions have always been very close to each other. This paper introduces development and common history of the concepts as well as changes of their relationship. The goal of the contribution is not only to consider the history of supervenience and emergence, but also to find appropriate distance to introduce thesis explaining the current relationship between them. The thesis could be simply formulated as follows: New use of the concept of emergence can be understood as a continuation of the idea of supervenience deprived of its fundamental problems.

Keywords: supervenience, emergence, philosophy of mind, cognitive science

1 Introduction

This paper aims to discuss relationship between two important notions of philosophy of mind: supervenience and emergence. The initiative question is how it is possible that these originally closely related concepts get into the opposition according to many authors. I must add that the history of these concepts is very long and rich, particularly when I take into account intuitive ideas of relationships which were later denoted as supervenient or emergent relationships. For this reflection I will focus on the context in which these terms appear in cognitive science of the twentieth century. Due to the tendency which this contribution focuses on notion of supervenience will be given first.

The simple and intuitive idea of supervenience was very attractive. The special type of relationship between two levels of properties is based on very intuitive idea that there is one kind of property or fact (supervenient property) that may only be present in virtue of the presence of some other kind of properties (subvenient base). New property appears and changes depending on the particular arrangement of underlying properties. It still maintains a degree of autonomy for supervenient properties. Supervenience seemed to be very promising concept for many areas, especially for mind-body problem. It was an attractive alternative to reductionst physicalism. It is possible to speak about boom of supervenience in cognitive science in eighties of the twentieth century. This boom was connected with famous names as Jaegwon Kim, Terence Horgan, John Haugeland, David Kelogg Lewis, Simon Blackburn, Brian P. McLaughlin, Barry Leower etc. But the initial optimism faded. Supervenience was largely stigmatised by criticism and many other difficulties. Supervenience proved to be unable to answer the fundamental question of the nature of this concept, unable to resist the strong criticism. Supervenience with its problems is weakened or refused by many authors. On the other hand emergence appears in philosophical papers more and more frequently. From nineties there is a tendency to strictly distinguish between supervenience and emergence. It is interesting, because these two notions were used as synonyms in the twenties of twentieth century.

Ambition of this paper is not to full specify the nature of the relationship of the concepts. Such a goal would require a much deeper insight and detailed examination of the issue. As it turns out, the nature of the relationship largely depends on the level of description or better on the distance that we choose. The paper would like to point out some interesting aspects of development of this relationship and would like to introduce a possible perspective on the role of the modern notion of emergence.

In this paper I will introduce the development of the relationship between supervenience and emergence, than I will discuss the issue of their differences and common features. I will try to show the information about development of their relationship as a guide for interpreting emergence as a new way to revive what once seemed attractive in supervenience.

2 Supervenience vs. Emergence

It is undisputable that there is very close relationship between supervenience and emergence. Denotations of these words, terms, concepts has been changing for a very long time as well as their mutual position has been changing. Let me briefly outline some important points of development of these concepts.

2.1 Development of the Relationship

McLaughlin [4] describes the use of terms emergence and supervenience by describing history of British emergentism. Father of this tradition was John Stuart Mill. His “heteropatic effect” [5] is labeled as ”emergent” by George Henry Lewes. “Emergent” is an effect that is not only the sum of all effects of each of its causes if they acted separately. Among the number of representatives of British emergentism I will present Lloyd Morgan in the abstract. Morgan in [6] used words “supervene” and “emerge” as different stylistic variants of the same. The next step in the development of the relationship between supervenience and emergence is Van Cleve’s notion of an “emergent property” [7] which refers to supervenience.

Modern concept of emergence occurs in many contexts, we can meet this term e.g. in connectionist functionalism in artificial intelligence. In the nineties these terms can no longer be claimed to be synonyms. Or as it turns out, they could be, but they are not. The nature of their relationship is increasingly confused and a number of questions appears. Is there any relationship of subordination and superiority between them? Is it not too simplistic to determine supervenience as term that could be employed to explicate the notion of emergence as McLaughlin in [4] does it? Equally simplistic idea is that emergence can be considered as a general term that refers to the appearance of “new properties” and supervenience gives stronger framework to emergence, because it indicates directly to the relationship between these “new” or supervenient properties and properties that could be called basic or subvenient.

2.2 Opposition of the Concepts?

It was stated that from nineties emergence appears more frequently in papers, while supervenience unable to successfully respond criticism loses its attractiveness. Attempt to define the simple idea of supervenience fails and this concept is gradually fragmenting into many types and categories (as weak, strong, global, local etc.) of supervenience, which could provide a more accurate definition. Because of this fragmentation supervenience loses its simplicity, clarity and appeal. Supervenience has been faced with many problems, the notion seemed to be somehow undetermined in many aspects. Supervenience fails to answer the basic question: What is it? What does it consist in? Why to believe that some properties supervene on another (base) properties? How can be explained that mental properties supervene on physical properties? Therefore the concept of supervenience had to recede into the background. Even supporters and proponents of this concept could not adequately answer questions about supervenience and provide a solid foundations for this notion. On the other hand idea of this kind of relationship remained to be an interesting possibility and inspiration.

There are tendencies to prefer emergence, e.g. in [1]. There is no need to be so radical (emergence, not supervenience), but it is necessary to highlight the fact, that there is obvious tendency to advert to their differences, to distinguish between supervenience and emergence, e.g. in[1],[8]. In this context next question appears: In what do they differ in fact? Or how much do they differ? The answer leads us to the thesis of the paper.

2.3 Modern Emergence as Continuation of Supervenience

I do not want to refute the views promoting differentiation and prioritization of emergence — they absolutely make sense when we consider problems and fragmentation of the notion of supervenience. My position is based on the opinion, that it also makes good sense not to shatter the basic idea, to stay at a general level, to consider only the basic idea of these concepts, the core idea that is not marked by stated problems and criticism. Both notions have problems with clear definition, but they share the same core idea. The thesis of modern emergence shows a striking resemblance with core idea of supervenience. There are many common features that are connected with not entirely successful supervenience as well as with modern emergence, features that support my statement, that there is no big difference between core idea of supervenience and core idea of emergence: property or fact that appears as something additional, unexpected; dependency relationship between different levels of properties maintaining certain autonomy of supervenient or emergent property, multiple realizability which was very inviting for functionalism and artificial intelligence etc.

If we consider an intuitive idea that connects both concepts and all of the above the main thesis of the paper can be formulated. New use of the concept of emergence can be understood as a continuation of the idea of supervenience deprived of its fundamental problems. Emergence is an attempt to keep the good and attractive things about supervenience and distance from problems that this notion was not able to cope with. (We put aside the fact how successful this second attempt was.)

3 Conclusion

View of the development of concepts of supervenience and emergence can provide interesting conclusions. Supervenience attracted much attention, but it failed to succeed. Time of emergence came. Two close concepts had to be distributed in order that the common idea can develop. Exploration of the history of these concepts can illuminate some aspects of their development, can answer asked questions, may allow a new interpretation, which is rather than big discovery a proof that shows how strongly the relationship which is in our discursus known as supervenient of emergent is deep-seated in our thinking or mind. In adequate distance, where it is possible to concern core ideas of these concepts, new emergence is just another stage of development of the concept of a hierarchical relationship between properties, a relationship that preserves some autonomy to those properties that are new, additional and dependent on the basic, subvenient properties which are arranged in a certain way. In this light it is possible to say that new use of the concept of emergence can be understood as a continuation of the idea of supervenience deprived of its fundamental problems.

References

1. Butterfield, J.: Emergence, reduction and supervenience: a varied landscape (September 2010), http://philsci-archive.pitt.edu/5549/, forthcoming in Foundations of Physics
2. Humphreys, P.: Emergence, not supervenience. *Philosophy of Science* 64, 337–345 (1997)
3. Kim, J.: Supervenience and mind: selected philosophical essays. Cambridge University Press, Cambridge (1993)
4. McLaughlin, B.P.: Emergence and supervenience. *Intellectica: Emergence and explanation* 2(25), 25–43 (1997)
5. Mill, J.S.: Of the composition of causes. In: System of Logic (1859)
6. Morgan, C.L.: Emergent Evolution. Henry Holt and Co. (1923)
7. Van Cleve, J.: Mind–dust or magic? Panpsychism versus emergence. *Philosophical Perspectives* 4, 215–226 (1990)
8. Vision, G.: Re-Emergence: Locating Conscious Properties in a Material World. The MIT Press, Cambridge, Mas. (2011)

Neuroinformatics - Data Management and Analytic Tools for EEG/ERP Research

Jan Štěbeták, Petr Brůha, and Roman Mouček

Department of Computer Science and Engineering, University of West Bohemia,
Univerzitní 8, 306 14 Pilsen, Czech Republic
{stebjan, pbruha, moucek}@kiv.zcu.cz

Abstract. Brain research is currently a rapidly growing scientific field. Large amounts of data, various data formats and non-standardized domain descriptions lead to incompatible results and interpretations of neuroscientific data and to difficult communication between laboratories and scientific teams. Neuroinformatics, a newly established scientific field, provides integration across all levels and scales of neuroscience. Our research group has contributed to the building of a neuroinformatics infrastructure by developing and integrating data management and analytic tools for EEG/ERP (electroencephalography, event-related potential) research.

Keywords: neuroinformatics, electroencephalography, event-related potentials, INCF, data management, semantic web, analytic tool

1 Introduction

Understanding the brain is one of the greatest challenges of modern science. Due to the rapid development and affordability of powerful computers, neuroscience attracts more and more researchers and produces a large amount of remarkably diverse data. This natural diversity is caused by the large variety of scientific procedures, scales, and recording techniques used. An obvious, yet unpleasant side effect of extensive research in neuroscience are difficulties with data management, data granularity, terminology, hierarchy of domain concepts, and performance of analytic tools. A successful integration of diverse data and related findings obtained by different laboratories and multiple experimental techniques is essential for understanding the functioning of the brain and, as a result, treating diseases.

A newly established scientific field, neuroinformatics, provides this integration across all levels and scales of neuroscience. It encompasses the tools and techniques for data acquisition, sharing, publishing, storage, analysis, visualization, modeling and simulation [1]. The International Neuroinformatics Coordinating Facility (INCF) was established to develop this international neuroinformatics infrastructure, which promotes the sharing of data and computing resources within the international research community. The primary objective of the INCF is, however, to help develop scalable, portable, and extensible applications that can be used by neuroscience laboratories worldwide [2].

Our research group at the Department of Computer Science and Engineering, University of West Bohemia, a member of the Czech INCF National Node, specializes in the research into electroencephalography (EEG) and event-related potentials (ERP). These techniques were expected to become obsolete when hemodynamic methods (e.g. PET and (f)MRI) were developed, but due to their high temporal resolution they are currently viewed as important complements to them. Apart from performing EEG/ERP experiments, we contribute to the development of the international neuroinformatics infrastructure by creating software and hardware solutions for acquisition, management, sharing, analysis and visualization of EEG/ERP data.

This paper focuses on two aspects of building a software infrastructure for EEG/ERP research: data management with emphasis on data modeling and semantic web issues, and development and integration of analytic tools for processing the EEG/ERP signal. We note that some terms, such as semantics, ontology or concept, are used as they are understood in software engineering and semantic web research.

2 State of the Art

This section shortly describes the state of the art in the development of neuroinformatics infrastructure, the semantic web solutions for EEG/ERP data management and methods for EEG/ERP signal analysis.

The neuroinformatics infrastructure is being built in several INCF national nodes in parallel. The INCF portal [1] includes e.g. a software center for easy storage and sharing of neuroinformatics software tools, a content management system for national nodes presentation and provides access to supercomputing resources for the neuroinformatics community.

The Neuroscience Information Framework (NIF) is a dynamic inventory of registered Web-based neuroscience resources containing data, materials, and tools. It advances neuroscience research by enabling access to public research data and tools through an open source environment [3]. Currently more than 4,000 sources are registered in NIF.

The British Carmen portal [4] provides storage of experimental data, metadata and analysis code. Users can also analyze their experimental data. Within the Japan National Node, neuroinformatics databases are organized and shared with the public in platforms. The platforms serve as storages of experimental data and papers, some of them provide analytic tools.

Semantic web solutions for EEG/ERP data management focus mainly on building ontologies. The ontology built within the NEMO project provides formal semantic definitions of concepts in ERP research, including ERP patterns, spatial, temporal, functional (cognitive/behavioral) attributes of these patterns, and data acquisition and analysis methods and parameters [5].

The methods for analyzing a pure EEG signal are used especially in medicine for detecting epileptic waveforms and sleep disorders, or for indicating brain death. The cluster analysis and adaptable neuronal networks are widely used [6]. The fundamental method for obtaining ERP components from the EEG signal is averaging. The methods used for ERP components detection include the matching pursuit algorithm, wavelet transform, and Hilbert-Huang transform [7].

3 EEG/ERP Portal

Since there is no ontology describing conditions and circumstances during encephalography experiments (e.g. weather conditions, used hardware and software or subject's impairment) and no reasonable and easily extensible tool for long-term storage and management of EEG/ERP data (metadata), our research group has contributed to the building of neuroinformatics infrastructure by developing the EEG/ERP portal. The portal offers the set of features (Fig. 1). The set of accessible features depends on the specific user role.

4 EEG/ERP Data Management - Semantic Web Solution

The EEG/ERP Portal has been developed using common design and programming approaches (three-layer architecture, relational database, object relational mapping) and technologies (Oracle, Hibernate, Spring, Spring Security). In the description of the domain we used experience from our experiments, medical expertise provided by the University Hospital in Pilsen, our results of cooperation with other INCF national nodes and, last but not least, our analysis of related scientific papers published in international journals.

Fig. 1. EEG/ERP Portal - User Interface Preview

The domain ontology was at first built as a relational data model mapped to an object-oriented code by the Hibernate framework. An equivalent UML model was derived from the code. In general, expressional means of common data models are considered to be too restricted to serve as suitable domain descriptions. Moreover, recognition of the portal as a well-known data source needs providing the domain description (ontology) in the semantic web form (e.g. registration within NIF). To get this description, transformations from the relational model and the object oriented code were investigated in parallel. However, these transformations are very difficult to implement since an underlying semantic gap between both descriptions has to be bridged. As a result, we used the D2RQ tool [8], the Jena framework [9] and the OWL API [10] as the basic building blocks and designed and implemented both transformations. The transformation from the code to the semantic web description was enriched by new transformation mechanisms based on Java annotations and extension of used frameworks. The resulting library was integrated into the portal to provide the domain description.

5 Analytic Tools

We provide methods to detect ERP components and the actual time of its occurrence in the EEG/ERP portal. In the case of the wavelet transformation the user chooses a wavelet which is translated and dilated until it covers the original signal. The matching pursuit method decomposes the signal into a set of Gabor's atoms. Both methods are integrated into the EEG/ERP portal as JAR libraries. The computational complexity is linearly dependent on the size of the analyzed file and on the chosen wavelet or on the number of atoms. Currently, we are working on the way of displaying the results of these methods.

6 Conclusion

This abstract introduces neuroinformatics as a scientific field integrating information across all levels and scales of neuroscience and presents the contribution of our research group to building the infrastructure for EEG/ERP research.

Acknowledgments

The work was supported by the Ministry of Education, Youth and Sport of the Czech Republic under the grant ME 949 and by the UWB grant SGS-2010-038 Methods and Applications of Bio- and Medical Informatics.

References

1. International Neuroinformatics Coordinating Facility (2011), `http://www.incf.org/about/what-is-neuroinformatics`
2. International Neuroinformatics Coordinating Facility (2011), `http://www.incf.org/about/what-we-do/our-mission`
3. Neuroscience Information Framework (2011), `http://www.neuinfo.org/`
4. The British Carmen portal (2010), `https://portal.carmen.org.uk/`
5. Neural Electro Magnetic Ontologies (2011), `http://nemo.nic.uoregon.edu/wiki/NEMO`
6. Krajča, V.: Automatic classification and visualization of epileptic EEG by supervised and unsupervised algorithms, *ČVUT FMBI, FNB, lecture notes* (2011)
7. Ciniburk, J., Mouček, R., Mautner, P., Řondík, T.: ERP components detection using wavelet transform and matching pursuit algorithm. *DCII*, Prague (2010)

8. D2RQ (2011), http://www4.wiwiss.fu-berlin.de/bizer/D2RQ/
9. Jena framework (2011), http://jena.sourceforge.net
10. OWL API (2011), http://owlapi.sourceforge.net/

Hand-drawn Objects with Structure as Means of Communication with Machines

Daniel Průša and Václav Hlaváč

Center for Machine Perception, Department of Cybernetics,
Faculty of Electrical Engineering, Czech Technical University
121 35 Prague 2, Karlovo náměstí 13
{prusapa1, hlavac}@cmp.felk.cvut.cz

Abstract. We study hand-drawn objects with structure as a way of human-machine communication. The objects include various schemas like building plans, electronic circuits, filled forms or notations like mathematical or chemical formulae. They usually carry a quite complex information in a form which is intelligible to the addressed side, thus we consider them as important means of information representation and transmission. We identify the main principles relating to the structure recognition as it is performed by humans and we outline how the principles can be simulated by machines.

Keywords: human-machine communication, pattern recognition, 2D grammars, structure, syntax and semantics

1 Introduction

Humans use various ways of communication in personal contact. Speech is the most frequent one. It is often complemented with facial expressions and gestures. Gestures can carry supplementary information as well as play a major role, for example for deaf persons who use sign language.

People would prefer to have the same possibilities when communicating with machines. The above mentioned ways are natural, comfortable and effective. And really, we can find many cases when such methods are adopted. For example, voice is commonly used to control smart phones.

When the complexity of information to be exchanged is growing, people tend to supplement their explanation by drawings. Consider a man and a woman who are going to furnish their apartment. They will use a sheet of paper to outline a kitchen design to reach an agreement on its final form.

Another example is a teacher explaining to students what is a derivative of function. This can hardly be done without a blackboard.

We can say that drawing a scheme or mathematical formulae also represents a way of communication. Moreover, we can find many examples when this approach could be beneficially applied to speak to machines. Let us list some existing or potential scenarios.

- *Tablet PC.* A user working with a word processor needs to enter a mathematical formula. He would prefer to write it in a natural way by hand, e.g. using an electronic pen, instead of composing the formula within the application by selecting and combining particular parts piece by piece.
- *Machine processing number lottery tickets.* A bettor fills in a ticket. He marks bet numbers by crosses. The ticket is processed by a machine. Betting numbers are detected, a certificate is printed and passed to the bettor.
- *Autonomous humanoid robot.* A senior citizen owns a companion robot. He would appreciate the opportunity to navigate the robot by drawing a plan.

2 Structure

What is structure? Let us cite Wikipedia to provide a general definition.

> *"Structure is a fundamental, tangible or intangible notion referring to the recognition, observation, nature, and permanence of patterns and relationships of entities."*

All the objects we mentioned so far can be characterized as planar patterns composed of hand-drawn entities. Under the understanding of the structure we mean the identification of entities and disclosure of relationships between them.

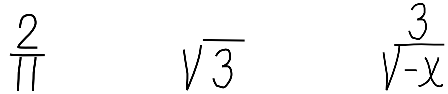
2.1 How Humans Recognize Structure

To realize how people perceive structure, it is necessary to pay attention to two phenomena known from linguistics – *syntax* and *semantics*. Semantics is the study of meaning and focuses on the relation between basic units such as words, phrases, signs and symbols. Syntax describes the combinatorics of the units, without reference to their meaning.

Sentence "Boys runs quickly." contains a syntax error. The subject or predicate has to be singular. On the other hand, sentence "Boys run musically."

is well formed, but the adjective does not make sense, its usage is a semantic error.

We can find syntax and semantics for each domain of the studied structural objects. Their knowledge is crucial in identifying entities as well as determining their role within the structure. An evidence for sentences is given in [8]. Readers have significant error rates when recognizing characters out of context. For manuscript, resp. cursive writing, Suen in [8] reports error rate of 2.4%, resp. 4.4%. Another example, related to a planar structure, is depicted in Fig. 1.



$\frac{2}{11}, \frac{2}{\Pi}$ $\qquad$ $\sqrt{3},\ V\overline{3}$ $\qquad$ $\frac{3}{V-x}, \frac{3}{\sqrt{-x}}$

Fig. 1. Syntax enables to resolution of ambiguities occurring in hand-written mathematical formulae. The expected, resp. incorrect interpretation is shown bellow each formula.

2.2 How Machines Can Recognize Structure

Hand-drawn objects recognition comes under the scope of *pattern recognition*. Two different data types can be considered as inputs:

- *Raster images*, taken e.g. by a scanner (we speak about *off-line* data).
- *Sequences of strokes*, hand-written on a tablet (*on-line* data).

The following tasks have to be completed by a machine to perform structural recognition.

1. *Segmentation.* It is necessary to partition image into areas, resp. strokes into groups that form one entity.
2. *Entities recognition.* Segmented areas, resp. strokes have to be classified. Methods of Optical Character Recognition (OCR) are applied.

3. *Structural analysis.* Knowledge of syntax and semantics has to be utilized to determine dependencies between entities.

Syntax can be represented by mathematical formalisms. One of the most famous are *context-free grammars*, introduced within the theory of formal languages [1]. Their initial ambition was to describe natural languages [2], but they found a better use for programming languages. Extensions of the grammars were proposed to model planar structures [3, 4].

It is harder to mathematically express semantics. Selected semantic restrictions could be represented by hard-coded rules. On the other hand, even the structural analysis based only on the syntax can give good recognition results.

2.3 Comparison, Humans vs. Machines

Humans are more effective in structure recognition than machines. Their perception is robust, they benefit from a deep knowledge of semantics which, together with logic, helps them to deal with different dialects or corner cases.

On the other hand, there are situations that machines can handle better than humans. As the structure size grows in the number of units, it becomes confusing for people. They can easily fail when trying to understand it. Even relatively small structural objects could be hard to decipher when the same units repeat several times. An example is given in Fig. 2. A human reader will tend to write additional markers into such expressions to make sure that the syntax is correct.

$$((((7 * (2 + 1))) + (2 + 3) * 3)) + ((23 - 4) * 5))))$$

Fig. 2. Is bracketing in the expression correct?

Large size objects do not pose a problem to machines. Conversely, they provide more syntactic dependencies in the object, leading to a better chance to recognize the whole structure correctly.

3 Our Approach to Structural Recognition

We develop so called *structural construction paradigm* [7] – a framework for structural recognition that puts emphasis on the segmentation driven by structural analysis. Segmenting entities standalone is a prone to errors from which it is difficult to recover. To model the syntax we proposed a stochastic variant of two-dimensional grammars called *coordinate grammars*. They allow to express spatial relationships between entities.

To prove that the general concept leads to an effective and reliable implementation, we applied it to the domain of mathematical formulae [5, 6]. We focus primarily on the on-line inputs, however, the proposed structural analysis could be applied to raster images as well.

An important precondition for tuning the method is to have a sufficiently large database of annotated formulae. Since there are no publicly available samples of on-line formulae, we built own database. It servers several purposes. Statistical methods were applied to extract parameters for the grammar rules. It is also a benchmark for correct rate of the implementation. Our future plans include to study possibilities of semantic dependencies extraction.

4 Conclusion

We gave arguments proving that hand-drawn objects with structure play an important role in human-machine communication. We also described principles that machines can follow to understand the meaning of structure. Although several recognition methods were successfully implemented for particular objects, there are still new challenges. A large variety of domains and entity dependencies makes the task difficult. Universal, robust and effective approaches are subject to further research.

Acknowledgment

The authors were supported by the Grant Agency of the Czech Republic under project P103/10/0783.

References

1. Hopcroft, J. Ullman, J.: Formal languages and their relation to automata. Addison-Wesley (1969)

2. Chomsky, N.: Syntactic structures. Mouton and Co, The Hague (1957)
3. Liang, P., Narasimhan, M., Shilman, M., Viola, P: Efficient geometric algorithms for parsing in two dimensions. *Proceedings of the 8th International Conference on Document Analysis and Recognition*, 1172–1177 (2005)
4. Průša, D.: Two-dimensional languages. Faculty of Mathematics and Physics, Charles University in Prague (2004)
5. Průša, D., Hlaváč, V.: Mathematical formulae recognition using 2D grammars. *Proceedings of the 9th International Conference on Document Analysis and Recognition*, vol. II, 849–853 (2007)
6. Průša, D., Hlaváč, V.: Structural construction for on-line mathematical formulae recognition. *Proceedings of the Iberoamerican Conference on Pattern Recognition*, Springer Verlag, 317–324 (2008)
7. Schlesinger, M.I., Hlaváč, V.: Ten lectures on statistical and structural pattern recognition. *Computational Imaging and Vision* 24, Kluwer Academic Publishers (2002)
8. Suen, C.Y.: Handwriting generation, perception and recognition. *Acta Psychologica* 54, 295–312 (1983)

Beyond AI: Towards Smart Workplaces

Peter Mikulecký

Faculty of Informatics and Management, University of Hradec Králové
50003 Hradec Králové, Czech Republic
peter.mikulecky@uhk.cz

Abstract. Area of Ambient Intelligence (AmI) that can be naturally considered as an area beyond Artificial Intelligence exploiting its achievements largely aims to make various digital devices so embedded and hidden in environment that humans can use them even without thinking about them. One of recently studied application area of AmI seems to be the area of intelligent offices, or more generally, intelligent workplaces. The paper is focused on presenting an overview of recent activities in the area of smart offices and other intelligent workplaces. An overview of recently published important results in the area is presented and some open problems are reminded.

Keywords: ambient intelligence, smart environments, smart workplaces

1 Introduction

In the last decade a new self-contained and highly applicable area appeared, based mainly on the *Artificial Intelligence* achievements. The area of *Ambient Intelligence (AmI)*, that can be considered naturally to be an area beyond Artificial Intelligence, aims to make digital devices so embedded and natural that humans can use them without even thinking about them. There is a number of recent *AmI* applications, among them well known intelligent homes, health care in hospitals or houses for elderly, intelligent classrooms, or offices. Many of these intensively investigated and developed applications can be considered also as a kind of intelligent workplaces aiming to support workers in fulfilling specific but also routine tasks.

The paper will be focused on presenting an overview of recent activities in the area of smart offices and other intelligent workplaces. We shall bring an overview of recently published important results in the area and we wish to present also some open problems.

2 Recent Situation

The basic idea behind *AmI* is that by enriching an environment with technology (e.g., sensors and devices interconnected through a network), a system can be built such that acts as an "*electronic butler*", which senses features of the users and their environment, then reasons about the accumulated data, and finally selects actions to take that will benefit the users in the environment.

Cook, Augusto, and Jakkula note in [3], that *AmI* incorporates artificial intelligence research into its purview, encompassing contributions from machine learning, agent-based software, and robotics. *AmI* research can therefore include work on hearing, vision, language, and knowledge, which are all related to human intelligence.

According to [10], one of most interesting and useful applications of *AmI* seems to be *Smart Offices* and *Smart Decision Rooms*. Smart offices and decision rooms contribute to reducing the decision cycle, and offer connectivity wherever the users are, aggregating the knowledge and information sources. The topics as smart offices and intelligent meeting rooms are well studied and they intend to support the decision making activity, however, they have received a new perspective from the *AmI* concept. This concept enables a different way to look at traditional offices and decision rooms, where it is expected that these environments support their inhabitants on a smart way, promoting an easy management, efficient actions and, most importantly, to support the creation and selection of the most advantageous decisions.

Le Gal [5] defines a smart office as an environment that is able to help its inhabitants to perform everyday tasks by automating some of them and making the communication between user and machine simpler and effective. Smart offices can be defined further on as an environment that is able to adapt itself to the user needs, release the users from routine tasks they should perform, to change the environment to suit to their preferences and to access services available at each moment.

Freitas and others pointed out in [4], that the aim today is to develop systems that support distributed and asynchronous meetings, naturally allowing a ubiquitous use that can add flexibility to the global organisational environment of today. This kind of rooms named by Intelligent Meeting Rooms (IMR) can be considered as a sub domain of Smart Rooms for the workplace context. Their aim is mainly to support multi-person interactions in the environment in real time.

In accordance with the methodology proposed in [9] by Ramos, Augusto and Shapiro, it is expected that *Smart Offices* and *Intelligent Decision Rooms* integrate *AmI* environments covering the following tasks:

- interpreting the environment state;
- representing the information and knowledge associated with the environment;
- modelling, simulating and representing entities in the environment;
- planning decisions or actions;
- learning about the environment and associated aspects;
- interacting with humans;
- acting on the environment.

When the necessity of user preferences appears, usually new location-based services can be adapted to accomplish this task. Related to this, it is proposed in [1] that a smart environment takes care for storing and sending the personal information. The person in this approach is always accompanied by a mobile virtual object in the smart environment. So location based services adapted to personal profiles can be offered.

Specific place among challenging applications of *AmI* have application in educational process that can be beneficial at every level of education. Educational environment certainly is a workplace deserving a special attention and focus. One of these applications is the *Smart Classroom* project [11]. It aims to build a real-time interactive classroom with tele-education experience by bringing pervasive computing technologies into traditional distance learning. The goal of *Smart Classroom* project is to narrow the gap between the teacher's experience in tele-education and that in the traditional classroom education, by means of integrating these two currently separated education environments together. A more general overview of the *AmI* possibilities in education brings also our recently published book chapter [7].

3 Where We Go Now?

Bureš and Čech [2] argue that the achievement of ambient intelligence postulates an adequate shift in thinking that concerns also managerial work. The paper therefore presents experience on how to test the meaningfulness of teaching systems thinking for managers and increasing thus the level of acceptance of new technologies.

One of our earlier papers [8] pointed out that managers, in order to be capable of producing the best possible strategic decisions, should have the right information in the right time. However, without having the appropriate knowledge the production of good decisions would not be easy, if not impossible. It is, therefore, quite sensible to think about such a managerial workplace, where the manager would have the best possible working conditions in various meanings of this formulation. The paper [8] was therefore devoted to an analysis of managerial needs for a qualitatively higher working environment, based on the idea of ambient intelligence, oriented on intelligent environment design.

Based on results presented in our paper [6] we wish to stress the role of knowledge and of its management in intelligent workplaces of various kinds. Such environments inevitably need to be rich in knowledge; therefore a synergy of approaches and techniques from ambient intelligence as well as from knowledge management is necessary. An analysis is necessary of what are the basic common features of intelligent (and therefore knowledge rich) workplaces, what are they relations to ambient intelligence approaches, and what other serious problems could arise.

A specification of basic features of such an intelligent workplace could be helpful in overcoming some barriers and stressful situations typical for managerial decision making. In our opinion, the workplace should, among its other features:

– to ensure broad but focused (and personalized) access to relevant information and knowledge resources, supporting thus both learning needs of the manager as well as creation of his/her decisions;
– to offer as much relief from stress as possible by avoiding all the usual stressful situations (or more precisely their potential sources);
– to ensure broad and up to date technical support for all technically based activities in the workplace.

The problems and works mentioned above are just a collection of some applications and approaches which are somehow interesting and important for our further research direction. The collection certainly is not complete and a lot of other important problems could be found. However, we hope that some inspiration for focusing research work can be found here. We think that at least the importance of the area for common practice has been exhibited.

Acknowledgments

The research has been partially supported by the Czech Scientific Foundation project No. P403/10/1310 SMEW.

References

1. Bagci, F., et al.: The refective mobile agent paradigm implemented in a smart office environment. *Pers Ubiquit Comput.* (11), 11–19 (2007)
2. Bureš, V., Čech, P.: Complexity of ambient intelligence in managerial work. In: *ITiCSE 2007: 12th Ann. Conference on Innovation and Technology in Computer Science Education.* ACM Press, New York (2007)
3. Cook, D., Augusto, J., Jakkula, V.: Ambient intelligence: Technologies, applications, and opportunities. *Pervasive and Mobile Computing* 5(4), 277–298 (2009)
4. Freitas, C., Marreiros, G., Ramos, C.: IGTAI - an idea generation tool for ambient intelligence. In: *3rd IET International Conference on Intelligent Environments*, 391–397 (2007)
5. Le Gal, C.: Smart Environments: Technology, Protocols and Applications. Wiley (2005)
6. Mikulecký, P.: Remarks on ubiquitous intelligent supportive spaces. In: *Recent Advances in Applied Mathematics and Computational and Information Sciences: Proc. of the 15th American Conference on Applied Mathematics and Proc. of the International Conference on Comp. and Information Sciences*, Vols. I and II. WSEAS Press, Athens, 523–528 (2009)
7. Mikulecký, P., Olševičová, K., Bureš, V., Mls, K.: Possibilities of ambient intelligence and smart environments in educational institutions. In: Handbook of Research on Ambient Intelligence and Smart Environments: Trends and Perspectives. Information Science Reference/IGI Global, 620–639 (2011)
8. Mikulecký, P., Olševičová, K., Bureš, V., Mls, K.: Ambient intelligence in decision support. In: *Strategic Management and Its Support by Information Systems, Proceedings of 7th Intl. Conference.* VSB-Tech. Univ. Ostrava, 48–58 (2007)
9. Ramos, C., Augusto, J., Shapiro, D.: Ambient intelligence the next step for artificial intelligence. In: *IEEE Intelligent Systems*, 15–18 (2008)
10. Ramos, C., Marreiros, G., Santos, R., Freitas, C.: Smart offices and intelligent decision rooms. In: Handbook of Ambient Intelligence and Smart Environments. Springer Science+Business Media, Heidelberg, 851–880 (2010)
11. Shi, Y., Qin, W., Suo, Y., Xiao, X.: Smart classroom: Bringing pervasive computing into distance learning. In: Handbook of Ambient Intelligence and Smart Environments. Springer Science+Business Media, Heidelberg, 881–910 (2010)

The Usage of "Formal Rules" in the Human Intelligence Investigations

Tzu-Keng Fu

Department of Computer Science,University of Bremen
Bremen, Germany
tzukeng.fu@gmail.com

Abstract. As the result of this paper, we will say that the usage of the formal rules in unconscious processes is possible for the study of human intelligence by an investigation proposed by psychoanalyst Matte Blanco raised in the second half of the 20^{th} century. That is what we would like to propose for the usage of "formal rules" in the whole picture of the human intelligence to against the Dreyfus' arguments for anti-mechanism. In this paper, we explore the bi-logic framework with the general rules of unconscious processes. The anti-mechanism in this way should have a re-examination in an interdisciplinary manner with psychoanalysis.

Keywords: bi-logic framework, Matte Blanco theory, paraconsistent logic, unconscious processes

1 Introduction

A certain sense of anti-mechanism was proposed by Dreyfus that human intelligence and expertise rely mainly on the unconscious instinct than conscious symbol manipulation. And, these unconscious skills could never be taken in formal rules. ([9], [10]) However, like some psychoanalysts, we do not agree with the second half of this conclusion. In the studies of the unconscious processes, a formal perspective for unconscious processes by some very basic tools provided in modern logic was taken in Matte Blanco's theory (MBT, thereafter) on the logic of unconscious. ([6], [7], [18]) It has shed new lights to the formulation of the basic rules in unconscious: the principle of symmetry and the principle of generalization, by some basic set-theoretical notions in the modern logic.

2 Bi-logic Framework and MBT

Two principles were observed in the unconscious processes as follows:

(i) **Principle of Generalization**. The System Unconscious treats an individual thing (person, object, concept) as if it were a member or element of a set or class which contains other members; it treats this set or class as a subclass of a more general class, and this more general class as a subclass or subset of a still more general class, and so on. ([6], p.38)

(ii) **Principle of Symmetry**. The System Unconscious treats the converse of any relation as identical with the relation. In other words, it treats asymmetrical relations as if they were symmetrical. (ibid.)

The principle of generalization could be attributed to the set theory, of which it fits well with the axiom schema of separation in ZFC ([13], p. 3); the principle of symmetry could be attributed to a logical system with an inference rule of symmetrization. It is debatable whether a system with these two rules is deserved to be “a logic” from the trend of logic nowadays ([2], [11]), nevertheless it is not a problem to have a description for the unconscious processes in rule. Generally speaking, the bi-logic framework represents the inter-connection between the kind of logical mode (asymmetric mode) that people has followed in ordinary and other parts of human minds whose behavior is not governed by this mode but instead another type of logical mode (symmetric mode).

To study the unconscious processes and emotional processes in human beings become potentially reliable by the bi-logic framework. MBT, in literature, is one research to deal with unconscious and logic (or the logic of unconscious) by describing the Freudian unconscious processes with some basic notions in the mathematical logic. ([6], [7], [18]) It takes the bi-logic framework to do a two poles structural analysis of mental structure by proposing a distinction of conscious processes and unconscious processes in Freudian psychoanalysis theory, moreover to explain the five main Freudian characteristics for unconscious processes: the timelessness, the replacement of external by internal reality, the condensation, the displacement, and the absence of mutual contradiction, as arising out of the symmetrization as the formulation of principle (ii). In other words, the five characteristics of unconscious processes functioning have been re-formulated as arising out of the symmetrization in thought. MBT reflected the bi-logic framework naturally.

MBT defined the fundamental processes by a few precise mathematical notions: sets, numeration, symmetry, asymmetry, dimension and infinity, whose interactions will be further used in order to describe other complex dynamic mental processes. As we have mentioned above, MBT let the human minds be usefully conceived as partly functioning with the idea of two distinct polarized modes of knowing in a bi-logic perspective such that the approach it takes will characterize the unconscious processes precisely, and will moreover formulate the so called the "logical mode" of unconscious processes in a mathematical sense. With the conventional logical mode in the conscious processes, the whole mental life further described in the manner of stratification. So, MBT has described the overall structural aspects of mind *stratified* in a mathematical sense as follows:

(1) conscious and well-delimited objects stratum,
(2) more or less conscious emotions stratum,
(3) symmetrization of the class stratum,
(4) formation of wider classes which are symmetrized stratum,
(5) the mathematical limited stratum.

Based on the mathematical sense that MBT has taken, we explore the potential transition between two logical modes formulated such that the study about the transition between unconscious and conscious processes mentioned in MBT will be addressed. Such transition is another building block of bi-logic framework, whilst emotions serve as this building block in MBT. In this way, we study the transition between the underlying logics by the logic translation investigations in modern trend of theoretical computer science and mathematical logic. ([14], [15], [16]) We formulate these logics as the paraconsistent logics by the Freudian negation and the contradictions found in the human minds with the help of the argument in [4]. In brief, there are three common basic definitions of paraconsistency ([3], [5], [17]):

(a) with the rejection of *ex-contradictio sequitur quod libet* (EC)
(b) with the distinction between triviality and inconsistency
(c) with the semantic formulation of the principle of contradiction (PC)

By these definitions, we get that the paraconsistency should hold in the logical mode of unconscious processes since the observable phenomena in unconscious processes, the absence of mutual contradiction and negation and the co-presence of contradiction. We further get that bi-logic framework should also hold the same property. Emotion, which has been conceived as the combination of thinking and feeling by MBT serves as the place of transition happening between thinking (conscious/ asymmetric mode in general) and feeling (unconscious/ symmetric mode in general) and would be studied by the logic translation between paraconsistent logics.

Note 1. A similar idea, based on different methodology, has also been proposed by Hameroff ([12], p. 230), i.e., transition from nonconscious (superpositioned quantum information) to classical information, with consciousness the transition itself.

3 Conclusion

As the result of this paper, we have indicated a potential formal method for the bi-logic framework as the instance of formal investigations in the human unconscious processes. Nevertheless, it is not sufficient to be as a sound argument to against Dreyfus's anti-mechanism argument that “the unconscious skill can never be taken in formal rules.” However, the rapid development of various human intelligence investigations makes us still optimistic that Dreyfus' proposal can be eliminated in practice. For example, the institution theory will serve as the new framework for study the logic translation interaction between the transition between two logic modes formulated in this essay, i.e., to see the logic translation between paraconsistent logics. ([14], [15], [16]) A potential works in theoretical computer science with unconscious studies in an interdisciplinary manner is to paraphrase this specific bi-logic framework in this way.

Finally, a debatable issue in the modern theories of mind should be argued before structuring the unconscious/conscious processes bi-logically in an interdisciplinary manner, that is whether the statements about structuring human minds can be justified such that we are able to consider unconscious/conscious processes distinction structurally.

References

1. Béziau, J-Y.: What is “Formal Logic” ?, *Revista Brasileira de Filosofia* 232 (2009)
2. Béziau, J-Y.: Logica Universalis, 2^{nd} edition, Birkhäuser, Basel (2007)
3. Béziau, J-Y.: Paraconsistent Logic!, *Sorites* 17, 17-26 (2006)
4. Béziau, J-Y.: S5 is a Paraconsistent Logic and so is First-Order Classical Logic, *Logical Investigations* 9, 301-309 (2002)
5. Béziau, J-Y.: What is Paraconsistent Logic? In: Batens D. (ed.) Frontiers of Paraconsistent Logic. Research Studies Press, Baldock, 95–111 (2000)

6. Blanco, M.: The Unconscious as Infinite Sets: An Essay in Bi-Logic, Duckworth, London (1975)
7. Blanco, M.: Thinking, Feeling, and Being, Routledge, London (1988)
8. Bomford, R.: The Symmetry of God, Free Association Books (1999)
9. Dreyfus, H.: What Computers Can't Do: The Limits of Artificial Intelligence (revised) (1979)
10. Dreyfus, H.: What Computers Can't Do: The Limits of Artificial Intelligence (1972)
11. Gabbay, D.: What is a Logical System?, Oxford University Press (1994)
12. Hameroff S.: Consciousness, Neurobiology and Quantum Mechanics: The Case for a Connection. In: Tuszynski, J. (ed.), The Emerging Physics of Consciousness. Springer-Verlag, 193-253 (2007)
13. Jech T.: Set Theory: The Third Millennium Edition, Revised and Expanded. Springer (2003)
14. Kutz, O., Mossakowski T., and Dominik L.: Carnap, Goguen, and the Hyperontologies, Logica Universalis, *Special Issue on "Is Logic Universal?"* 4(2), 255-333 (2010)
15. Mossakowski T., Goguen J., Diaconescu, R., and Tarlecki, A.: What is a Logic Translation?, *Logica Universeralis* 3(1), 95-124, 2009.
16. Mossakowski T., Goguen J., Diaconescu, R., and Tarlecki, A.: What is a Logic? (revised version) In: Béziau, J.-Y. (ed.), Logica Universalis. Birkhäuser Verlag Basel, 111-133 (2007)
17. Priest, G.: Paraconsistency and Dialetheism. In Gabbay, D. and Woods J. (eds.), Handbook of the History and Philosophy of Logic (2000)
18. Rayner, E.: Unconscious Logic: An Introduction to Matte Blanco's Bi-Logic and its Uses, Harvard University Press (1995)

Rationality {in|through|for} AI

Tarek R. Besold

Institute of Cognitive Science, University of Osnabrück
49069 Osnabrück, Germany
tbesold@uni-osnabrueck.de

Abstract. Based on an assessment of the history and status quo of the concept of rationality within AI, we propose to establish research on (artificial) rationality as a research program in its own right, aiming at developing appropriate notions and theories of rationality suitable for the special needs and purposes of AI. We identify already existing initial attempts at and possible foundations of such an endeavour, give an account of motivations, expected consequences and rewards, and outline how such a program could be linked to efforts in other disciplines.

Keywords: rationality, research program, human intelligence

1 Introduction: Rationality

Research in rationality and rational behavior has a rich history, both within more abstract disciplines like philosophy or economics, as well as in more individual-centered fields as psychology and cognitive science. Over the years, many quite distinct frameworks for modeling rationality (and establishing a normative theory) have been proposed. Breaking these distinct approaches down to their underlying theoretical foundations, four main types of models can be identified: logic-based models (cf. e.g. [1]), probability-based models (cf. e.g. [2]), heuristic-based models (cf. e.g. [3]), and game-theoretically based models (cf. e.g. [4]). Unfortunately, it shows that the definitions of rationality arising from within the distinct approaches are in many cases almost orthogonal to each other (as are the frameworks), and only partially cover human behavior and human-like rationality (cf. e.g. [5], [6], [7]).

2 AI & Rationality

Rationality & (Natural) Intelligence In many, often folk-psychologically motivated accounts of human cognition, the notion of intelligence is mostly seen as closely intertwined with a concept of rationality. Whilst in many cases a certain form of intelligence is taken as a precondition for rationality and rational behavior, rational behavior and rationality also may be considered one possible indicator for an agent's intelligence. Often, both phenomena are even understood as following some direct proportionality (although there are positions questioning that view, cf. e.g. [8]).

Motivation Artificial intelligence offers a wealth of concepts and notions called rational or claimed to reflect rationality. Treating rationality within a framework of artificial intelligence actually may have good reasons: On the one hand, subscribing to a stance close to the human-like intelligence or "strong AI" research programs, modeling (human-style) rationality clearly has to be considered one of the milestones for reaching the overall goal of a "truly" intelligent AI system. Lately, a test for an artificial system's behavior in rationality tasks has even been proposed and accepted as a sub-task of a modernized decomposition [9] of Turing's famous Test for machine intelligence [10]. On the other hand, staying closer to the "weak AI" ideas, having a feasible concept of rationality within a system allows agents to interact with humans in a more natural way, predicting human's expectations and interactive behavior. Also, from a metaperspective, AI accounts of rationality might provide inspiration, modeling tools and testbeds for theories of rationality arising within related disciplines.

Earlier Work & Status Quo Several prominent researchers within the field of artificial intelligence explicitly addressed the relation between AI and rationality, as well as the role rationality can play within artificial intelligence (cf. e.g. [11], [12]). Also, certain conceptions of rationality have played an important role at different turning points within the development of AI as a scientific field, sometimes emerging from new paradigms within AI, sometimes directly contributing to the creation of new stances and perspectives (cf. e.g. [13] for a compact treatment of three main positions).

The different families of models for rationality listed in Sec. 1 (partly with exception of the heuristics-based approach) can presently be found within existing AI systems and theories. Still, when having a closer look at the latter, it shows that the underlying notions of rationality have stayed close to their fields of origin, having been adapted only to a minimal degree (if at all). Also, possible deficits and shortcomings have been brought along without questioning or fixing. In consequence, AI systems e.g. mostly fall short in tasks such as predicting or exhibiting behavior resembling human-like rationality, a crucial need in all domains concerned with close interaction/cooperation between a human user and an AI system (rational agents communication for cooperative dialogues [14], adaptive and cooperative wheelchairs [15], etc.).

3 The Vision: Rationality for AI

Recently, some researchers in cognitive science and decision theory (cf. e.g. [16], [17], [8]) are questioning the completeness and suitability of the classical approaches to rationality. One possible consequence of this stance is the proposition to use humans as gold-standard for actually existing rational agents, and consequently base models of rationality and rational behavior on cognitive capacities. Clearly, this brings along a fundamental shift in the type of theory, aiming not for theories of normative nature, but trying to build a positive theory of human rationality (as e.g. also mentioned in [18]).

First theoretical and empirical studies on using computational cognitive systems for these modeling tasks seem to provide support for the practical possibility and valuable applicability of that approach (cf. e.g. [19], [20]).

Following that example, additionally being inspired from within AI e.g. by [21], we argue for establishing work on rationality as a proper, coordinated research program within artificial intelligence, not only limiting its focus to taking theories of rationality from respective neighboring disciplines and adapting them so that they can be applied within an AI framework, but to actively pursue research on (artificial) rationality on its own.

The advantages of such an approach seem significant: From an AI-internal perspective, it allows for approaching rationality within AI from a more holistic perspective, possibly amalgamating existing accounts into theories and conceptions more suitable for the use in artificially intelligent systems (as, e.g., already to be rudimentarily found in the development of probabilistic dynamic epistemic logics [22], an attempt at integrating probabilistic and logical views of rationality), and also providing feedback to neighboring disciplines, directly guiding further research in rationality concepts for AI (playing an integrative role similar to the artificial general intelligence program within the development of "strong AI", cf. e.g. [23]). From a more high-level point of view, it gives a basis for closer cooperation with neighboring fields, allowing for more and deeper (also not only unidirectional, but bidirectional) interaction. Such joint efforts might span a very wide range of scenarios:

- Concrete project-based joint ventures for example in human-computer interaction.

– Rather specific applications of AI methods and systems for mitigating shortcomings of theoretical frameworks in economics (cf. e.g. [24]).
– Inspiration, modeling tools and testbeds for theories of rationality within other disciplines (as, for example, it has already been the case in the work of Herbert Simon, cf. e.g. [25]).
– Direct new contributions to longstanding concepts of rationality (cf. e.g. the "algorithmic rationality" attempts at offering a framework for including computational costs in otherwise game-theoretical notions of rationality [26]).
– Fundamental research on human-like rationality conducted together with researchers on the cognitive science-side of system design and modeling, epistemology and philosophy of cognition (cf. e.g. [27]).

4 Conclusion

Summarizing, we want to put forward the claim and demand that targeted research on theoretical and applied aspects of rationality within AI should be conducted, using methods from AI and, in the long run, aiming at solving some of the fundamental problems the overall artificial intelligence program tries to answer. Whilst research on rationality has – for historical, but also for social reasons within scientific communities – by now mostly been conducted almost exclusively in the fields of economics, decision theory and psychology, we are convinced that a proper research program within AI could be fruitful not only for the purposes of AI itself (here also allowing for new approaches and takes on long-standing challenges), but also would provide valuable feedback and new perspectives for the already established "players in the field of rationality".

Acknowledgments

The author wishes to thank Maricarmen Martínez Baldares for her helpful recommendations and comments on an earlier draft.

References

1. Evans, J.: Logic and human reasoning: An assessment of the deduction paradigm. *Psychological Bulletin* 128, 978–996 (2002)

2. Griffiths, T., Kemp, C., Tenenbaum, J.: Bayesian Models of Cognition. In: The Cambridge Handbook of Computational Cognitive Modeling. Cambridge University Press (2008)
3. Gigerenzer, G., Hertwig, R., Parchur, T., eds.: Heuristics: The Foundation of Adaptive Behavior. Oxford University Press (2011)
4. Osborne, M., Rubinstein, A.: A Course in Game Theory. MIT Press (1994)
5. Tversky, A., Kahneman, D.: Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review* 90(4), 293–315 (1983)
6. Wason, P.C., Shapiro, D.: Natural and contrived experience in a reasoning problem. *The Quarterly Journal of Experimental Psychology* 23(1), 63–71 (1971)
7. Arrow, K.: Rationality of self and others in an economic system. *The Journal of Business* 59(4), 385–399 (1986)
8. Gilboa, I.: Questions in decision theory. *Annual Reviews in Economics* 2, 1–19 (2010)
9. Besold, T.R.: Turing revisited. In: Philosophy and Theory of Artificial Intelligence (PT-AI 2011). (2011)
10. Turing, A.: Computing Machinery and Intelligence. *Mind* LIX(236), 433–460 (1950)
11. Russell, S.: Rationality and intelligence. *Artificial Intelligence* 94, 57–77 (1995)
12. Doyle, J.: Rationality and its roles in reasoning. *Computational Intelligence* 8(2), 376–409 (1992)
13. Clark, A.: Artificial Intelligence and the Many Faces of Reason. In: The Blackwell Guide to Philosophy of Mind. Blackwell (2003)
14. Sadek, M.D., Bretier, P., Panaget, F.: ARTIMIS: Natural dialogue meets rational agency. In: *Proceedings of IJCAI-97*, Morgan Kaufmann, 1030–1035 (1997)
15. Galluppi, F., Urdiales, C., Sandoval, F., Olivetti, M.: A study on a shared control navigation system: human/robot collaboration for assisting people in mobility. *Cognitive Processing* 10(2), 215–218 (2009)
16. Kokinov, B.: Analogy in decision-making, social interaction, and emergent rationality. *Behavioral and Brain Sciences* 26(2), 167–169 (2003)
17. Besold, T.R., Gust, H., Krumnack, U., Abdel-Fattah, A., Schmidt, M., Kühnberger, K.: An argument for an analogical perspective on rationality & decision-making. In von Eijck, J., Verbrugge, R., eds.: *Proceedings of the Workshop on Reasoning About Other Minds: Logical and Cognitive Perspectives (RAOM-2011)*, Groningen, The Netherlands. Volume 751 of CEUR Workshop Proceedings, CEUR-WS.org (2011)
18. Binmore, K.: Rational Decisions. Princeton University Press (2011)
19. Petkov, G., Kokinov, B.: JUDGEMAP - Integration of analogy-making, judgment, and choice. In: *Proceedings of the 28th Annual Conference of the Cognitive Science Society (CogSci)*, 1950–1955 (2006)
20. Gust, H., Krumnack, U., Martinez, M., Abdel-Fattah, A., Schmidt, M., Kühnberger, K.: Rationality and general intelligence. In: *Proceedings of the Fourth Conference on Artificial General Intelligence (AGI-11)*. (2011)

21. Russell, S.J., Wefald, E.: Do the right thing - studies in limited rationality. MIT Press (1991)
22. Kooi, B.P.: Probabilistic dynamic epistemic logic. *Journal of Logic, Language and Information* 12, 381–408 (2003)
23. Wang, P., Goertzel, B.: Introduction: Aspects of artificial general intelligence. In Goertzel, B., Wang, P., eds.: Advances in Artificial General Intelligence: Concepts, Architectures and Algorithms - Proceedings of the AGI Workshop 2006. Volume 157 of Frontiers in Artificial Intelligence and Applications., IOS Press, 1–16 (2007)
24. Dixon, H.: Some Thoughts on Economic Theory and Artificial Intelligence. In: Surfing economics: Essays for the enquiring economist. Palgrave (2001)
25. Frantz, R.: Herbert Simon. Artificial intelligence as a framework for understanding intuition. *Journal of Economic Psychology* 24, 265–277 (2003)
26. Halpern, J.Y., Pass, R.: Algorithmic rationality: Adding cost of computation to game theory. *ACM SIGecom Exchanges* 10(2), 9–15 (2011)
27. Pollock, J.: Epistemology, rationality, and cognition. In Dancy, J., Sosa, E., Steup, M., eds.: A Companion to Epistemology: Second Edition. Wiley-Blackwell (2010)

Selfish Genes and Evolutionary Computation

Jan Zelinka

University of West Bohemia, Faculty of Applied Sciences, Dept. of Cybernetics
Univerzitní 8, 306 14 Plzeň, Czech Republic
zelinka@kky.zcu.cz

Abstract. This paper focused on the description of the Evolutionary Computation (EC) and its role in artificial intelligence (AI). We will try to describe parts of the evolution theories which are important not only for explanation of certain biological phenomena but also for evolution realized by a computer algorithm or even for the concept of evolution in general. We will present our opinion about the role of reproduction in the origin of evolution and selfishness. We would like to point out that the introduction of fitness function eliminates every possible kind of selfish gene, may be fundamentally wrong. Some space will be devoted to describing the problem of selfish gene in EC and the metaevolution. We will try to erode the common believe that metaevolution is powerful method with only technical limits. A selfish gene by itself cannot be blindly regarded as an ultimate contribution which makes way for the true AI from an EC.

Keywords: selfish gene, genetic algorithms, evolutionary computation

1 Evolutionary Computation

This paper would first like to focus on the description of the Evolutionary Computation (EC) and its role in artificial intelligence (AI) [1, 2]. First and foremost, we would like to specifically emphasize that we do not see neither the EC as an omnipotent optimizing method nor the selfish gene [3, 4] as a powerful concept which can provide a simple holistic model of the world. Instead, we intend to present some clues that could lead us to more reserved but still positive view on at least the EC. We will start the paper with reflection on the role of the evolutionary computation in AI when AI is seen as an optimization problem. The first thesis, for which we will provide some arguments, is the assertion that the EC is an approximation of the (generally unknown) optimal solution method. The criticism of the believe that such a solution exists and we are only not able to find it analytically (or, alternatively, we see the analytical solution to be difficult to search for) was our motivation to propose an antithesis which states that the result found by an EC algorithm is not a mere approximation of an ideal analytical solution but instead a new and original solution by itself. It is exactly this antithesis that supports the (traditional) view of the EC as a remarkable approach to AI problems. End of the first part will be devoted to the critique of this antithesis.

2 Concept of Selfishness

Concepts of evolution, gene and selfishness which are seen as abstract structures or abstract properties of abstract structures will be discussed in the second part. We will try to describe parts of the evolution theories which are important not only for explanation of certain biological phenomena but also for evolution realized by a computer algorithm or even for the concept of evolution in general. This part ends with description of notions in evolution theories that allow to grasp a concept of selfishness. We will consider the role of a fitness function which is seen as foreign and inappropriate element in the biological evolution theories, yet even those theories cannot avoid taking it into account when the concept of selfishness or evolution stability is defined against the concept of optimality. On the other hand, the fitness function plays a central role in the EC algorithms.

3 Origin of Selfishness

The next part is focused on interpretation of the selfish gene concept and on hypothesis about the origin of selfishness as a phenomenon. Principles and different kinds of selection mechanisms are discussed in this part. We will present our opinion about the role of reproduction in the origin of evolution and selfishness. It will be shown that it is just the mechanism of reproduction that is the main or possibly the only cause of evolution and even selfishness. Several thought experiments and computer simulations will be presented in order to show the clues that led us to this opinion, i.e. that evolution and even selfishness are possible without the concept of selection and also without communication or interaction between individuals (genes or their vehicles), of course with an exception of the reproduction contact itself.

4 Selfish Gene and Evolutionary Computation – State of the Art

Next, we will deal with selfish gene as a phenomenon in EC and its reflection in related literature. First, we want to draw attention to the fact that, even in extensive monographs, any connection between the selfish gene notion and the EC is usually totally disregarded or the selfish gene concept is just mentioned in passing without the real consequences for the EC algorithms. We believe that it is because most authors presume that the introduction of a fitness function totally eliminates selfish gene and its harmful effects. Fortunately, there are certain significant exceptions to this approach [5–9] and thus the short resumes of the publications dealing with selfish gene concept within EC will be presented the end of this chapter.

5 Selfish Gene, Evolutionary Computation and Metaevolution

We will further try to describe the formation of a selfish gene phenomenon in EC in the light of findings described previously. Let us stress out that we do not intend to describe a way in which an EC algorithm with a selfish gene can be programmed. Instead, we would like to point out that the assumptions introduced in the previous chapters, i.e. that the introduction of fitness function eliminates every possible kind of selfish gene, may be fundamentally wrong. The genesis of a selfish gene in EC will be shown in several experiments, followed by a possible explanation of this fact and some generalizations. Then some space will be devoted to describing the problem of selfish gene in EC and the metaevolution, i.e. evolution of EC algorithm. Metaevolution is something that cannot be avoided altogether by any EC algorithm because every such algorithm has some parameters to be appropriately adjusted and something like small evolution is usually applied, though not intentionally. But on the other hand, this metaevolution process is usually not algorithmized. We will try to erode the common believe that metaevolution is powerful method with only technical limits (e.i. CPU speed or computer cluster size) which will be sooner or later removed. We see the role of a selfish gene in EC mainly in the fact that it sets a limit on the metaevolution and especially on joining of the evolution and metaevolution which is performed by genes including some EC parameters such parameters concerned with reproduction. The possibility of selfish gene usage in the EC will be discussed in the last part. We will present our opinion that selfish gene occurrence does not automatically ensure emergence of nature in EC and most probably does not even ensure that emerged artificial intelligence will appear to be natural. Consequently, a selfish gene by itself cannot be blindly regarded as an ultimate contribution which makes way for the true artificial intelligence from an EC. However, we will try to write our opinion about some alternative ways towards this goal.

6 Conclusion

In our paper the role of selfish gene in EC is considered. Our hypotheses about the origin of selfishness and our experiments lead us to opinion that the selfish gene can not be excluded from the EC. We will demonstrate the potential utility of the effects of selfishness in several examples. But the main importance of selfishness we see in the fact that selfishness highly limits a metaevolution for which will try to give some arguments and examples.

Acknowledgments

This research was supported by the Ministry of Education of the Czech Republic, project No. MŠMT LC536.

References

1. Altenberg, L.: The evolution of evolvability in genetic programming (1994)
2. Koza, J.R.: Genetic programming (1997)
3. Dawkins, R.: The selfish gene. Popular Science. Oxford University Press (1989)
4. Dawkins, R.: Climbing mount improbable. Penguin science. Penguin (1997)
5. Corno, F., Sonza Reorda, M., Squillero, G.: The selfish gene algorithm: a new evolutionary optimization strategy. In: *SAC: ACM Symposium on Applied Computing*, 349–355 (1998)
6. Zhang, J., Bushnell, M.L., Agrawal, V.D.: On random pattern generation with the selfish gene algorithm for testing. In: *Digital Sequential Circuits, in Proc. International Test Conf.*, 617–626 (2004)
7. Onoyama, T., Maekawa, T., Sakurai, Y., Tsuruta, S., Komoda, N.: Selfish constraint satisfaction genetic algorithm for planning a long-distance transportation network. JCP 3(8), 77–85 (2008)
8. Barta, Z., Flynn, R., Giraldeau, L.A.: Geometry for a selfish foraging group: a genetic algorithm approach. Proceedings of the Royal Society London, B 264, 1233–1238 (1997)
9. van der Meijs, F.: Implementation of the parallel selfish gene genetic algorithm (2008)

The Role of Externalisation in Human Enhancement: Are All Our Thoughts in Our Head?

Eva Zackova[1], Jan Romportl[2], Marek Havlik[1], and Michal Polak[1]

[1] Department of Philosophy, Faculty of Philosophy and Arts, University of West Bohemia, Pilsen, Czech Republic
{zacka, mshavlik, miha}@kfi.zcu.cz

[2] Department of Cybernetics, Faculty of Applied Sciences, University of West Bohemia, Pilsen, Czech Republic
rompi@kky.zcu.cz

Abstract. The paper discusses Clark's conception of extended mind and critically analyses his four criterions of externalised cognitive functions. On the basis of the interdisciplinary research in the fields of AI and cognitive science it proposes a redefinition of these criterions. The theoretical claims are supported by Gazzaniga's experiments with split-brain phenomena and our development of AI-based speech prosthetics for laryngectomy patients.

Keywords: human enhancement, extended mind, artificial intelligence, split-brain phenomena, speech synthesis, laryngectomy

1 Introduction

In computer science we can trace prognoses of man-machine symbiosis back to the 1960s when conceptions of so-called intelligence amplification occurred [1, 2]. They intended to reach the man and computer technology mergence in order to support and enhance current cognitive functions of human organism while further preserve actual human experience. Thereby the influence of the means used by human to facilitate pragmatic and epistemic actions (through artefacts and especially language) has been emphasised and opened to further discussion in the field of theory of mind.

The idea of mind being tightly coupled with its environment has already been — at least since the Maturana and Varela's autopoietic theory [3] — perceived and acclaimed as a sound and plausible alternative to cognitivistic theories of mind. The argumentation in favour of the extended mind [4, 5] is compatible with this paradigm of understanding of mind, and we believe that it is very important in the discussion about possibilities and frontiers of human enhancement.

In addition to theoretical analysis of the extended mind model, our paper would like to support our claims about externalisation of "truly internal" cognitive functions with examples of neuroscience experiments with split-brain phenomena [6], and to illustrate our claims on an example of an AI-based system whose aim is to be able to adequately and in real time generate prosody for speech of patients after laryngectomy, who lost their ability to generate speech without external devices and to express their own prosody at all.

2 Extended Mind Conception

We have focused on the conception of extended mind, which was firstly published in Clark's and Chalmers' essay *Extended Mind* (1998) and later more deeply elaborated by Clark in *Supersizing the Mind* (2008). The idea of extended mind is frequently compared with the embodied perspective [7], which Clark calls brainbound model because from this point of view all human cognition depends directly on neural activity alone, and mind processes are reduced just to a brain or neural activity [5].

Clark based his extended model on so-called parity principle, which held that "if a process in the world works in a way that we should count as a cognitive process if it were done in the head, then we should count it as a cognitive process all the same" [5]. In brief, "if a certain state plays the same causal role in the cognitive network as a mental state, then there is a presumption of mentality" [5]. Hence we can infer a thesis that not all our thoughts, not all our cognition have to be located in our head.

3 Criterions of External Cognitive Process

Clark defines four basic criterions of what he still counts as a cognitive process even though it is realized outside our brain. In our research we came to a conclusion that they actually do not support the extended mind conception very well. As a result we redefine them and suggest our own "hard-wired system criterion".

The first Clark's criterion says that the resource must be reliably available and typically invoked [5]. The criterion of typical invocation means that one would never say "I don't know, I have to check my external memory" when one needs to consult an external resource to get the answer. Instead, the resource is automatically invoked, and only in the case that the answer is not found, then one would say that s/he does not know — just like in case of biological memory.

The second criterion says that "any information thus retrieved must be automatically endorsed and trustworthy as something retrieved clearly from biological memory" [5]. Automatic endorsement criterion seems reasonable to us, even though we believe that the external memory can be even more trustworthy than the biological one. However, the key issue, which is commonly ignored, is that this criterion is related just to a *memory function*. From our point of view, this is a serious deficiency of Clark's theory. Obviously, we know other types of cognitive processes which are more complex than retrieving information from our memory — as for example in case of speech competence, attention, recognition, learning, reasoning and so on.

The third criterion says that "the information contained in the resource should be easily accessible as and when required" [5]. We believe that this one is plausible enough, it is quite close to the first criterion and we have accepted them both.

The last criterion claims that "the information must have been consciously endorsed at some point in the past and be there as a consequence of this endorsement" [5]. We strongly disagree with this condition. According to us, the information does not have to be consciously endorsed every time. Firstly, a general endorsement can be done instead of a recurrent particular endorsement. For instance, we can decide to trust all information from a particular source. Hence we do not have to make the endorsement for each particular piece of information or process. Secondly, and more importantly, we hold that we can externalise the endorsement itself (for example in cases of attention disorders or in case of certain levels of externalised speech competence).

4 Hard-Wired System Criterion

Since we regard Clark's criterions as unsatisfactory and not enough supportive for the extended mind conception, we have suggested a criterion of a "hard-wired system" (or simply "hard-wired system criterion"), which is still based on the aforementioned parity principle and Clark's original requirements that we consider as reasonable, such as easy accessibility, reliability and typical invocation of the resource, but which we have enriched with mutual adaptation (of the externalised system and of the mind processes) and functional dependency (of the brain cognitive processes on the external cognitive processes).

To summarise our idea, the hard-wired criterion claims that we can speak about extended mind when functioning of the external system (which is based on the parity principle) influences functioning of the brain and mind in such a way that other mechanisms of the brain adapt themselves so that in case the system is disconnected, it will cause serious deficiencies in cognitive abilities of the person.

5 Split-Brain Drawing as an Example of the Externalisation

There is a great portion of scientific evidence that can be used as examples for our claims. For instance, Gazzaniga describes several surgeries that were performed in order to stop epileptic seizures of a person. Surgeries of this type involved sectioning the *corpus callosum* (a bundle of neural fibers connecting the left and right cerebral hemispheres) of the person's brain, thus creating the split-brain phenomena.

Sophisticated experiments based on the separate stimulation of the left or the right hemisphere showed that if the corpus callosum is sectioned, the patient is not capable of speaking about a stimulus in his right hemisphere and neither he is aware of it. If the stimulus is shown to his right hemisphere and the patient is asked what he saw, the patient usually answers that he saw nothing [6].

For example, a stimulus of a house is put to the right brain hemisphere and because the patient is not aware of the stimulus and is not able to speak about it, he is asked to draw it. He starts to draw a picture of the house with his left hand that is controlled by the right hemisphere. When the picture is finished, he can use the picture as a piece of information for the left hemisphere and thus is able to say that it is a house [6].

Corpus callosum is a device for transferring information from one hemisphere to the other but has nothing to do with memory function. Externalisation of this transferring function through drawings that stand as stimulation for the left hemisphere is a good example of externalisation of cognitive functions and also a good example of the parity principle.

6 Speech Prosthetics as a Real-Life Application of Extended Mind

To foster the extended mind conception defined by our hard-wired system criterion more deeply, we illustrate it by a concrete example of a real-life application.

We work on a system that will allow patients after total laryngectomy to use intelligent speech prosthetics based on TTS (text-to-speech), ASR (automatic speech recognition) and AI technologies [8].

The system will transform the distorted and hardly comprehensible speech of laryngectomees to naturally sounding synthesised speech. However, the patients will have absolutely no control over the output prosody, which will be completely estimated by the artificial system [9]. This means that the internal cognitive process, normally responsible for prosody generation, will no longer be used and will be substituted by the external cognitive process realized by the speech prosthetic system.

We believe that this is a good example of AI-based externalisation of a cognitive function of human mind and that it could be regarded not just as treatment but also as enhancement. Furthermore, this example involves a cognitive function that is more complex than in case of memory, which gives further evidence for our critics of Clark's memory-oriented criterions of external cognitive processes.

7 Conclusion

We believe that we can provide enough empirical evidence for the conception of extended mind and for validity of the assertion that our thoughts can be not only in our heads. On the basis of our work in the field of AI we can claim that systems actively externalising cognitive functions are not waiting hidden in the distant future but can most probably be achievable even now through a fostered symbiosis of human and complex state-of-the-art AI systems.

Acknowledgments

The paper was supported by the University of West Bohemia under the grant SGS-2011-042.

References

1. Engelbart, D. C.: Augmenting Human Intellect: A Conceptional Framework. Prepared for: Director of Information Sciences Air force Office of Scientific Research, Washington (1962)
2. Licklider, J. C. R.: Man-Computer Symbiosis. In: Taylor, R. W. (ed.) In Memoriam: J.C.R. Licklider: 1915-1990. *Digital Systems Research Center Reports* 61, Palo Alto, CA (1990)

3. Varela, F. J., Thompson, E., Rosch, E.: The Embodied Mind. MIT Press, Cambridge MA — London (1991)
4. Clark, A., Chalmers, D.: The Extended Mind. *Analysis* 58, 1, 7-19 (1998)
5. Clark, A.: Supersizing the Mind: Embodiment, Action, and Cognitive Extension. Oxford University Press, New York (2008)
6. Gazzaniga, M.: The Social Brain. Basic Books, New York (1985)
7. Rupert, R.D.: Cognitive Systems and the Extended Mind. Oxford University Press, New York (2009)
8. Hanzlíček, Z., Matoušek, J.: Voice conservation: Towards creating a speech-aid system for total laryngectomees. In this volume.
9. Romportl, J.: On the Objectivity of Prosodic Phrases. *The Phonetician*, vol. 96, The International Society of Phonetic Sciences, 7-19 (2010)