AISB 96

Adaptation in Artificial and Biological Systems

Volume 1

Editors: Tim Kovacs and
James A. R. Marshall

University of BRISTOL

EPSRC
Engineering and Physical Sciences Research Council

Proceedings of AISB'06: Adaptation in Artificial and Biological Systems
Tim Kovacs and James A. R. Marshall (eds.)

ISBN 1 902956 97 5 – Volume 1
ISBN 1 902956 98 3 – Volume 2
ISBN 1 902956 96 7 – Volume 3

Published by the Society for the Study of Artificial Intelligence and the Simulation of Behaviour

Printed by the University of Bristol, Bristol, UK

Contents

Foreword

The Society for the Study of Artificial Intelligence and Simulation of Behaviour (SSAISB) is the UK's largest Artificial Intelligence society and, having been established in 1964, one of Europe's oldest. Consequently the SSAISB has a long history of sponsoring "Good Old Fashioned AI". The convention theme we have selected for AISB'06 is "Adaptation in Artificial and Biological Systems", a theme that reflects an approach to AI and related disciplines that is much more of the late 20th and early 21st centuries. We have chosen this theme both because it captures what we feel is an important zeitgeist in computer science, mathematics, engineering and the life sciences, and because Bristol has very strong interdisciplinary research groups in all these areas. Indeed, world-leading bio-mimetic research in the Bristol area dates back at least to the 11th century with the flight, and crash, of Eilmer of Malmesbury [1]. More recently Grey Walter built his first bio-inspired autonomous turtle robots in Bristol between 1948 and 1949.

The development of Western science since the 17th century has seen a transition from polymaths and a field of general science to rigid disciplinary boundaries. In the 21st century we expect this fragmented approach to science to tend back to the earlier holistic approach. Recent decades have seen exciting developments in the study of biology from a systems perspective, making use of advances in computing and mathematical techniques. At the same time there has been an increased interest in applying solutions from biological systems to engineering problems. For example the uncertain, massively parallel computing environment presented by the internet is much closer to the kind of environment biological systems interact with than the highly centralised computing paradigm initiated by Babbage when he designed the first calculation engines.

These recent developments in biological study and biological inspiration are a continuation of a much older tradition, the seed of which can be traced back very far indeed – Eilmer's flight was inspired by that of Daedalus and Icarus. More practical was Marc Burnel's tunnelling shield said to have been inspired by shipworms which bore through wood. The shield was first used to excavate a tunnel under the Thames which, begun in 1825, still carries the East London line of the London underground. Even a paradigm as sophisticated as artificial life was articulated as far back as 1787 when Goethe described to a friend the idea for an archetypal plant thus; "With this model and the key to it, one will be able to invent plants... which, even if they do not actually exist, nevertheless might exist and which are not merely picturesque or poetic visions and illusions, but have inner truth and logic. The same law will permit itself to be applied to everything that is living" (letter of 1787, quoted in [3], p.14).

The importance of adaptation has long been recognised by some of the greatest figures in the history of AI. In 1950, Alan Turing proposed that artificial intelligence be pursued through adaptation when he outlined his idea of an artificial child [2]. In 1975, John Holland published an influential monograph on 'Adaptation in Natural and Artificial Systems'; needless to say it is no coincidence that the theme for AISB'06 coincides with this title very closely! Adaptation is fundamental to AI, as adaptation is key to intelligence. Seemingly intelligent behaviour is brittle unless it is adaptive. Robust intelligence in biological systems has arisen as a result of adaptation on multiple levels, including evolutionary, social, and individual.

While biological inspiration has long played a role in engineering artificial systems, the flow of ideas and tools in the other direction has been increasing. Neural networks were inspired by brain function and in turn are used to model it. So too with reinforcement learning. Genetic algorithms were inspired by evolutionary processes and now form the basis of models used to investigate evolutionary theory. Algorithms derived from social insect research find applications in engineering, while computer models of insect colonies advance understanding of the decision-making capabilities of these natural systems. More radically, recent work on DNA and cellular computing has blurred the lines between the implementation details of artificial and natural systems.

The convention theme for AISB'06, and most of its constituent symposia, reflects this rich interaction between the study of adaptation in the artificial and the biological. We hope you find these proceedings to be an illuminating record of current research activity in the area. Before we leave you to enjoy them we would like to report the widespread support we received for retaining the format of AISB. Obtaining feedback on research in its early stages can clearly be of enormous benefit. The AISB conventions provide a venue in which work-in-progress can be presented, something that is all too rare in computer science, in contrast to the life sciences. At the same time the option to publish abstracts and papers under a non-exclusive copyright allows us to disseminate a record of the event.

Finally we would like to thank all the people who have helped us make AISB'06 possible, including the symposia organisers, members of the SSAISB committee, the organisers of AISB’05 and in particular those members of the University of Bristol who have assisted in various ways, from the Machine Learning and Biological Computation group, Department of Computer Science, and from the AI Group, Department of Engineering Maths. Special thanks go to Robert Egginton and Tobias Larsen for helping produce these proceedings, Sophie Benoit for administrative assistance, and the MLBC volunteers for manning registration desks and providing other help during the convention itself. We are grateful to the Engineering and Physical Sciences Research Council and to HP Labs for their support.

Tim Kovacs
James Marshall
March 14th 2006, Bristol, UK

References

[1] William of Malmesbury (c. 1125). *Gesta regum Anglorum / The history of the English kings*, ed. and trans. R.A.B. Mynors, R.M. Thomson, and M. Winterbottom, 2 vols., Oxford Medieval Texts (1998-9).
[2] A.M. Turing (1950). Computing machinery and intelligence. *Mind* 59, 433-460.
[3] B. Mueller and C.J. Engard (1952). *Goethe's Botanical Writings*. Honolulu HI: University of Hawaii Press.

Evolutionary Design of Complex Chemical Systems

Mark A. Bedau
Protolife SRL, Venice
Reed College, Portland OR, USA
European Centre for Living Technology, Venice

Abstract

Complex chemical systems are difficult to design, largely because of unanticipated and unwanted side reactions. This is an instance of the well-known reality gap problem that afflicts evolutionary design. This talk describes a very general design method that circumvents the reality gap. The method is illustrated in two contexts: a dissipative particle dynamics (DPD) model chemistry for self-assembling lipid structures, and a laboratory realization of such a chemistry. This methodology has commercial value as a general and automated high-throughput method for creating such things as designer biosensors or designer liposomes for drug delivery. It also shows significant scientific promise as a way to attack the holy grail of wet artificial life - creating a living artificial cell wholly from non-living chemical materials.

Life and Mind: A Union without Divorce?

Maggie Boden
Centre for Research in Cognitive Sciences
University of Sussex

Abstract

Why should people primarily interested in the human mind bother with A-Life? Why can't psychologically-oriented cognitive scientists ignore these technological advances? There are two answers, one more problematic than the other.

The first answer is that A-Life, considered as a methodological sub-species of AI, studies a number of phenomena that are psychologically interesting. These include distributed cognition (ant trails, and the like), situated robotics (motor behaviour in cockroaches, for instance), computational neuroethology (such as aspects of mating in crickets and hoverflies), and evolutionary systems (from co-evolution in Tierra to evolutionary computer art). 'Natural' ethology and psychology can learn a lot from artificial models of such matters.

The second answer is that mind requires life. If that's true, it follows that the former can't be properly understood without understanding the latter. One could even say that AI is, at least in principle, a sub-species of A-Life.

That mind requires life is often stated, and even more commonly assumed. Certainly, all the minds we know about are found in living things. But the link, and its supposed necessity, is very rarely explicitly argued. And when it is, the arguments are usually pretty thin.

One problem here is that the concept of life itself is unclear. Does it necessarily involve embodiment, for example? And if so, why? How does embodiment differ from mere physicality, which every robot satisfies? Is situatedness enough? Or is metabolism needed too?

Even assuming that we know what we mean by "life", or anyway that we can recognize it when we see it, what has it got to do with mind? In autopoietic theory, for instance, the life-mind link is repeatedly stated but not clearly justified. In evolutionary philosophies of intentionality, it is tacitly assumed. If one accepts such accounts, one must conclude that evolutionary A-Life may be highly relevant for understanding mind. (That it is relevant for understanding neural selection is a different, though related, matter.)

From Individual to Collective Intelligence

Nigel R. Franks
School of Biological Sciences
University of Bristol

Abstract

Social insect colonies exhibit both individual and collective intelligence. I will illustrate this with decision-making during house hunting in rock ants. Each worker, among the 200 or so in a colony, has less than 100,000 neurones (compared to 1011 neurones in humans) yet these ants employ the most sophisticated of all consumer strategies when choosing a new nest. Indeed, they can choose the best-of-N among alternative nests even though each has many different and important attributes. I will show how information can cascade through these social networks enabling colonies to benefit from both individual and collective intelligence. They use quorum sensing to facilitate collective intelligence and to achieve flexible speed accuracy trade-offs. These ants also fulfil all of the criteria of teaching. We have been able to show teaching through experimental manipulations and detailed quantitative analyses. Certain information is so valuable in these ant societies that individual workers conserve and propagate it by teaching others. Indeed, these ants and humans are the only animals in which teaching has been demonstrated.

Artificial Consciousness and the Simulation of Behaviour

Owen Holland
Department of Computer Science
University of Essex

Abstract

In the last few years a new discipline has begun to emerge: machine consciousness. This talk will describe the background to this movement, and will present a line of thought showing how the problem of constructing a truly autonomous and intelligent robot may also constitute an approach to building a conscious machine. The basis of the theory is that an intelligent robot will need to simulate both itself, its environment, and its own behaviour in order to make good decisions about actions, and that the nature and operation of the internal self model may well support some consciousness-related phenomena.

As part of an investigation into machine consciousness, We are currently developing a robot that we hope will acquire and use a self-model of the right kind. We believe that this requires a robot that does not merely fit within a human envelope, but one that is anthropomimetic - with a skeleton, muscles, tendons, eyeballs, etc. - a robot that will have to control itself using motor programs qualitatively similar to those of humans. The early indications are that such robots are very different from conventional humanoids; the many degrees of freedom and the presence of active and passive elasticity do provide strikingly lifelike movement, but the control problems may not be tractable using conventional robotic methods.

The project is limited to the construction and study of a single robot, and there are no plans for the robot to have any encounters with others of its kind, or with humans. Without any social dimension to its existence, and without language, could such a robot ever achieve a consciousness intelligible to us?

Artificial Immune Systems and Immune System Modelling

4th April 2006

Organisers

James Marshall, University of Bristol
Tim Kovacs, University of Bristol
Steve Cayzer, HP Labs
Jim Smith, Univ. of the West of England

Programme Committee

Uwe Aickelin, University of Nottingham
Mick Bailey, University of Bristol
Peter Bentley, University College London
Steve Cayzer, HP Labs
Ed Clarke, University of York
Darren Flower, Edward Jenner Institute for Vaccine Research
Jungwon Kim, University College London
Tim Kovacs, University of Bristol
James Marshall, University of Bristol
Lindsay Nicholson, University of Bristol
Martin Robbins, University of Wales
Jim Smith, Univ. of the West of England
Jon Timmis, University of York
Jamie Twycross, Univ. of Nottingham

Contents

Gene Libraries: Coverage, efficiency and diversity

Steve Cayzer
HP Laboratories
Bristol BS34 8QZ
UK
steve.cayzer@hp.com

Jim Smith
University of the West of England
Bristol BS16 1QY
UK
james.smith@uwe.ac.uk

Abstract

Gene libraries are a biological mechanism for generating combinatorial diversity of antibodies. However, they also bias the antibody creation process, so that they can be viewed as a way of guiding lifetime learning mechanisms. In this paper we examine the implications of this view using two AIS performance measures: coverage and avoidance of self. We show how low numbers (in our case 2) gene libraries may drastically reduce computational expense while having negligible effect on performance. In addition to this efficiency/coverage trade-off, we also illustrate a trade-off between safety (ease of self avoidance) and diversity. The implications of these findings are discussed.

1 Introduction

In artificial immune systems, gene libraries can be used to take advantage of the fact that antigens are not uniformly distributed in non-self space. From a computational point of view, libraries introduce initialisation bias and provide a 'species memory' to tackle the antigen mapping task. What could this mean for AIS? Could gene libraries be used to intelligently seed our algorithm? In a previous paper (Cayzer et al 2005) we postulated that gene libraries might:

1. improve non-self space coverage – through better placement of detectors (antibodies), over and above random creation;
2. reduce the cost of detector generation by more effectively avoiding self;
3. map the antigen population more accurately; and
4. help deal with co-evolving antigens

In that paper, we showed that option 2 is somewhat easier to achieve than option 1. Here we extend these results, showing that there is a trade-off between self avoidance and diversity, and between encoding efficiency and coverage. Future work will concentrate on hypotheses 3 and 4, with the aim of shedding light on the sort of real world problems for which gene libraries should be considered.

2 Gene libraries for coverage

The most naïve way of looking at antibody creation is a way of covering a multidimensional area (antigen space). This is somewhat complicated by the necessity of avoiding self. We tested a number of different library configurations using 8 bit r contiguous matching on antibodies/antigens of 32 bits, as shown in table 1:

Number libraries	Segments in each library	Size of each segment	Number antibodies	Genome size
1	1089	32	1089	34848
2	33,33	16,16	1089	1056
3	11,11,9	10,11,10	1089	321
4	6,6,6,5	8,8,8,8	1080	184

Table 1: configuration of gene libraries. We kept the number of antibodies and their size (almost) constant in each case. Each row shows how we created these antibodies using a combination of gene library segments, and how we changed the segment size and number of genes per library in each case. Genome size is calculated as the sum of (#segments* size of segment) for each library.

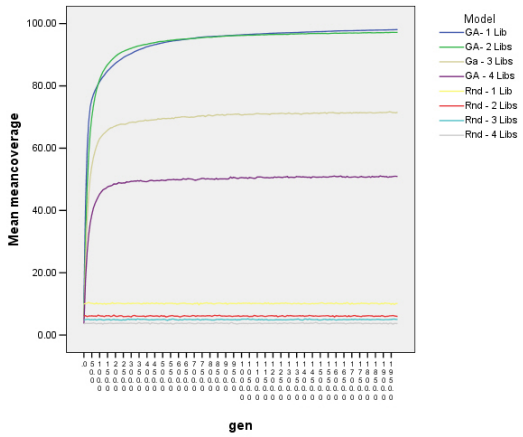


Figure 1: Mean coverage over 2000 generations (x axis). Each result shows the % antigens matched (y axis) by antibodies created from a varying number of libraries. The results using random creation are shown for comparison. Values averaged over 25 runs. Over the last 500 generations, 1 library is statistically superior (1% level, one sided t test)

Coverage was assessed using a static universe of 1024 antigens while the self set comprised 128 proteins. Figure 1 shows how the use of gene libraries comprehensively outperforms random creation on this basic task. Interestingly, although the use of 1 library gave the best result (97.8%), 2 libraries provide a comparable (97.0%) performance.

2 Gene libraries for avoiding self

In the human immune system, avoidance of self is essential to protect against autoimmune reactions. Could gene libraries provide a bias to assist negative selection; that is, make the creation process cheaper? Our previous results (Cayzer et al 2005) showed that gene libraries indeed had a profound effect on the cost of negative selection. However, subsequent analysis (figure 2) shows that this is at the expense of reducing diversity. In other words, one gets a high proportion of 'safe' (non self reactive) antibodies – but also a large number of duplicates. Clearly there is a trade-off between coverage and cost of creation.

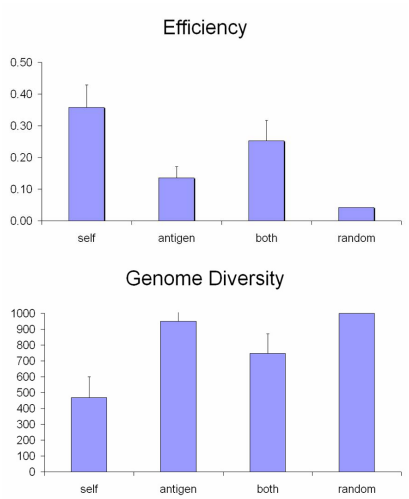


Figure 2: Effect of using avoidance of self as a fitness function (self), as opposed to coverage (antigen), combined (both) or simply using a random creation strategy. The top figure shows that AIS individuals can evolve gene libraries with a far higher (36%) chance of producing valid antibodies than one whose fitness function measures only coverage (antigens; 13%) and far above random creation (5%). All differences are statistically significant (wilcoxon). Experiments were run using 3 libraries, 16 bit strings and 6 bit r-contiguous matching. In the lower figure, the 'self' AIS individuals have roughly half the diversity of the others (unique number of antibodies; 470 cf 950 antigen, 983 random). All differences significant except antigen/random.

3 Coverage and encoding efficiency

In the above analysis, one library gave superior coverage. Yet this is at considerable cost. One library gives no combinatorial advantage at all, requiring a genome of almost 35 thousand bits (table 1). Conversely, the same number of antibodies can be created from 4 libraries using less than 200 bits. Even using 2 libraries gives a 35 fold reduction in genome length, and in view of the comparable performance this is one result we will be investigating further.

4 Mapping antigen

We are currently investigating the performance of gene libraries when faced with an inhomogeneous antigen environment. Our expectation is that depending on the nature of self and antigen space, gene libraries may prove either a boon or a burden. Our results should give guidance as to the region of problem space amenable to the use of gene libraries.

5 Conclusions

Gene libraries are clearly beneficial for introducing initialisation bias to antibody creation. Whether they are superior to state of the art algorithms, or indeed computationally tractable, is the subject of ongoing investigation.

In this paper we have shown the use of gene libraries to perform a coarse grain mapping on antigen space. Although the results are clearly superior to random creation, it seems an expensive way to achieve a simple result, even when one includes the complication of avoiding self (see Cayzer et al 2006 for fuller discussion). There are specialised algorithms (Wierzchon 2002) which may be more suitable for this task; the true advantage of gene libraries may only be evident in more complex environments. The representation and mapping operators may also make a significant difference.

We have also have shown that there are two important trade-offs: diversity vs safety; and encoding efficiency vs coverage. Our results suggest that a low number of gene libraries appears to provide considerable advantage in efficiency while not proving overly detrimental to coverage. Whether the same benefit can be demonstrated in dynamic environments is a topic for future investigations.

References

Steve Cayzer, Jim Smith, James Marshall & Tim Kovacs (2005) What have gene libraries done for AIS? Proceedings ICARIS-2005, 4th International Conference on Artificial Immune Systems, LNCS 3627, pp 86-99, Springer-Verlag, Banff, Canada.

Wierzchon, S (2002) Deriving a concise description of non-self patterns in an artificial immune system. In: New Learning Paradigms in Soft Computing Physica-Verlag (2002) 438-458

Artificial Immune Tissue using Self-Organizing Networks

Jan Feyereisl* and Uwe Aickelin*
*School of Computer Science
University Of Nottingham, Nottingham,
NG8 1BB, UK
jqf,uxa@cs.nott.ac.uk

Abstract

As introduced by Bentley et al. (2005), artificial immune systems (AIS) are lacking tissue, which is present in one form or another in all living multi-cellular organisms. Some have argued that this concept in the context of AIS brings little novelty to the already saturated field of the immune inspired computational research. This article aims to show that such a component of an AIS has the potential to bring an advantage to a data processing algorithm in terms of data pre-processing, clustering and extraction of features desired by the immune inspired system. The proposed tissue algorithm is based on self-organizing networks, such as self-organizing maps (SOM) developed by Kohonen (1996) and an analogy of the so called Toll-Like Receptors (TLR) affecting the activation function of the clusters developed by the SOM.

1 Introduction

A number of immune inspired systems have been developed over the years. From negative selection based algorithms to the self vs. non-self (Forrest et al., 1996) and the danger model (Aickelin et al., 2003). Bentley et al. (2005) argue that tissue is one missing component of AIS, as it is the first line of defence against viruses and bacteria, which possibly initiates the activity of the whole immune system.

1.1 Tissue

Tissue is any part of a multi-cellular organism, which provides an environment, that can be affected by viruses and bacteria and thus initiate an immune response. It is an intermediate layer between a problem and the actual immune system, which provides a certain interpretation of the occurring problem to the AIS in order to better protect itself.

1.2 TLRs

TLRs are a set of receptors on the surface of immune cells, such as dendritic cells, which act as sensors to foreign microbial products essential to their existence. When encountering one or more of such products, they trigger a cascade of events potentially resulting in an immune response. Different combinations of activated TLRs perform different actions.

2 SOM and Intrusion Detection

SOMs have been used as part of an IDS on a number of occasions, nevertheless their main application so far has been in the area of network packet analysis. Our proposed method looks at the use of the SOM algorithm in a number of distinctly different ways. Firstly, the SOM algorithm is only a part of an overall tissue algorithm comprising of a set of functions analogous to biologically real tissue, e.g. the notion of inflammation, TLRs, antigens, etc... Secondly, the aim of an artificial tissue is not to act as an IDS on its own, but rather as an initial pre-processing of system data. Thus it supplies the AIS with 'interesting data', making it easier, quicker and more reliable for the AIS to make a decision about a potential threat to the system. As in the human body, the artificial tissue is an environment in which the initial interactions and alarms are raised when 'something' is happening.

3 The Link

There are four main areas of the biological analogy; Tissue, cells, TLRs and inflammation. A general overview of the proposed algorithm design can be seen in Figure 1.

3.1 Artificial Tissue

Tissue is a layer between the problem and the AIS, represented in terms of a pre-processing algorithm. It

is an environment, within which malignant organisms (i.e. malicious code) invade cells in order to survive and eventually cause damage. In this way, tissue acts as an encoding and reduction layer for the incoming data into the AIS. It analyses the data based on an immunological concept and only passes the 'interesting' data to the AIS. By 'interesting', we mean data which is of potentially unknown nature to the tissue environment. Tissue can be seen as a grid of neurons within a SOM.

3.2 Artificial Cells

Tissue comprises of cells, each of which might have slightly different functionality. We can imagine an artificial cell in terms of a neuron within a self organizing map. This means, that a cell has a number of inputs, which are used to compare the incoming data to the tissue to the data that the cell holds, in order to find the most suitable cell to which to relate. Such a cell comes in contact with data that is similar to the cells' content and is eventually adjusted, as well as its neighbouring cells, according to the incoming data, based on the SOM algorithm. The outcome of this automatic cell 'growth' results in the tissue being compartmentalized according to similar types of cells, based on the correlation of the multidimensional input features incoming into the tissue. In other words, similar system behaviour is grouped together within the tissue. This results in a constantly updated map, which holds information about the normal behaviour of a system as a whole. Once an unusual action occurs, this should affect cells within the tissue, that have not been affected before or that have not been affected in such a dramatic way.

3.3 Artificial TLRs

The analogy of TLRs is based on the enhancement of the functionality of the tissue cells described above. In the immune system, TLRs sense specific predefined chemicals, which are released by malignant organisms. In a similar way, we can specify a set of potentially hazardous system features, each of which can be represented as a receptor. The TLRs will be associated with the cells within the tissue, as in real life, and based on their activation, they will affect the 'growth' of the cell in a more dramatic way. Similarly to the natural functionality of TLRs, the artificial receptors will have a different impact on the underlying cell if a combination of them are activated at the same time.

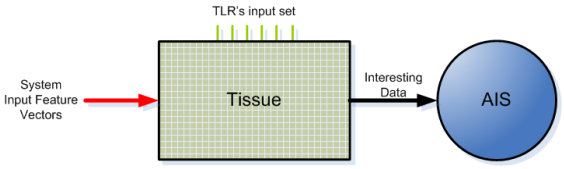


Figure 1: SOM Tissue Design, Cells represented as intersections of white lines, Inflammation as the bandwidth of the tissue I/O streams

3.4 Artificial Inflammation

Inflammation proposes the possibility of signalling where the AIS should possibly focus its attention on, or where priority is to be set, thus possibly enabling the notion of problem locality. For example as a result of a rapid cell 'growth', the system can increase or decrease the priority of an associated process. Similarly a technique described by Somayaji and Forrest (2000) can be used in order to give the AIS a better chance at making a correct decision.

Acknowledgments

This research is partially funded by the ARTIST network (EPSRC GR/S56621/01).

References

Uwe Aickelin, Peter J. Bentley, Steve Cayzer, Jungwon Kim, and Julie McLeod. Danger theory: The link between AIS and IDS? In *International Conference on Artificial Immune Systems (ICARIS), LNCS 2787*, pages 147–155. Springer-Verlag, 2003.

P. Bentley, J. Greensmith, and S. Ujjin. Two ways to grow tissue for artificial immune systems. In *International Conference on Artificial Immune Systems (ICARIS), LNCS 3627*, pages 139–152, 2005.

Stephanie Forrest, Steven A. Hofmeyr, Anil Somayaji, and Thomas A. Longstaff. A sense of self for UNIX processes. In *IEEE Symposium on Security and Privacy*, pages 120–128, Oakland, CA, 1996. IEEE Computer Society Press.

T. Kohonen. *Self-Organizing Maps*. Springer-Verlag, Berlin, 1996.

Anil Somayaji and Stephanie Forrest. Automated response using system-call delays. In *Proceedings of the Ninth USENIX Security Symposium, August 14–17, 2000, Denver, Colorado*, page 185, 2000.

Dendritic Cells for Real-Time Anomaly Detection

Julie Greensmith* and Uwe Aickelin*
*School of Computer Science, University of Nottingham, UK
jqg, uxa@cs.nott.ac.uk

Abstract

Dendritic Cells (DCs) are innate immune system cells which have the power to activate or suppress the immune system. The behaviour of human DCs is abstracted to form an algorithm suitable for anomaly detection. We test this algorithm on the real-time problem of port scan detection. Our results show a significant difference in artificial DC behaviour for an outgoing portscan when compared to behaviour for normal processes.

1 Introduction

Intrusion detection systems (IDS) are a method used in computer security for detection of unauthorised use of machines. The Danger Project proposed by Aickelin et al. (2003) aims to improve on results previously seen with artificial immune systems (AIS) by applying concepts from the Danger Theory to IDS. Danger theory proposes that exposure to danger signals or pathogenic bacteria causes the activation of the immune system, not pattern matching of antigen. The cells responsible for combining these various signals are Dendritic cells. We use the 'signals plus context' processing power of Dendritic Cells (DCs) to perform anomaly detection.

Abstraction of certain properties thought key to DC function was performed, with algorithmic details and sources of biological inspiration detailed in Greensmith et al. (2005) . The properties we abstract from DC behaviour include their existence in different states, depending on their environmental conditions. As immature DCs, they collect multiple antigens and are exposed to signals, derived from dying cells in the tissue (safe or danger signals). DCs can combine these signals with bacterial signatures (PAMPs) to generate different output concentrations of costimulatory molecules, semi-mature cytokines and mature cytokines. Exposure to signals generates an increase in co-stimulatory molecules and causes the maturation of a DC to two different states: mature and semi-mature. DCs process a multitude of signals generated by the presence of bacteria or generated by damage to the tissue. PAMPs, based on a pre-defined signature, and danger signals (released on damage to the tissue) cause an increase in mature DC cytokines. Safe signals cause an increase in semi-mature DC cytokines and have a suppressive effect on both PAMPs and danger signals.

A key feature of the DCs is an ability to combine signals with antigen. In order to provide an environment suitable for the collection of signals and antigen, we use a system developed for the Danger Project (Aickelin et al. (2003)) known as `libtissue`. Using `libtissue` we can create a tissue compartment, to house a population of DCs. This compartment, known as a tissue server, is used to update the DCs on exposure to signals and antigen. A tissue client transforms raw values into normalised signal concentrations. A weighted signal processing function (described in detail in Greensmith et al. (2005)) is used to combine these signals to determine the output signal concentration of a DC. The exposure of a DC to PAMPs, danger or safe signals causes an increase in co-stimulatory molecules (CSM) on the DC. Once the CSM value exceeds a given threshold, the DC 'matures' and is removed from the sampling population. The concentrations of mature or semi-mature cytokines expressed by the DCs are calculated. Antigen is collected during the period of signal exposure in the sampling pool. Once a DC has matured, any antigen collected is labelled with presented DC context (mature or semi-mature) for each antigen. A mean percentage mature antigen value can be calculated, indicating the number of times an antigen was presented in a mature context.

2 Port Scan Experiment

For this experiment an ICMP (ping) scan is used to provide an example of malicious behaviour. System calls for a monitored secure shell (ssh) session are captured using a tissue client. This includes normal processes such as the controlling shell, x-forwarding

agent and ssh demons. The process IDs of these system calls form the antigen. Signals are captured from different aspects of system behaviour. PAMPs are signature based and are derived from the number of ICMP errors received per second. Danger signals are derived from the number of packets per second sent by the machine. Safe signals are represented as the rate of change of packets per second, based over a 2 second moving average.

These incoming signals are used to convert DCs to either semi-mature or mature, measured by the relative concentrations of their output cytokines. For the duration of the experiment, the IDs of running processes, the output cytokines and the presented antigen are recorded. Post-hoc analysis allows us to measure the mean % mature context antigen for each process. Each experiment is repeated 10 times and an average value for the mean % mature context antigen is calculated, per process.

Three experiments are performed, using different combinations of signals and variations on the weight of the suppressive safe signal. Experiment 1 uses danger and safe signals alone, with a -1 value for the suppression by safe signals; experiment 2 uses danger and safe signals in combination with PAMPs; and finally, experiment 3 uses PAMPS, danger and safe signals, but with an increased value of suppression, a value of -2, to allow for exploration between this value and the detection of normal processes.

2.1 Results and Analysis

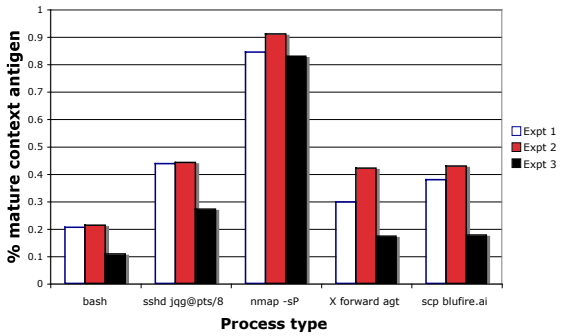


Figure 1: Experimental results, showing the % mature context antigen for each process

The five processes of interest presented in Figure 1 include: the bash shell from which the scan was performed; the ssh demon; the nmap performing the port scan; the graphical forwarding agent for the remote shell; and finally, the file transfer (scp). The detection of the nmap was significantly greater than the value derived for the normal file transfer, especially in experiment 3. The addition of signals did not significantly alter the mean % mature antigens of the nmap process. Conversely, the mean % mature antigens for the normal file transfer was significantly reduced when the safe signal weight was changed to -2 (all significance assessed through paired t-tests, with 95% confidence demonstrated).

In each experiment the nmap process generated significantly more mature context antigen than any other process. The addition of PAMPs did not significantly increase the detection of the 'anomalous' nmap but combined with a higher safe signal weight, lowered the detection of the normal processes. In future experiments, a much higher level of safe signal could be used without reducing the detection of the misbehaving process (lower rate of false positives).

3 Conclusions

In this paper we have demonstrated the use of a Dendritic cell inspired algorithm on a small-scale, real-time problem. The promising results shown in the port scan experiments imply that the DC algorithm plus `libtissue` framework can be used for the purpose of anomaly detection under real-time conditions. Future work involves the use of a replay client, so data from real-time experiments can be captured for the purpose of testing the parameters and limitations of the system in a controlled manner.

Acknowledgements

This project is supported by the EPSRC (GR/S47809/01), Hewlet Packard Labs, Bristol, and the Firestorm intrusion detection system team. `Libtissue` developed by Jamie Twycross.

References

U Aickelin, P Bentley, S Cayzer, J Kim, and J McLeod. Danger theory: The link between ais and ids. In *Proc. of the Second Internation Conference on Artificial Immune Systems (ICARIS-03)*, pages 147–155, 2003.

J Greensmith, U Aickelin, and S Cayzer. Introducing dendritic cells as a novel immune-inspired algorithm for anomaly detection. In *Proc. of the Fourth International Conference on Artificial Immune Systems (ICARIS-05)*, pages 153–167, 2005.

Alternative Representations for Artificial Immune Systems

James A. R. Marshall* Tim Kovacs*
*Department of Computer Science
University of Bristol
Woodland Road
Bristol BS8 1UB
marshall@cs.bris.ac.uk kovacs@cs.bris.ac.uk

Abstract

Artificial Immune Systems (AIS) have been proposed to solve binary classification problems; distinguishing between instances of self and of non-self. For any classification system such as AIS, the choice of classifier representation used impacts substantially on the kind of classification problem that can be handled. Despite this, most classication systems such as AIS make use of only one representation, frequently without explicit consideration of its suitability to the classification problem of interest. As many AIS make use of binary encoding, AIS are thus frequently being applied to boolean functions. One other notable field of classification system research is also applied extensively to boolean functions, namely Learning Classifier Systems (LCS). Here we consider boolean functions and the suitability of different representations for their classification. We compare a simple representation proposed for use in AIS, Hamming-distance based matching, with a traditional representation from Learning Classifier Systems (LCS), binary classifiers with wildcards. These different representations realise different shapes in a high-dimensional instance space; hyperspheres (AIS) and hyperplanes (LCS). In fact, hyperplanes are a more general case of the kind of classifiers implemented for use with the r-chunks matching rule in AIS (Balthrop et al., 2002). We consider the different characteristics of these representations, analysing how their size (number of instances covered) and instance space size (number of distinct classifiers) varies differently with both problem size (instance string length) and classifier size (hyperplane dimension or hypersphere radius). As well as these differences, we consider how hyperspheres and hyperplanes differ in the way in which they generalise. These differences are likely to mean that the traditional hypersphere-based AIS representation is of limited applicability. For example, it is likely that many boolean functions cannot be well covered by sets of general hypersphere classifiers, unless specificity of matching is used to arbitrate between multiple classifiers matching a single instance. Similarly, differences in the generalisation mechanisms of the two representations mean that increasing the dimensionality of a problem, through increasing the number of its attributes, will have different consequences according to which representation is used. As well as the usefulness *per se* of analysing differences between classifier representations, implementing a suite of alternatives may enable a classification system to learn to use the most appropriate one when faced with a particular classification problem (Marshall and Kovacs, 2006), or even to evolve new representations. While the maintenance of detectors using different representations has already been proposed for AIS, this has been achieved by random permutations on bit order of the detector strings (Hofmeyr and Forrest, 2000). Our results show, that as the number and size of classifiers is different when using hyperplane or hypersphere representations (Marshall and Kovacs, 2006), a simple permutation on the detector string is insufficient to map between the different representations. The different representations have different informational content, and hence offer genuine differences to each other. We thus propose going beyond simple permutation in maintaining diversity of representations in an AIS.

References

Balthrop, J., Esponda, F., Forrest, S. and Glickman, M. (2002) Coverage and Generalization in an Artificial Immune System. GECCO 2002.

Hofmeyr, S. A., Forrest, F. (2000) Architecture for an Artificial Immune System. Evolutionary Computation 8, 443-473.

Marshall, J. A. R. and Kovacs, T. (2006) A representational ecology for learning classifier systems. Under submission to GECCO 2006.

Stress, Resource Allocation and Mortality

John McNamara[*]
[*]Department of Mathematics
University of Bristol
Bristol BS8 1TW
John.McNamara@bristol.ac.uk

Kate Buchanan[†]
[†]School of Biosciences
Cardiff University
Cardiff CF10 3TL
BuchananKL1@cf.ac.uk

Abstract

We model the optimal allocation of limited resources of an animal during a transient stressful event such as a cold spell or the presence of a predator. The animal allocates resources between the competing demands of combating the stressor and bodily maintenance (e.g. maintaining immune function). Increased allocation to combating the stressor decreases the mortality rate from the stressor, but if too few resources are allocated to maintenance, damage builds up. A second source of mortality (disease in the case of reduced allocation to immune function) is associated with high levels of damage. Thus, the animal faces a trade-off between the immediate risk of mortality from the stressor and the risk of delayed mortality due to the build up of damage. We analyse how the optimal allocation of the animal depends on the mean and predictability of the length of the stressful period, the level of danger of the stressor for a given level of allocation, and the mortality consequences of damage. We also analyse the resultant levels of mortality from the stressor, from damage during the stressful event, and from damage during recovery after the stressful event ceases. Our results highlight circumstances in which most mortality occurs after the removal of the stressor. Results also highlight the importance of the predictability of the duration of the stressor and the potential importance of small detrimental drops in condition. Surprisingly, making the consequences of damage accumulation less dangerous can lead to a reallocation that allows damage to build up by so much that the level of mortality caused by damage build up is increased. Similarly, because of the dependence of allocation on the dangerousness of the stressor, making the stressor more dangerous for a given level of allocation can decrease the proportion of mortality that it causes, whilst the proportion of mortality caused by damage to condition increases. These results are discussed in relation to biological phenomena.

References

McNamara, J. M. & Buchanan, K. (2005) Stress, resource allocation and mortality. Behavioral Ecology 16(6):1008-1017

Simulation of the Immune System to Investigate Rare Outcomes of Infection

David Nicholson [*]
[*] Computational, theoretical and structural group
Dept. of Chemistry, Imperial College, London
d.nicholson@imperial.ac.uk

Lindsay B. Nicholson[†]
[†]Cellular and Molecular Medicine
University of Bristol
l.nicholson@bristol.ac.uk

Abstract

We report here a simple simulation of the immune system in which we analyse the behaviour of responder cells in the presence of target cells. If target cells are recognised, the responder cells divide and become effector cells with the capacity to kill targets. Variable parameters determine the behaviour of the cells within the simulation, and many simulations using the same parameters ensure that statistical variability is achieved. When the number density of responding cells is increased, target cells are cleared more efficiently, but there is an increased likelihood of a prolonged response. This leads to a significant increase in the probability of persistence in circumstances in which the immune response fails to clear infection during the first period of clonal expansion, which suggests that pathogen independent mechanisms might contribute to the development of chronic infection.

1 Introduction

The mammalian immune system is comprised of many parts. In the face of external stimulation, these interact together in a response which is often effective in eliminating infection. Defining the immune response is a challenge that has been addressed in many ways, and we have a detailed understanding of many of the mechanisms that regulate immunity down to the level of molecular structure and interaction. But complex systems may be more than the sum of their parts, and investigating the collective behaviour of an immune response may yield additional insights to those obtained from the study of its components.

One way to connect constituent elements to the whole is to build models. By extracting essential elements and putting them together in a system whose behaviour can be controlled and analysed, we may be able to make a quantitative analysis of the relationship between the components and the system as a whole. Model building has played an important role in our understanding of the dynamics of infection. The most widely used models describe elements of the system in terms of partial differential equations that express the evolution in time of various components in the system. This approach is very successful at describing global properties, but it is not able to address effects dependent on local non-equilibrium aspects such as co-operation and clustering, or describe rare behaviours arising due to stochastic variation. We have designed a simple simulation model in which spatial distribution forms a central feature and have used it to model responses to a pathogen.

2 Results

The model we have chosen simulates the behaviour of responder cells in the presence of target cells. If the target cells are recognised, the responder cells expand and become effector cells with the capacity to kill targets. Effector cells proceed through a fixed number of divisions and then die unless they encounter further stimulation because of the continued presence of target cells. Variable parameters are defined in an input file and determine the behaviour of the cells within the simulation. A random number

seed, which generates a sequence of pseudo random numbers, is used for probability testing. The seed is set prior to each run from a predetermined sequence and averages over many runs (typically 100) ensure that statistical variability has been achieved.

In this model the results of all our simulations are defined by a bounded environment. When the simulation is complete, we observe three possible outcomes. Target cells may have been eliminated, they may have filled the system, or they may be present at intermediate levels. In the model we assume that an optimal immune response is one that produces clearance. We find that even under parameter sets where there is a high probability of clearance, per sistence beyond the duration of the simulation is observed. When we test what factors are associated with persistence we find that low peak proliferation, but high total proliferation which leads to higher average number densities, plays a significant role. The number density of responder cells also influences the probability of persistence. Based on these observations, we propose that chronic immune responses may be fundamentally different from acute responses in terms of their population biology, and suggest that pathogen independent mechanisms might contribute to the development of chronic infection. These results may also have implications for understanding the development of chronic autoimmunity.

A hybridized AIS for Anomaly Detection: Combining Negative Selection and Association Rules (Extended Abstract)

Arlene Ong and David Clark*
*King's College London
Department of Computer Science
Strand, London WC2R 5LS
{ong,david}@dcs.kcl.ac.uk

Jungwon Kim†
†Department of Computer Science
University College London
Malet Place, London WC1E 6BT, UK
J.Kim@cs.ucl.ac.uk

Abstract

This paper proposes a hybrid, supervised, anomaly detection system using association rules and a variant of the negative selection algorithm (NSA). The aim of this hybridization is two-fold: first, to provide more comprehensible detection results and second, to improve the quality of the NSA detector generation process so as to obtain more accurate detection results. We have applied the algorithm to both the UCI and the Kdd '99 cup data sets. Our experimental results show that the algorithm can obtain high detection rates although with varying false positive rates. The detection results, as expected, demonstrated very good comprehensibility.

1 Introduction

Anomaly detection systems (ADSs) are popular in many application domains including fraud detection and intrusion detection. In general, ADSs can be classified into supervised and unsupervised. A supervised ADS relies on the availability of a clean system/data log for training. There are two steps to a supervised ADS. The first is to define and select a representation of self, and the second is to tag deviations from self as anomalies. Regardless of the application domain and the type of ADS employed, a successful ADS should be accurate and comprehensible. In this paper, we propose a hybrid supervised ADS with these properties by combining the negative selection algorithm (NSA) and association rule mining (ARM).

Our motivation for building the hybrid ADS is twofold. First, many current techniques such as neural networks and statistical measures rarely provide good, comprehensible, detection results. In most ADSs, users still have to review and act upon the anomalies detected. Having more comprehensible detection results will greatly improve user's understanding and speed up any action required. We believe that an association rule based system provides better comprehension because it is well known that rules are more easily understood by a user than raw transactions are. Second, there have been many criticisms on the NSA Aickelin et al. (2004) Stibor et al. (2005), questioning the NSA's suitability for anomaly detection. We believe that part of the weakness of the NSA lies in the initial definition and representation of self.

Since a supervised ADS needs to identify and represent the self present in a system/data log, selecting the right representation for the "normal" pattern (self) is crucial. Maxion and Tan (2000) state "If detector performance is indeed a function of environment regularity, it would be critical to match detectors to environmental characteristics". Currently, the NSA does not do more than just use the raw, self data as this "normal" pattern and we believe that the provision of some structure in the definition of self should assist in targeting detectors of anomalies more accurately and ultimately lead better detection rates. We argue that association rules are a good representation of self, since it shows the regular behaviour within self data and that it should be used to guide the detector generation process.

Our hybrid ADS first transforms the training data into a set of association rules. Then candidate detectors are generated in the form of itemsets (potential anomalous signatures) that detect anomalous individual transactions.

2 Algorithm Description

There are three main components in our hybrid supervised ADS. The first component is the self profile builder. We define our self profile as a set of association rules, generated using a Apriori algorithm (Agrawal et al., 1996). The second component of the

system is the candidate detector generation. A candidate detector is an itemset that combines the antecedent attribute-value pairs of an association rule and the contradicting consequent attribute-value to the same association rule. Then, candidate detectors are compared to the self data in the process of filtering candidate detectors. If a candidate detector matches any self data instance, it is eliminated. The remaining candidate detectors become the final detectors representing potential anomalies. For example, {venomous=no, milk=**no**} is a detector generated from the association rule, {venomous=no → milk=yes}. The final component of the system is the detection component of the system. The task of this component is to apply the final detectors to the test data. A user pre-defined threshold level is used for raising alarm. Basically, if a data instance "matches" more than a threshold of detectors, an alarm is raised. Also, if a detector "matches" more than a threshold of data instances, an alarm will also be raised. The data instances detected along with the detectors and the paired association rule are presented to a user for judgment.

3 Discussion and Results

We have redefined the boundaries of self with the introduction of association rules and in doing so we have split the space of anomalies into two; interesting anomalies (non-self that matches detectors) and non-interesting anomalies (non-self that are not detected by detectors). More specifically, interesting anomalies are data instances that indicate patterns that are infrequent and contradict frequent occurring highly correlated rules. For example, we applied the algorithm to the zoo data set from the UCI repository (Blake and Merz, 1998) where the class *mammals* is treated as self. One of the successful detectors generated is {fins=no,hair=yes,venomous=yes}. On it's own, it is difficult to comprehend why the detector represents anomalous behaviour. However, if we look at the paired association rule, {fins=no,hair=yes→venomous=no}, we note that the specific contradiction lies with the attribute *venomous*. Interestingly, the detector demonstrates a specific behaviour of the "insect" subclass of the anomalous class.

By introducing the creation of a self profile into the algorithm, we effectively increase the storage space required. It is our conjecture that, the self profile aside, our hybrid algorithm is more computationally efficient than the original NSA because it needs less iterations to generate the candidate detectors. On the other hand, clearly, there is a cost involved in applying Apriori algorithm to generate the self profile in the first place.

We also ran our algorithm on the larger Kdd' 99 cup (Lee et al., 1999) data set to test the accuracy of detection. First, we randomly split the normal data into 10 sets, using a single set as self for training and 9 other set for testing along with the anomalous data. In our best results, we obtained a detection rate of approximately 95% with 1% false positive rate. In the worse case, for the same level of detection rate, we obtained a false positive rate of 38%. The volatility of the false positive rate is far from ideal. However we argue that this could be due to our over optimism of the amount of training data required and that further experiments are required in which the amount of training data is increased. In addition, the system currently discussed is a static system, however we have seen signs of suitability for further extending the system to ensure continuous learning and adaptability of the the association rules and detectors to cope with changing data behaviour.

References

R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, and A. I. Verkamo. Fast discovery of association rules. In *Advances in Knowledge Discovery and Data Mining*, 1996.

Uwe Aickelin, Julie Greensmith, and Jamie Twycross. Immune system approaches to intrusion detection - a review. In *ICARIS*, pages 316–329, 2004.

C. L. Blake and C. J. Merz. UCI repository of machine learning databases, 1998.

Wenke Lee, Salvatore J. Stolfo, and Kui W. Mok. A data mining framework for building intrusion detection models. In *IEEE Symposium on Security and Privacy*, pages 120–132, 1999.

Roy A. Maxion and Kymie M. C. Tan. Benchmarking anomaly-based detection systems. In *First International Conference on Dependable Systems and Networks*, NY, USA, 2000.

T. Stibor, P. Mohr, J. Timmis, and C. Eckert. Is negative selection appropriate for anomaly detection? pages 321–328, June 2005.

An Immune Inspired Network Intrusion Detection System Utilising Correlation Context

Gianni Tedesco* Uwe Aickelin*
*School of Computer Science & IT (ASAP)
University of Nottingham
NG8 1BB
gxt@cs.nott.ac.uk

Abstract

Network Intrusion Detection Systems (NIDS) are computer systems which monitor a network with the aim of discerning malicious from benign activity on that network. While a wide range of approaches have met varying levels of success, most IDSs rely on having access to a database of known attack signatures which are written by security experts. Nowadays, in order to solve problems with false positive alerts, correlation algorithms are used to add additional structure to sequences of IDS alerts. However, such techniques are of no help in discovering novel attacks or variations of known attacks, something the human immune system (HIS) is capable of doing in its own specialised domain. This paper presents a novel immune algorithm for application to the IDS problem. The goal is to discover packets containing novel variations of attacks covered by an existing signature base.

1 Introduction

Network intrusion detection systems (NIDS) are usually based on a fairly low level model of network traffic. While this is good for performance it tends to produce results which make sense on a similarly low level which means that a fairly sophisticated knowledge of both networking technology and infiltration techniques is required to understand them.

Intrusion alert correlation systems attempt to solve this problem by post-processing the alert stream from one or many intrusion detection sensors (perhaps even heterogeneous ones). The aim is to augment the somewhat one-dimensional alert stream with additional structure. Such structural information clusters alerts in to scenarios sequences of low level alerts corresponding to a single logical threat.

A common model for intrusion alert correlation algorithms is that of the attack graph. Attack graphas are directed acyclic graphs (DAGs) that attempt to represent the various types of alerts in terms of their prerequisites and consequences. Typically an attack graph is created by an expert from a priori information about attacks. The attack graph enables a correlation component to link a given alert with a previous alert by tracking back to find alerts whose consequences imply the current alerts prerequisites. Another feature is that if the correlation algorithm is run in reverse, predictions of future attacks can be obtained.

In implementing basic correlation algorithms using attack graphs, it was discovered that the output could be poor when the underlying IDS produced false negative alerts. This could cause scenarios to be split apart as evidence suggestive of a link between two scenarios is missing. This problem has been addressed in various systems by adding the ability to hypothesise the existence of the missing alerts in certain cases. Ning et al (2004) go as far as to use out of band data from a raw audit log of network traffic to help confirm or deny such hypotheses.

While the meaning of correlated alerts and predicted alerts is clear, hypothesisd results are less easy to interpret. Presence of hypothesised alerts could mean more than just losing an alert, it could mean either of:

1. The IDS missed the alert due to some noise, packet loss, or other low level sensor problem
2. The IDS missed the alert because a novel variation of a known attack was used
3. The IDS missed the alert, because something not covered by the attack graph happened (totally new exploit, or new combination of known exploits)

This work is motivated specifically by the problem of finding novel variations of attacks. In our case a

variation is determined to be an attack which exploits the same vector as an attack detected by an existing rule. The basic approach is to apply AIS techniques to detect packets which contain such variations. A correlation algorithm is taken advantage of to provide additional safe/dangerous context signals to the AIS which would enable it to decide which packets to examine. The work aims to integrate a novel AIS component with existing intrusion detection and alert correlation systems in order to gain additional detection capability.

2 Intrusion Alert Correlation

Although the exact implementation details of attack graphs algorithms vary, the basic correlation algorithm takes an alert and an output graph, and modifies the graph by addition of vertices and/or edges to produce an updated output graph reflecting the current state of the monitored network system.

For the purposes of discussion, an idealised form of correlation output will be defined which hides specific details of the correlation algorithm from the AIS component. This model, while fairly simple, adequately maps to current state of the art correlation algorithms. Due to space constraints we do not describe the full model here.

3 Danger Theory

The advent of Polly Matzingers Danger theory in has inspired a great deal of research in to the functioning of the innate immune system. A subsystem of the human immune system (HIS) which is apparently able to distinguish between benign and pathogenic material within the organism and initiate an adaptive immune response.

For this purpose our "libtissue" AIS framework, a product of a danger theory project (Aickelin et al, 2003), will model a number of innate immune system components such as dendritic cells in order to direct an adaptive T-Cell based response.

Dendritic cells (henceforth DCs) are of a class of cells in the immune system known as antigen presenting cells. They differ from other cells in this class in that this is their sole discernable function. As well as being able to absorb and present antigenic material DCs are also well adapted to detecting a set of endogenous and exogenous signals. These biological signals are abstracted in our system under the following designations:

1. Safe: Indicates a safe context for developing tolleration
2. Danger: Indicates a change in behaviour that could be considered pathological
3. Pathogen Associated Molecular Pattern (PAMP): Known to be dangerous

All of these environmental circumstances, or inputs, are factors in the life cycle of the DC. A sufficient concentration of signals may trigger maturation along one of two differented pathways. One of which is associated with a reactive and the other with a tolerogenic T-cell response.

In the proposed system, DCs are seen as living among the IDS environment. This is achieved by wiring up their environmental inputs to certain changes in the IDS output state. Populations of DCs are tied to the prediction vertices in the correlation graph, one DC for each predicted attack. Packets matching the prediction criteria of such a vertex are injested by the corresponding DC.

A prediction veretex can either be upgraded to an exploit vertex, changed to a hypothesised vertex, or be deleted depending on subsequent alerts. These possibilities will result in either a PAMP, danger or safe signal respectively.

These signals initiate maturation and consequent migration of the DC to a virtual lymph node where they are exposed to a population of T-cells generated using the IDSs signature base in much the same way as in a gene library. This is combined with partial matching algorithms to find a T-cells to bind to the antigen being presented by the DC.

Upon successful binding, the original packet corresponding to the culprit antigen is tagged and logged much like a normal alert.

References

U Aickelin, P Bently, S Cayzer, J Kim and J McLeod. "Danger Theory: The Link between AIS and IDS?" 2nd International Conference on Artificial Immune Systems. 2003. 4th International Conference on Articial Immune Systems, 2005.

Peng Ning, Dingbang Xu. Christpher G. Healey and Robert St. Amant. "Building Attack Scenarios through Integration of Complementary Alert Methods" Proceedings of the 11th Annual Network and Distributed System Security Symposium. 2004.

Experimenting with innate immunity

Jamie Twycross (jptcs.nott.ac.uk) and Uwe Aickelin (uxacs.nott.ac.uk)

Abstract

`libtissue` is a software system for implementing and testing AIS algorithms on real-world computer security problems. AIS algorithms are implemented as a collection of cells, antigen and signals interacting within a tissue compartment. Input data to the tissue comes in the form of realtime events generated by sensors monitoring a system under surveillance, and cells are actively able to affect the monitored system through response mechanisms. `libtissue` is being used by researchers on a project at the University of Nottingham to explore the application of a range of immune-inspired algorithms to problems in intrusion detection. This talk describes the architecture and design of *`libtissue`*, along with the implementation of a simple algorithm and its application to a computer security problem.

1 Introduction

One of the achievements of immunology over the last decade has been the uncovering of the innate immune system as of central important both as the initiator and the director of immune system processes Germain (2004). Artificial immune systems are beginning to take inspiration from this work and attempt to model some aspects of innate immunity. In Twycross and Aickelin (2005), the authors presented a conceptual framework for innate immunity. The framework highlighted a number of key general properties observed in the biological innate and adaptive immune systems, and discussed how such properties might be instantiated in artificial systems. The next logical step was to take these ideas and build a software system with which systems with these properties could be experimentally evaluated. This talks reports the progress made in taking that step.

2 Innate immunity

The authors have discussed innate immunity from a biological perspective in detail in Twycross and Aickelin (2005) and it is only briefly reviewed here. Cells are the principal actors in the immune system. Many immune system cells have access to their environment on two levels: the level of antigen and the level of signals. Antigen are the markers by which the immune system senses the structure of its environment. The structure is tightly coupled to the context of the environment, which is reflected by levels of signals. Perhaps it is too strong to say that a different structure always implies a different function, since there is almost certainly some duplication of function, but generally the immune system seems to follow this principle. Signals reflect what entities are doing on a higher level than antigen, which reflect what entities are doing on a structural level.

Almost all immune processes involve the interaction of groups of different types of cell. The type of a cell is really a label for its phenotypical and functional characteristics. The following characteristics were chosen as initial areas of experimental study: antigen processing, signal processing, cell binding, antigen matching and antigen response. The reasons for choosing these have already been discussed at detail in Twycross and Aickelin (2005).

3 Implementing innate immunity

`libtissue` is a software system which allows researchers to model and experiment with novel AIS algorithms and to apply them to realtime computer security problems. It is specifically designed to explore the characteristics of innate immunity described in the previous section. An AIS algorithm is implemented as a collection of cells, antigen and signals interacting within a tissue compartment. While designed for computer security problems in the first case, its design has been kept general with a view to applying it to realtime problems from other domains.

`libtissue` has a client/server architecture pictured in Figure 1. The AIS algorithm is implemented as a *`libtissue` server*, while *`libtissue` clients* provide input data to the algorithm and provide response mechanisms. This client/server architecture separates data collection by the *`libtissue`* clients

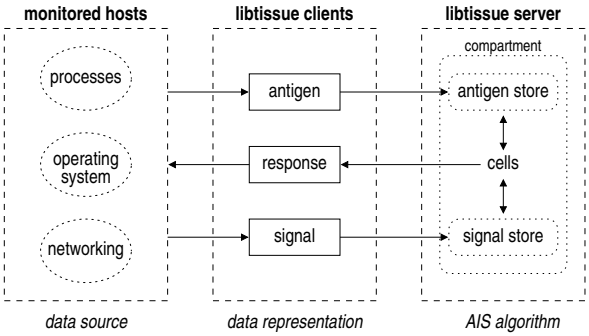


Figure 1: The architecture of *libtissue*. Hosts are monitored by *libtissue* antigen and signal clients, which in turn provide input data to the AIS algorithm, implemented as a *libtissue* server. Algorithms are able to change the state of the monitored hosts through response clients.

from data processing by the *libtissue* servers and allows for relatively easy extensibility of the existing system to new data sources. Client and server APIs exist, allowing new antigen and signal sources to be easily added to *libtissue* servers, and the testing of the same algorithm with a number of different data sources. Client/server communication is socket-based, allowing clients and servers to potentially run on separate machines, for example a signal client may in fact be a remote network monitor.

4 An example algorithm

A relatively simple AIS algorithm was implemented to validate *libtissue* and to illustrate how *libtissue* can be used to explore the behaviour of an artificial system on a realworld problem. This example has cells of two types, labelled type 1 and 2, and is shown in Figure 2. Type 1 cells are designed to emulate two key characteristics of biological APC cells: antigen and signal processing. In order to process antigen, each type 1 cell is equipped with a number of antigen receptors and producers. A cytokine receptor allows type 1 cells to respond to the value of an external signal. Type 2 cells emulate three of the characteristics of biological T cells: cellular binding, antigen matching, and response to antigen. To accomplish this, each type 2 cell has a number of cell receptors specific for type 1 cells, VR receptors to match antigen, and a response producer which is triggered when antigen is matched.

A tissue compartment is created and populated with a number of type 1 and 2 cells. The tissue compartment also stores antigen and signals received

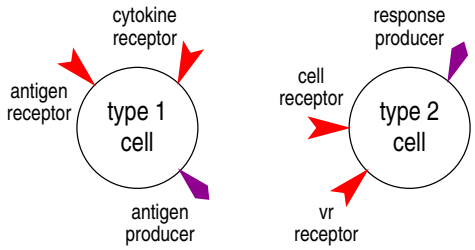


Figure 2: An example two-cell *libtissue* algorithm.

from *libtissue* clients, which provides the input data to the system. Type 1 cells ingest antigen through their antigen receptors and present it on their antigen producers. The period for which the antigen is presented is determined by a signal read by a cytokine receptor on these cells. Type 2 cells attempt to bind with type 1 cells via their cell receptors. If bound, VR receptors on these cells interact with antigen producers on the bound type 1 cell. If an exact match between a VR receptor lock and antigen producer key occurs, the response producer on type 2 cells produces a response.

5 Results

A number of experiments were carried out with the example algorithm on a realistic computer security problem, that of detecting anomalous process behaviour. The aim of these experiments was to validate *libtissue* and to highlight the methodology employed when attempting to understand the behaviour of algorithms implemented with *libtissue*. These experiments produced many interesting results, which will be presented in this talk.

Acknowledgments

This research is supported by the EPSRC (GR/S47809/01).

References

R N Germain. An innately interesting decade of research in immunology. *Nature Medicine*, 10(12): 1307–1320, 2004.

Jamie Twycross and Uwe Aickelin. Towards a conceptual framework for innate immunity. In *Proc. of the 4th International Conference on Artificial Immune Systems*, Banff, Canada, 2005.

Oil Price Trackers Inspired by Immune Memory

William Wilson*, Phil Birkin*, and Uwe Aickelin*
*School of Computer Science, University of Nottingham, UK
wow, pab, uxa@cs.nott.ac.uk

Abstract

We outline initial concepts for an immune inspired algorithm to evaluate and predict oil price time series data. The proposed solution evolves a short term pool of trackers dynamically, with each member attempting to map trends and anticipate future price movements. Successful trackers feed into a long term memory pool that can generalise across repeating trend patterns. The resulting sequence of trackers, ordered in time, can be used as a forecasting tool. Examination of the pool of evolving trackers also provides valuable insight into the properties of the crude oil market.

1 Introduction

The investigation of time series data to predict future information is a well studied area of research. This paper proposes an immune inspired solution to this problem. Inspiration for memory development was taken from the biological theory proposed by Dr Eric Bell. The theory indicates the existence of two separately identifiable memory populations (Bell, 2005), one long term and the other short term. Their differing characteristics make them ideal in recognising long and short term trends prevalent in time series data. These trends can then be sequenced for use in forecasting and prediction.

2 Development of long and short term memory

The flexible learning approach offered by the immune system is attractive as an inspiration for problem solving. However without an adequate memory mechanism the knowledge gained from the learning process would be lost. Memory therefore represents a key contributing factor in the success of the immune system. A difficulty arises in extracting immune memory properties however, because very little is still known about all the biological mechanisms underpinning memory development (Wilson and Garrett, 2004). Theories such as antigen persistence and long lived memory cells (Perelson and Weisbuch, 1997), idiotypic networks, and homeostatic turnover of memory cells (Yates and Callard, 2001) have all attempted to explain the development and maintenance of immune memory but all have been contested.

The attraction of the immune memory theory proposed by Dr Eric Bell is that it provides a simple, clear and logical explanation of memory cell development (Bell, 2005). This theory highlights the evolution of two separate memory pools. The first is a short term memory pool containing short lived, highly proliferative, activated cells that have experienced an antigen. The purpose of this pool is to drive the affinity maturation process to cope with the huge diversity of potential antigen mutations. The second pool consists of those short term memory cells that have evolved to homeostatically turnover to sustain knowledge of an antigen experience over the long term. This long term pool identifies and maintains knowledge of more generalised antigen trends.

3 Analysis of oil price trends

The price of WTI crude oil (a world marker price for oil price movements) was selected as the time series for investigation. This data set was chosen because there is considerable economic, financial and government interest in investigating oil price forecasting due to its influence on so many other market sectors. In addition, oil prices have historically exhibited a number of short and long term trend patterns which could map to our long and short term memory concepts, providing an ideal case study for this analysis.

4 An immune inspired forecasting solution

The proposed solution comprises a population of ”trackers” that correspond to B cells from the immune system. The trackers attempt to identify and

record trends in the oil price data. Price data, as measured by the change in price from one time period to the next, is encapsulated within an artificial antigen object and presented to the population of trackers. The antigens are constructed to show current and historical price movements over a particular period. In order to recognise price trends over time, each tracker is allocated a random length ”review period”. This allows the tracker population to identify a variety of potential price movements over a range of time intervals.

Following the traditional clonal selection approach (de Castro and Von Zuben, 2002), trackers attempt to bind to antigens, and undergo proliferation if successful. The resulting clones mutate in relation to the strength of the bind, with mutation taking one of three forms. One subset of clones has a random price value within their review periods mutated from its original value. A second subset has their review period extended by the addition of a randomly generated price movement to anticipate future potential price movements. A third subset of clones has a random price value removed from their review period to allow them to attempt a better fit to previously experienced antigens.

The degree proliferation is proportional to the strength of the bind and the length of the bound tracker. Initially trackers have relatively short review periods, to enable them to assess a wide variety of price trends. If successful, trackers proliferate and the review periods lengthen to anticipate additional price movements. Excessively long tracker review periods are prevented because trackers become more specific as they lengthen and are therefore less likely to bind. Without successful binds these trackers are likely to be removed via apoptosis. This leads to the evolution of a dynamic population of trackers.

The population of proliferating trackers can be seen to represent the short term memory of experienced price data, as knowledge of an identified price trend is carried forward through the generations of tracker clones. Interrogating the composition of this memory pool provides valuable insight into the dynamics of the oil market.

The process of filtering the short term memory pool to a long term memory subset is achieved through development of the ”‘tracker sequence”’. The tracker sequence is a list of trackers, ordered in time, that best represents the data presented up to the current point in time. Dominant tracker candidates, based on their degree of proliferation, are selected from the short term memory pool and transferred to the tracker sequence for use as a source of long term memory. Generalisations can be made in the tracker sequence for repeating patterns of trackers to highlight recurring price trends. The tracker sequence provides the forecasting mechanism in the system. When new price data becomes available the tracker sequence is examined to identify whether a previously identified trend is recurring again.

5 Conclusion

Inspiration was taken from the principles of memory within the immune system to build a system that would identify trends within an oil price time series. This data showed evidence of short term price fluctuations as well as exhibiting underlying long term trends. Detailed inspiration was taken from the theory of immune memory proposed by Dr Eric Bell which identifies two forms of memory, short term and long term. We indicate that these could in principle provide a mechanism to identify and map the short and long term trends evident in the crude oil market which could then be used for forecasting.

Acknowledgements

The authors would like to thank Dr Eric Bell from the University of Manchester for his valuable input.

References

E. Bell. University of Manchester, personal communication, 2005.

L. N. de Castro and F. J. Von Zuben. Learning and optimization using the clonal selection principle. *IEEE Transactions on Evolutionary Computation*, 6(3):239–251, 2002.

A. S. Perelson and G. Weisbuch. Immunology for physicists. *Rev. Modern Phys.*, 69:1219–1267, 1997.

W. Wilson and S. Garrett. Modelling immune memory for prediction and computation. In *3rd International Conference in Artificial Immune Systems (ICARIS-2004)*, pages 386–399, Catania, Sicily, Italy, September 2004.

A. Yates and R. Callard. Cell death and the maintenance of immunological memory. *Discrete and Continuous Dynamical Systems*, 1:43–59, 2001.

Associative Learning and Reinforcement Learning

3rd April 2006

Organisers

Eduardo Alonso, City University
Charlotte Bonardi, Univ. of Nottingham
Esther Mondragón, UCL

Programme Committee

Peter Dayan, University College London
Magnus Enquist, Stockholm University
Geoffrey Hall, University of York
Rob Honey, Cardiff University
Robin Murphy, UCL
Ulrich Nehmzow, University of Essex
Yael Niv, University College London
John Pearce, Cardiff University
Jose Prados, University of Leicester
Richard Sutton, University of Alberta

Contents

Pavlovian and Instrumental Q-learning: A Rescorla-Wagner-based approach to generalization in Q-learning

Eduardo Alonso
Department of Computing, City University
eduardo@soi.city.ac.uk

Esther Mondragón
Department of Psychology, University College London
e.mondragon@ucl.ac.uk

Niclas Kjäll-Ohlsson
Structurum Consulting
niclasko@gmail.com

1 Introduction

Traditionally, the Reinforcement Learning (RL) problem is presented as follows: An agent exists in an environment described by some set of possible states S, where it can perform any set of actions A. Each time it performs an action $a_t \in A$ in some state $s_t \in S$ the agent received a real-valued reward r_t that indicates the immediate value of this state-action transition. This produces a sequence of states, actions, and immediate rewards. The agent's task is to learn a control policy, $\pi : S \rightarrow A$, that maximizes the expected sum of rewards, typically with future rewards discounted exponentially by their delay. Unlike supervised learning, the learner is not told which actions to take, but instead must discover which actions yield the most reward by exploiting and exploring their relationship with the environment. Besides, actions may affect not only the immediate reward but also the next situation and, through that, all subsequent rewards. These two characteristics, trial and error search and delayed reward, are the two major features of RL.

Temporal-Difference (TD) is a set of RL methods that combine Monte Carlo ideas and Dynamic programming in that they can learn directly from row experience using simple sample models and, at the same time, bootstrap (*i.e.*, they updates estimates on other learned estimates, without waiting for a final outcome). We focus on the simplest TD technique, Q-learning (Watkins:1989), where the optimal expected long-term return is locally and immediately available for each state-action pair. A one-step-ahead search yields the long-term optimal actions without having to know anything about possible successor states and their values (according to the learning rule $Q(s,a) \leftarrow Q(s,a) + a[r + \gamma max_{a'} Q(s',a') - Q(s,a)]$, where $Q(s,a)$ is the value of the state-action pair, a is a constant learning rate, r is the reward and γ is a discount factor). Q-learning has been shown to converge with probability 1 to the optimal policy Q^*.

Regardless of its popularity in the machine learning community one single difficulty has so far prohibited the wide application of Q-learning and RL, that it does not allow for the "transfer" of learning between different yet similar situations. Despite the appar-

ent success of systems that have incorporated function approximation algorithms (*e.g.*, backpropagation in (Tesauro:1995)), for most practical tasks RL fails to generalise and, consequently, cannot be applied to large-size problems.

2 Rescorla-Wagner and Q-Learning

We present a associative learning based approach to deal with generalization in Q-learning: The Pavlovian and Instrumental Q-Learning (PIQL) framework. Our main insight is that the fundamental problem behind the lack of generalization results in Q-learning is the way RL represents the environment and its effects on the agent's behaviour. In RL an agent does not know anything about the environment or itself. Rewards are dictated by the environment, not part of the environment: Rewards are defined separated from outcomes. Consequently, the agent does not have knowledge about the motivational value of the outcomes and only estate-action values (no real goal-directed behaviours) are learned. In addition, learning is completely depended on the reward structure: If the reward changes, a new policy has to be relearned.

These assumptions are partly inherited from an early and insufficient psychological theory, Thorndike's Law of Effect, according to which animals form habits (stimulus-response associations) by learning by trial and error (Thorndike:1911). In PIQL, we propose to replace Thorndike's model with Rescorla and Wagner's model (RW, (Rescorla and Wagner:1972)). RW can be summarised in the famous learning equation $V\Delta CS = \alpha\beta[\lambda - \Sigma VCS_i]$ where α and β are constant parameters representing the Conditioned Stimulus (CS) and the Unconditioned Stimulus (US, the outcome O) respectively, and λ is the (US depending) asymptotic learning value. An operant interpretation of the equation simply replaces the CS for responses (R). Certainly, RW may find difficulties in explaining some phenomena, but still it is considered the most powerful model of associative learning (see (Miller *et al.*:1995) for an assessment of the model).

It can be argued that the main difference between RW and RL is that in RW each stimulus consists of a set of features that can be shared across stimuli. Hence, unlike RL, RW can deal with generalization. Nevertheless, we think that there are more profound differences between these two models [1], namely:

- RW does not make any distinction between rewards (returns) and outcomes. They are one and the same thing, stimuli with appetitive or aversive qualities;
- Learning does not refer exclusively to the formation of S-R associations but also to S-O (or S-S) and R-O associations;
- As outcomes are part of the association, expectance of a particular outcome can be explicitly anticipated and instrumental learning explained;

[1] It is important to distinguish between the psychological TD theory of reinforcement (Sutton and Barto:1990) and its application to the RL problem (Sutton and Barto:1998). The former is a real-time extension of RW's equation where the reinforcer's value decays with time. The RW generalization mechanism is left intact and the two models (TD and RW) are thus, in this respect, equivalent. However, when applied to the RL problem, TD (and Q-learning) is ripped off of its original psychological framework to comply with Markov Decision Processes upon which RL is mathematically formulated. Hence the differences between time-derivative models of RL and RW's model.

- S-O and R-O associations predict the actual outcome, and due to the agents internal drives and motivational state, the agent itself is able to place a value on the outcome.

On the other hand, there are striking functional similarities between RW equation and Q-learning (after all, both learning rules are examples of the delta –error correction rule) that we can use to our advantage.

3 The PIQL Model

The PIQL model has been presented in a bottom-up manner: Firstly, states are defined as vectors of stimuli with a modality intensity value and USs as stimuli with a motivational value (aversive or appetitive). In order to deal with generalization, a similarity function between stimuli, $sim(s_i, s_j)$, is represented as the normal distribution of the stimulus features intensity for a given modality. In so doing, we allow for the transfer of associative strength between stimuli that share similar features. The asymptotic learning value is then defined as

$$\lambda US = max(\forall s \in RV : sim(s, US)),$$

where RV is a set of reinforcer values.

With this toolbox we proceed to calculate the learning rule for 1st order classical (Pavlovian) conditioning according to a version of the RW rule

$$V\Delta CS = \alpha\beta[max(\forall s \in RV : sim(s, US)) - \Sigma VCS_i],$$

in which the predictive strength of the CS (V) is stored as (CS,US) pairs in a Pavlovian Memory (PM) that forms a causal model of the world. Similarly, we define the learning rule for 1st order instrumental learning taking into account the sign of the US (its motivational value, aversive or appetitive). Instrumental S-(R-O) associations are then formed and constitute an Operant Memory (OM). These are similar to S-R associations, but unlike in RL,

- The agent has to value outcomes itself, and
- It can now predict actual outcomes in addition to their reward value.

In a second stage, the two memories PM and OM are used to compute 2nd order conditioning. The value of an outcome on an instrumental chain is defined as

$$\lambda OM = \lambda US \times s.category + \delta max(\forall o \in O : \Sigma_{ns \in NSc} Vns \rightarrow (R \rightarrow o)),$$

and the learning rule specified as

$$Vs \rightarrow (R \rightarrow o) \leftarrow Vs \rightarrow (R \rightarrow o) + \alpha\beta[\lambda OM - AOS]$$

where λOM is the instrumental outcome value stored in memory and the Aggregate reward Operant Strength (AOS) is updated according to

$$AOS \leftarrow \Sigma_{s \in Sc} Vs \rightarrow (R \rightarrow o)$$

The new Vs $\rightarrow o$ is calculated in parallel in the same way (just replacing λOM with λPM and AOS with APS, an Aggregate Pavlovian Strength).

The resulting PIQL look-ahead algorithm works as follows:

1. Starts initializing Sc (Current stimuli compound), NSc (Next stimuli compound) and r (response elicited to get from Sc to NSc) to $Nothing$, APS and AOS to 0, and δ (the decay parameter, γ in reinforcement learning) to, say, 0.9;

2. For each episode, the agent perceives Sc and chooses a response r according to $cR(Sc) \leftarrow (\forall r \in R, \forall o \in (Sc \rightarrow O) : max_r(\Sigma_{s \in Sc} Vs \rightarrow (r \rightarrow o) \times Vs \rightarrow o))$;

3. Then it perceives NSc and for each ns in NSc, updates APS and AOS, and calculates λPM and λOM;

4. With these values, it then updates $Vs \rightarrow ns$ and $Vs \rightarrow (R \rightarrow ns)$ for each s in Sc until an optimal policy is found.

The essence of the PIQL algorithm is that it uses the RW equation that maintains the sum of associative strength for all stimuli in a compound towards any other stimulus. Because each stimulus (feature) of the compound has a separate value, features can be shared across compounds. Hence, we predicted that PIQL should be able to generalize.

4 Experiments and Results

The PIQL algorithm was tested in a Grid simulator against convergence (where agents have to approach or avoid appetitive and aversive stimuli respectively), and, most importantly, generalization. Firstly, it was understood that as the algorithms converged to the optimal policy the error of the PM, the OM and the Q-values would decrease (increase in aversive scenarios) and, as a consequence, the average absolute fluctuation in associative strength would also decrease (increase in aversive scenarios). We found that, under certain conditions (that the decay parameter is in range $< 0, 1 >$ and that continued exploration is allowed) convergence is guaranteed. This is an important baseline result as convergence is a minimum requirement indicating that learning is occurring.

Secondly, we assumed that an algorithm which managed to gain savings effect from "sharing associative strength" will generalize. Generalization should be manifested in faster convergence to the optimal policy if the algorithm is successful in using the redundant associations to its benefit.

Several experiments were carried out in three Grid domains of increasing size: Each experiment had a Experimental Condition and a Control Condition (run 8 times each) and consisted of 2 phases: A training phase, P1, and a test phase, P2. In the Experimental Condition the domains in both phases shared some features allowing for generalization to occur, whereas in the Control Condition the features during P2 where different from those in P1. A **null hypothesis (H0)** was then defined:

> The mean Optimal Policy Re-found (OPR, the number of epochs before the OP was found and converged to) will not differ during P2 for Experimental Condition and Control Condition. Rejection point was established at $p < 0.05$.

An ANOVA was performed to test the hypothesis. The results were conclusive: OPRs were statistically different. H0 had to be rejected and the alternative hypothesis was accepted: The OPR was reached significantly faster in the Experimental Condition that in the Control Condition. That is, the Experimental Condition shows generalization between

phases. As expected, Q did not generalize, that is, there were no significant differences.

5 Further Work

Further research will focus on extending our generalization results and expanding the PIQL framework to solve the exploration *vs* exploitation equilibrium problem in RL[2]. In particular,

1. The extension of results will pursue the following objectives:
   - The generalization results will be tested in more complex environments (*i.e.*, in domains with more complex differences from P1 to P2);
   - New experiments will test generalization inside one Grid-world layout (*i.e.*, redundant stimuli spread across locations);
   - The PIQL rule would incorporate XOR capability to allow for sharing across locations. It can be rightly argued that in PIQL generalization comes at the price of ignoring how to deal with exclusive conditions (that Q-learning solves). This is a direct consequence of using elemental associations. To solve this problem, we will explore how to adapt PIQL using configural cues (Wagner and Rescorla:1972) or alternative elemental solutions.
2. The expansion of the model phase will focus on implementing successful exploration following two complementary paths:
   - So far, PM and OM have equal weight and, at the same time, run independently in our algorithm. It is our understanding that exploration would need to be guided by models the agents may have learned about their environment. This model must not depend on the agent's behaviour but act as an underlying causal model of the world. That is, we propose to implement our reward maximising PIQL on top of learned S-S associations. These associations would then constrain exploration avoiding unsuccessful paths and improving instrumental generalisation and overall speed[3]. In order to do so, we would need to understand better the nature of the relations between classical conditioning and instrumental conditioning. In particular, we plan to extend IPQL in the light of psychological and computational incentive models of learning (such as (Dickinson and Balleine:2002) and (Dayan and Balleine:2002)) and new hybrid

[2]Unlike other machine learning paradigms, RL agents explore (state-action pairs for which the outcome is unknown) and exploit (those state-action pairs for which rewards are known to be high) in the assumption that trying sub-optimal paths can eventually bring optimal results. However, unsuccessful exploratory policies slow down the process. Finding the right balance between exploration and exploitation (that is, a compromise between solution quality and speed) is still an open problem.

[3]This proposal is related to the work by (Singh *et al.*:2005) on intrinsically motivated reinforcement learning in that intrinsic rewards (our reinforcer values) are used to develop general skills through exploration. In our case, instead of the formation of habits, we will study how S-S models help agents guide goal-oriented behaviour.

models of associative learning (LePelley:2004);

- The exploration-exploitation equilibrium dilemma comes down to the problem of convergence in dynamic environments. Q-learning convergence with probability 1 requires stationary environments. To cater for stochastic environment dynamics the step size parameter α must be held constant. The problem of exploration-exploitation balance can be then reformulated as one of readjusting the policy to a non-stationary environment. Perhaps surprisingly, the study of variable learning rates both in reinforcement learning (eligibility traces (Singh and Sutton:1996)) and associative learning (attentional (CS-processing) models (Mackintosh:1975; Pearce and Hall:1980)) has not yield the expected results. We propose a reevaluation of α values according to CS activation states (Wagner:1981) and Holland's rules of association representation (Holland:1990).

3. In addition, we will exploit the interdisciplinary nature of our proposal in what we have called the psychology phase. The framework described in the previous sections is for all practical purposes a computational model of 1st and 2nd order (sequential) learning. Although RW model can be easily extended to explain such phenomena this is the first time that an explicit computational extension of the model that deals with goal-directed behaviour has been presented. We shall test the psychological plausibility of our model. We have already proved that PIQL correctly simulates acquisition, generalization, avoidance and second-order conditioning. We will need to test its predictive power against other phenomena such as extinction, blocking, overshadowing, latent inhibition, negative patterning and conditioned inhibition.

Our research complements alternative computational approaches to associative learning (see (Balkenius and Morén:1998) for a comparative study). However, unlike previous work in the area, we take a more psychological approach and directly adjust the RW model (that is a computational model in its own right) to deal with goal-oriented behaviour rather than replacing it with computational (typically neural networks) models of associative learning. We believe that both approaches are valid but that ours has been neglected and that, as a consequence, psychologists have largely ignored the latter as alien to their theories and methods.

References

Balkenius C. and Morén J. (1998), *Computational models of classical conditioned: a comparative study*, LUCS 62, Lund University Cognitive Studies.

Dayan, P. and Balleine, B.W. (2002), Reward, motivation and reinforcement learning. *Neuron*, **36**, 285-298.

Dickinson, A. and Balleine, B.W. (2002), The role of learning in the operation of motivational systems. In H. Pashler and R. Gallistel (Eds.), *Stevens' Handbook of Experimental Psychology* (Third ed., Vol. 3: Learning, motivation, and emotion, pp. 497-533). New York: John Wiley & Sons.

Holland, P.C. (1990), Event Representation in Pavlovian Conditioning: Image and Action. *Cognition*, 37, 105-131.

Mackintosh, N.J. (1975), A theory of attention: Variations in the associability of stimulus with reinforcement. *Psychological Review*, **82**, 276-298.

Pearce, J.M. and Hall, G. (1980). A model for Pavlovian conditioning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, **87**, 532-552.

Le Pelley, M.E. (2004), The role of associative history in models of associative learning: A selective review and a hybrid model. *The Quarterly Journal of Experimental Psychology*, 57B(**3**), 193-243.

Miller R.R., Barnet R.C. and Grahame N.J. (1995). Assessment of the Rescorla-Wanger model. *Psychological Bulletin*, 117(**3**), 363-386.

Rescorla, R.A. and Wagner, A.R. (1972), A theory of Pavlovian conditioning: The effectiveness of reinforcement and non-reinforcement. In A.H. Black and W.F. Prokasy, (editors), *Classical Conditioning II: Current Research and Theory*. New York: Aleton-Century-Crofts.

Singh, S. Barto, A.G. and Chentanez, N. (2005), Intrinsically Motivated Reinforcement Learning. To appear in *Proceedings of Advances in Neural Information Processing Systems 17 (NIPS)*.

Singh S.P. and Sutton R.S. (1996), Reinforcement learning with replacing eligibility traces. *Machine Learning*, **22**, 123-158.

Sutton, R.S. and Barto, A.G. (1998), *Reinforcement Learning: An Introduction*, Cambridge, MA: The MIT Press.

Sutton, R.S. and Barto A.G (1990), Time-Derivative Models of Pavlovian Reinforcement. In Gabriel M. and Moore J. (Eds.), *Learning and Computational Neuroscience: Foundations of Adaptive Networks*, 497-537, The MIT Press: Cambridge, MA.

Tesauro G. J. (1995), Temporal difference learning and TD-Gammon. *Communications of the ACM*, **38, 58-68**.

Thorndike, E.L. (1911), *Animal intelligence: Experimental studies*, New York: Macmillan.

Wagner, A.R. (1981), SOP: A model of automatic memory processing in animal behaviour. In N.E. Spear and R.R. Miller (Eds.), *Information processing in animals: Memory mechanisms*, 5-47, Hillsdale, NJ: Erlbaum.

Wagner, A.R. and Rescorla, R.A. (1972), Inhibition in Pavlovian conditioning: Application of theory. In R.A. Boakes and M.S. Halliday (Eds.), *Inhibition and Learning*, 301-336, Academic Press: New York.

Watkins C.J.C.H. (1989), *Learning from Delayed Rewards*. PhD Thesis, Cambridge University.

Model of Reinforcement Learning in the Mouse Reaching and Grasping Experiment

Shahzia Anjum
Mammalian Genetic Unit
Medical Research Council
s.anjum@har.mrc.ac.uk

Rafal Bogacz
Department of Computer Science
University of Bristol
R.Bogacz@bristol.ac.uk

Valter Tucci
Mammalian Genetics Unit
Medical Research Council
v.tucci@har.mrc.ac.uk

Patrick M.Nolan
Mammalian Genetics Unit
Medical Research Council
p.nolan@har.mrc.ac.uk

Abstract

The Mouse Reaching and Grasping Performance Scale (MoRaG) experiment is aimed at observing the effect of certain gene mutations on the cognitive abilities of the mice. In this experiment a mouse is placed in a transparent cage with an opening so small that food pellet placed on the other side cannot be reached by the mouse by nose poke but only by hand-reach. The main observation noted down during the MoRaG experiment was the number of nose pokes performed by each mutant mouse in an attempt to reach the target food pellet. This nose-poking action was followed by the hand-reach action which eventually led the mice towards successful retrieval of the food pellet. The sequence and number of nose-pokes and hand-reach actions performed by the animal during the MoRaG experiment is a result of two parameters: (i) speed of learning and (ii) amount of preference for exploratory behaviour. To enable the easy quantification and analysis of these two parameters a computational model was built. The model assumes that the mouse selects one of two actions: hand-reach or nose-poke, and each action is associated with a weight determining the probability of its selection, which is updated according to the Rescorla Wagner rule. For each type of mutant mice used during the MoRaG experiment the two parameters (describing speed of learning and preference for exploration) have been estimated using maximum likelihood method. The model was able to replicate the behaviour of the mice as observed during the MoRaG experiment and to quantify the cognitive abilities of the mutant mice successfully. Thereby helping the scientist involved with the MoRaG experiment to assess the effect the genetic mutation had on the mice

1 Introduction

The ability to use limb movements to reach, grasp and retrieve food is widespread amongst mammals suggesting a common evolutionary origin of such movements.

The control of multi-joint arm movements, such as placing the hand on a visually detected target, grasping and retrieving, requires the transformation of visually derived information (position of the target). This information is then further transformed into a motor command to position the hand and perform the grasping action.

Reaching, grasping and releasing functions are therefore goal-directed movements. Such highly complex dynamical processes are comprised of two main steps: (1) planning, and (2) execution.

Planning requires the use of the cognitive abilities to learn from past experiences and applying them in decisions regarding future actions. On the other hand, execution of goal directed movements involve the effective use of limbs.

Both planning and execution of goal directed movements are controlled by the central nervous system. Genetic mutations can affect this central nervous system in a way that may diminish the ability to execute goal-directed movements, at different levels of severity.

To study the effects of specific gene mutations on the functioning of the central nervous system, the Mouse Reaching and Grasp-

ing performance scale experiment (MoRaG) was devised. This MoRaG experiment aims to monitor the motor activity and the cognitive abilities of mutant mice when placed in the experimental setup.
To determine the degree to which the mice have been affected by the mutations, close analysis of their behaviour during the MoRaG experiment have to be carried out.
In comparison to the task of observing the mutant mice behaviour, quantifying the effects of the mutation proved to be extremely time consuming given the large volumes of experimental data.

To help with the analysis and quantification of the effect of genetic mutation on the cognitive abilities of the mice, a computational model was built.
Before embarking upon the description of this computational model, the MoRaG experimental setup is described in detail. Following which a detailed description of the computational model is given.

2 The Mouse Reaching and Performance Scale Experiment

In the Mouse Reaching and Grasping Performance Scale (MoRaG) experiment, mutant mice are placed in Plexiglas chambers that are 11.4 cm in height, 6.4 cm wide and 3.8 cm thick. In the front wall of each cubicle there is a 9mm hole and through which a Plexiglas feeding plane can be accessed.
On this feeding plane, small food pellets (approximately 2-3 mm diameter) are placed in such a way that the mice can withdraw the food only by reaching out one of its paws.
The diagram below presents a pictorial view of the MoRaG experimental setup.

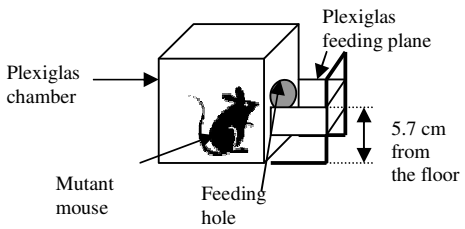


Figure I: Experimental setup for MoRaG

In the MoRaG experiment the upper-arm movement of the mouse is analysed in three sequential phases: 1) the reaching (the hand proceeds out of the trunk, on the horizontal plane, to approach an object); 2) the grasping (the object is grasped) and 3) the retrieval (the hand moves toward the mouth, with the retrieved object in its grasp).

To reach the food pellet, at first the mouse tries to poke its nose through the feeding hole. On failure, the mouse learns to reach out its paw to grasp the food.
A successful reaching is considered when the animal reaches the food pellets without any distortion or deviation from the target. The reaction time, that is the interval between the mouse seeing the food pellet and reaching it, is considered for all the trials except for the first one. This is mainly because at the first trial the mouse is completely unaware of the situation and the reaction time in that trial does not provide any information regarding the actual response time of the mouse.

A grasp is considered to be successful if the mouse can grasp the food pellet and most importantly, hold on to it.

Finally the behaviour of the mice during the experiment is measured using 26 parameters. The parameters measured during the MoRaG experiment are categorized as shown below:-

(1) Quantitative parameters (which are most relevant for the model): - During the MoRaG experiment, the quantitative parameters measured are the number of nose pokes before a hand reach action, number of successful reaches, number of successful grasps etc.

(2) Assessment of behaviour in the box: - Under this category are those parameters that were accessed before the mouse starts to perform any goal-directed motor behaviour. For instance body posture, sniffing and/or grooming and also the rate of activity in the box.

(3) Assessment of behaviour while the mouse is approaching the target: - For this category, the behaviour of the mice are observed under three phases - outward, reversal and inward phase. For the outward phase, parameters like trunk displacement, body rotation, and arm movements are measured. While for the reversal

phase, parameters like grasping preparation, grasping closure are measured. And finally for the inward phase, parameters like strategies are measured.
The MoRaG experiment was performed upon 6 different categories of mutant mice namely the C3H, BALB/c, SvPas, 129SVEVM, BI/6 and 129NM.

The assessment of mutant mice behaviour during the MoRaG experiment showed considerable differences in the cognitive and motor abilities of the mice. These differences can be attributed to the difference in their gene mutations.

In the MoRaG experiment, two action choices were available to the mutant mouse.
One option was to try and reach the target food pellet through nose pokes, generally referred to as the nose poke action throughout this paper. And the other option was to reach out its arm and try to grasp the target, referred to as the hand reach action in this paper.

It was observed during the MoRaG experiment that the mice usually tried to reach the food pellet through nose pokes and finally, realising that the food pellet was too far away, performed a hand reach action.

Repeated trials showed a constant decline in the number of nose pokes before the first hand reach suggesting the ability of the mice to learn from experience.

In a broader sense, the behaviour of the mice during the MoRaG experiment are influenced by two parameters - (i) speed of learning and (ii) amount of preference for exploratory behaviour. These two parameters may be influenced by different genes; hence it is useful to separate these two parameters on the basis of behavioural data. The means for such separation is provided by a computational model of reinforcement learning which has been successfully used to describe behaviour of animals and humans during sequences of choices between two alternative actions (Montague et al, 1995; Montague & Berns, 2002; O'Dorethy, 2004).This model provides probabilistic description for animal behaviour, hence the two parameters in question can be estimated using Maximum Likelihood method, as the values which maximize the likelihood of observed sequence of nose-pokes and hand-reaches given the model.

3 The Computational Model

As mentioned earlier, the computational model built for the MoRaG experiment was aimed at quantifying the degree to which the cognitive abilities of each group of mice were affected by the genetic mutations.

To enable close analysis of this learning ability of the mutant mice, their behaviour was modelled into a computational system as described below.

In the computational system, the two action choices, nose poke and hand reach, are given individual weights.
Since, the mutant mice always try to reach the food through nose pokes before attempting a hand reach, the nose poke (np) action was allotted an initial weight of $w_{np} = 1$. And intuitively, the weight associated with the hand reach (hr) action was initialized to $w_{hr} = 0$. The values of these weights controlled the probability of the associated action being selected.

Moreover, as a result of the weight initializations, the probability of selecting the nose poke action at the beginning of a trial was 1 while the probability of selecting the hand reach action at the beginning was 0.

Throughout each run of the computational system, the weights and probabilities associated with each action were continually updated.

If an action i is selected (where i is np or hr), the corresponding weight w_i is updated on the basis of the Rescorla–Wagner rule (Rescorla & Wagner, 1972) as shown below:

$$w_i = w_i + \text{Learning rate} * \delta \quad (1)$$

The above equation suggests that the change in a weight is a product of two terms: the learning rate, describing the speed of learning, and δ – equal to the difference between reward obtained and reward predicted by the animal. The model assumes that the weight associated with an action is actually equal to the predicted reward for this action. Hence δ is equal to (Montague et al., 1995):

$$\delta = \text{Reward obtained} - w_i \quad (2)$$

In the computational model built the '*Reward obtained*' is set to 1 when animal chooses

hand-reach and is set to 0 when the animal chooses the nose-poke.
Now, the probability of choosing the two actions is computed from the following equation (Montague et al., 1995):

$$\text{Probability of nose poke (Pns)} = \exp(\mu w_{ns}) / (\exp(\mu w_{ns}) + \exp(\mu w_{hr}))$$

$$\text{Probability of hand reach (Phr)} = \exp(\mu w_{hr}) / (\exp(\mu w_{ns}) + \exp(\mu w_{hr})) \quad (3)$$

The probability associated with each action depends on three parameters: two weights w_{ns}, w_{hr} associated with the actions, and μ – the Threshold value which characterizes the animal's preference for exploration: the higher μ, the less likely the animal is to choose action associated with lower weight.

Along the lines of the mutant mice behaviour observed during the MoRaG experiment, the probability of the nose poke action is at its peak at the start of the trial. And as the trial progresses, every time the nose poke action does not result in a reward, its probability of being selected the next time around is reduced.
On the other hand, the probability of the hand reach action is being simultaneously increased. This continues until there comes a point in time where the probability of selecting the hand reach action is higher than that of the nose poke action. In this case, the action hand reach will be selected, thereby ending the trial with the retrieval of the food target.

Having outlined the basics of a computational model that is capable of imitating the mutant mice behaviour, there are still two important factors to be decided upon. One of them is the value of the Learning Rate, used in equation 1 above, and second is the value of the Threshold parameter (μ) used in equation 2 above.
Since each category of mice used in the MoRaG experiment differs from one another in their type of genetic mutation, they will show varying cognitive abilities. Thus, each category of mutant mice would have their unique Learning rate and Threshold value. And these are the values that need to be estimated by the computational model.

The basic idea behind this is an optimization method where those values of Learning Rate and Threshold are sought that produce behaviour (i.e. sequences nose pokes and hand reaches) that is exactly the same as the behaviour of that category of mutant mice whose cognitive abilities are being quantified.

Since obviously this is a very tedious process if done manually, a more automated approach is adopted following along the lines of the Maximum Likelihood parameter estimation theory.
Very briefly, the main idea of the Maximum Likelihood parameter estimation theory is to calculate the parameter values that maximize the probability of the given sample of data.
The likelihood of data from a given category of mutant mice is equal to:

$$Likelihood = \prod_{t=1}^{nt} \prod_{h=1}^{nh} P_h\left(d_{t,h} \mid model\right) \quad (4)$$

where nh is the number of MoRaG trials performed by the mice, and each d_t is a record of the behaviour of that particular category of mutant mice during trial t. In particular, d_t is a vector of the length equal to the number of hand reaches nh performed by the mice in a given trial. Each entry $d_{t,h}$ in this vector contains the number of nose-pokes performed by the mice before hand reach h.
For example, during the first trial on say the C3H type of mutant mice, i.e. t=1, the C3H mice are repeatedly placed in the MoRaG experimental setup. For each of those times, the number of nose pokes they perform before a hand reach is recorded to form d_1.
Below is an example of what d_1 may look like:

Table I:Example MoRaG Data

Hand Reach Index, h	No. of Nose Pokes before a Hand Reach, $d_{1,h}$
1	12
2	9
3	4
4	0
5	0
6	0
7	0
8	0
9	0
10	0

The probability distributions $P_h(d|model)$ describe the probability of performing d nose pokes before hand reach h for given

parameters of the model. They are estimated by simulating the model with a given set of parameters 1000 times.
Basically, a table t compromising of the hand reach index (h) by nose pokes (d) is maintained.
During the 1000 repetitions for the given set of parameter values, this table is updated continually according to the observed model behaviour. This is done by incrementing the counter of t[h,d] by 1 every time the model performs d nose pokes before the h^{th} hand reach.
Thus, at the end of the 1000 repetitions, the probability of observing d nose pokes during the h^{th} hand reach can be easily estimated by the value of the counter in t[h,d]/1000.
These values are then used to compute the value of $P_h(d|model)$.
In simpler terms the entire computational model for the MoRaG experiment can be viewed as a two part system.
Part 1 of the system is given as input-randomly chosen values of Learning rate and Threshold parameters. It then uses these two values, as shown in equations 1, 2 and 3 above, to produce 1000 sequences of model generated behaviour that is similar to the mutant mice behaviour.
The generated behavioural data is then passed onto the second part of the computational model where the probability distributions, $P_h(d|model)$, are calculated. This ultimately produces the likelihood value, as shown in equation 4, for the given set of parameter values.

In a larger view, the main task of the second part of the model is to repeatedly try various combinations of input values of the Learning Rate and Threshold parameters and pass it on to Part 1 of the model. And each time, the outputs of Part 1, i.e. the behavioural data, are used by the second part of the model to calculate the likelihood for the set of parameter values given as input to the computational model at that time.
The combination of Learning Rate and Threshold values that produce the maximum likelihood is considered as the required value of these two parameters for the category of mutant mice being considered at that point.
The results or values of the Learning Rate and Threshold parameter, estimated by the computational model, for each category of mutant mice used in the MoRaG experiment are described in the following section.

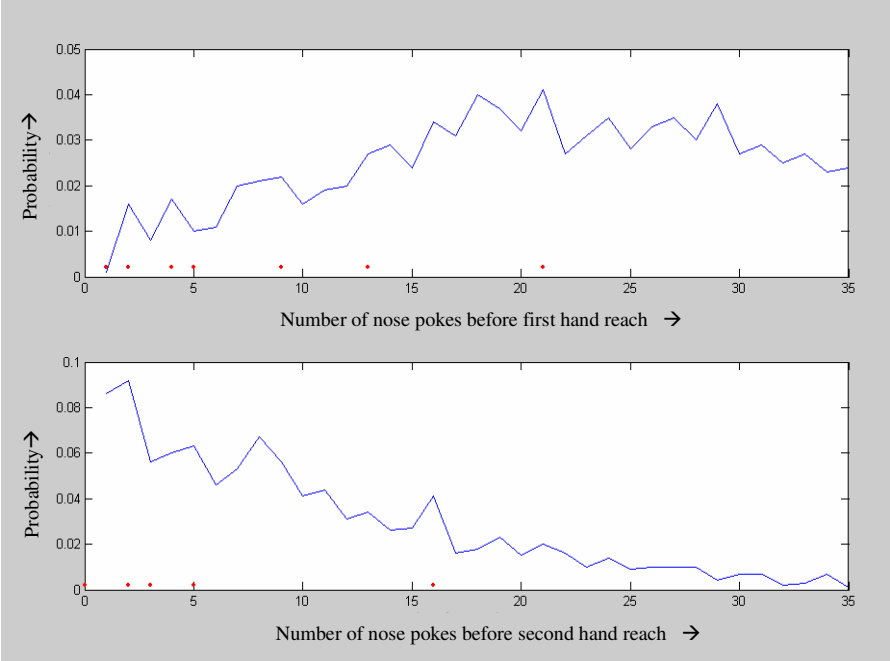
3.1 Results of the Computational Model

In the MoRaG experiment, there were six different types of mutant mice. The SvPas strain, the BALB/c strain, the C3H strain, the 129SVEVM strain, the BI/6 and the 129NM strain. For each of these strains or types of mutant mice, the Threshold value provides a measure of how deterministic this category of mice was when choosing between the two actions, nose pokes and hand reach. Thus, in more general terms, the Threshold value quantifies the animal's preference for exploratory behaviour after the genetic mutation.
The Learning Rate calculated for each of the six types of mutant mice quantify the ability of the mice to learn the correct action i.e. the hand reach action, in order to retrieve the food.
The table below shows the values of Threshold and Learning Rate calculated for each category of mice by the computational model.

Table II: Results

	Threshold	Learning Rate
SvPas	4.855385	0.025823
C3H	30.11	0.244
BALB/C	39.5032	4.2966
129SVEVM	7.8	0.00102
129NM	7.575	0.00877
BI/6	23.0	0.3

To verify the plausibility of the estimated learning rate and threshold values, the sequence of nose pokes and hand reach actions produced by the computational model are plotted. And on each such plot, the corresponding MoRaG data were visualized to see if the actual data points fit the model generated behaviour (as shown in Figure II).
It was found that while for the SvPas, C3H, 129SVEVM, BI/6 and 129NM the estimated threshold and learning rate values were plausible, for the BALB/c, the results were incorrect and the plot of the computational model generated behaviour could not fit the actual MoRaG observations for this type of mice. This failure was attributed to the fact that the



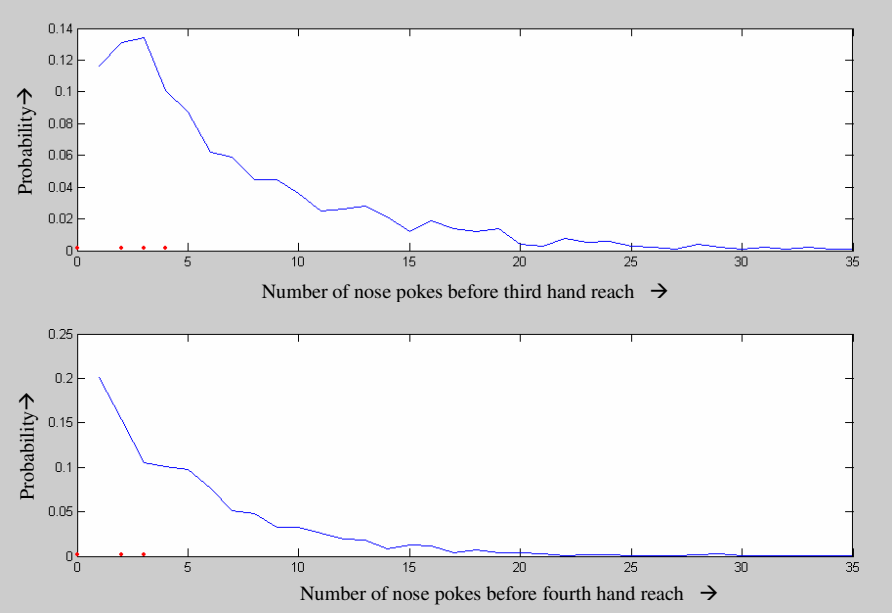


Figure II: Plots for testing the results of the SvPas Strain

MoRaG experimental data also displayed completely random behaviour by the mice under this BALB/c group and thus their behaviour could not be generalised as being of a certain type. Moreover, the learning rate and threshold factor value for each mouse in this group was varied and thus could not be generalised as one value.

4 Critical Evaluation and Summary

The computational model built to help the scientists in analyzing the effects of genetic mutations on mice was successful to a considerable extent.

However, it was observed from the MoRaG data that the BALB/c strain of mutant mice displayed behaviour opposite to what was observed with the rest of the mice. Out of the 10 BALB/c mice, 4 of them failed to learn from past trial experiences and the remaining, managed to learn the correct action to reach the food after 4-5 trials. In contradiction to this behaviour, it took just 1 to 3 trials for the other strains of mutant mice to learn the correct action to retrieve the food.

While the behaviour of the BI/6, C3H, 129NM, 129SVEVM and 129NM strains of mutant mice could be imitated by the model with the help of the Rescorla Wagner rule, the random behaviour of the BALB/c strain could not be reproduced by this model. Thus as a part of future development, the computational model can be modified such that it is able to quantify the results of the BALB/c type of mice as well.

In addition, the Threshold and Learning rates calculated by the computational model were computed using 10-14 sets of MoRaG experimental data for each type of mutant mice. Larger volumes of experimental data would obviously add to the confidence in the results of the computational model.

Moreover, during the quantification of the effects of mutation on the mutant mice, individual characteristics of the mice were not taken into consideration.
Thus, another scope of development is to calculate the Learning Rate and Threshold value for each mouse in the MoRaG experiment. This will help in studying the role of individual characteristics like age, weight etc on the degree to which genetic mutation can affect each mouse.

While examining the MoRaG data it was observed that the data did not contain information on whether the same mutant mouse has been tested upon on another date.
If this information could be gathered, the effect of mutation on the ability of the mice to remember past experiences and use the knowledge from those experiences to guide actions during future trials, could also be quantified.

In Conclusion, automating the process of identifying the effect of different types of genetic mutations on mice is capable of drastically cutting down the overall time spent on an experiment. Such was the case during the MoRaG experiment. However, there still remains ample scope of improving the current computational model, enabling it to better its own performance.

References

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R.J. (2004) *Dissociable Roles of Ventral and Dorsal Striatum in Instrumental Conditioning. Science,* 304: 452-454.

Montague, P.R., & Berns, G.S. (2002) *Neural economics and the biological substrates of valuation.* Neuron, 36: 265-284.

Montague, P.R., Dayan, P., Person, C, & Sejnowski, T.J. (1995) *Bee foraging in uncertain environments using predictive hebbian learning*. Nature, 377: 683-684.

Rescorla, W.A., Wagner, A.R. (1972). *A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement.* In Classical conditioning II: Current research and theory New York: Appleton.

Reinforcement learning in continuous time and space: Interference and not ill-conditioning is the main problem when using distributed function approximators

Bart Baddeley*
*Centre for Computational Neuroscience and Robotics
Department of Informatics
University of Sussex
Brighton, UK
bartbaddeley@aol.com

Abstract

If a problem is low-dimensional and discrete then it is straight-forward to apply reinforcement learning techniques to find optimal policies. Many interesting problems are continuous and/or high-dimensional though. Reinforcement learning in a continuous time, state and action formulation requires the use of function approximators. Often, local linear models have been preferred over distributed non-linear models for function approximation in this instance. Coulom (2002) suggests that a major reason for this is that the optimisation problem, that must be solved in order to train a non-linear, distributed function approximator, is often ill-conditioned, resulting in slow or unstable learning. We suggest that another reason for the difficulty in learning a value function using a distributed architecture, is the problem of negative interference, whereby learning of new data disrupts previously learned mappings. A continuous Temporal Difference (TD) learning algorithm, $TD(\lambda)$, (Doya, 2000), was used to learn a value function in a limited torque, pendulum swing-up task using a Multi Layer Perceptron (MLP) network for function approximation. Three different approaches were examined for learning in the MLP networks; 1) simple gradient descent, 2) vario-eta (Neuneier and Zimmermann, 1998), the method suggested by Coulom to reduce the effects of ill-conditioning of the learning problem, and 3) a *pseudopattern* rehearsal strategy (Robins, 1995), that attempts to reduce the effects of interference. Our results show that MLP networks can indeed be used for value function approximation in this task, but, require long training times. More interestingly, we found that vario-eta destabilised learning and resulted in a failure of the learning process to converge. Finally we showed that the *pseudopattern* rehearsal strategy drastically improved the speed of learning. The results indicate that interference is a greater problem than ill-conditioning for this task. And also, that most acceleration techniques that address the problem of ill-conditioning will actually exacerbate the problems of interference when attempting to learn a value function using a distributed feedforward neural network.

1 Introduction

A continuous formulation of the reinforcement learning problem provides a nice framework for addressing problems in motor learning. Maximising a, possibly, delayed reward over over a continuous series of continuous actions extended in time is exactly the situation we have in many motor learning situations, consider for example, throwing at a target. A current issue in reinforcement learning is therefore, how best to deal with continuous and/or high dimensional problems of the sort likely to be encountered in motor control? The issue is interesting for at least two reasons. Firstly, from a purely engineering perspective, a continuous formulation allows reinforcement learning techniques to be applied to a wider range of problems. Secondly, in TD models of the basal ganglia the TD error is proposed to represent the reward prediction error. If a continuous TD error could be shown to match a dopamine signal, then continuous TD models might ultimately shed light on the role of the basal ganglia in the control of movements.

One way to handle continuous states and actions is to discretise the state and action spaces and then use one of the standard discrete reinforcement learning algorithms to solve the problem [see Sutton and Barto (1998) or Kaelbling et al. (1996) for a survey

of reinforcement learning approaches]. The number of discrete states that must be mapped grows exponentially with the size of the input dimension and this approach is therefore unsuitable for all but the most basic of motor behaviours.

Alternatively, a function approximator can be employed to learn a value function and there are numerous examples of the successful use of function approximators in the reinforcement learning literature (Anderson, 1987; Baird, 1993; Tesauro, 1995; Gordon, 1995; Bertsekas and Tsitsiklis, 1996; Sutton and Barto, 1998; Sutton et al., 2000). In principal any function approximator can be used to learn and estimate a value function, however, local linear models have proved to be the method of choice in much of the work. There are good reasons for this. Local linear models are able to make better incremental use of training data than distributed or non-linear function approximation schemes. This is because linear models are easier to learn than non-linear models, and because local models are less susceptible to "unlearning" due to interference than distributed models are. Interference occurs when learning in a small region of input space disrupts learning outside this small area. Resistance to interference means that local models can employ higher learning rates safely, which in turn, allows faster convergence of the reinforcement learning algorithm.

The main drawback to applying local linear models is that the number of models grows rapidly with the dimension of the input. The smaller the region of trust for each individual local linear model is, then the greater the resistance to interference becomes, but at the cost of requiring more local models in order to span the input space. The problem becomes exponentially worse with increasing dimensionality, quickly leading to an unfeasibly large number of local models if the whole input space is to be covered. This would appear to be one of the main stumbling blocks preventing reinforcement learning from being scaled up for use in higher dimensional continuous problem domains.

In this paper we explore the suitability of learning a value function using a Multi Layer Perceptron (MLP) network. The main motivation for choosing to use a MLP is that in general they deal better with high dimensional problems than local linear models do (Barron, 1993). This can be particularly true if there are irrelevant dimensions and/or global constraints, such as one particular dimension passing a threshold signaling failure irrespective of the values of all other input variables. MLP networks are also generally more compact, meaning that there are fewer parameters that need to be determined by the training data. This provides another potential motivation for using MLP networks as opposed to local linear models.

The algorithm we employed in order to learn a value function in our experiments is Doya's continuous $TD(\lambda)$ (Doya, 2000). This is a continuous time and state formulation of Sutton original discrete $TD(\lambda)$ (Sutton, 1988). We implement the pendulum swing up task used in Doya's original work as a testbed for examining alternative schemes.

We employ a standard 3 layer MLP network with sigmoid activation functions and linear output units. Three different approaches are used to train the MLP networks; simple gradient descent, vario eta, a method that attempts to speed learning by reducing the ill-conditioning of the training problem, and finally a *pseudopattern* rehearsal strategy that is described below.

The approach using vario eta is similar to that employed by Coulom (2002) who uses continuous $TD(\lambda)$ in combination with a cascade correlation function approximator to learn a high dimensional motor control task. Coulom reports mixed results, with performance increasing initially, before a subsequent collapse. We note that the mean squared temporal difference (TD) error appears to diverge in his experiments. We examine Coulom's approach and posit a possible explanation for his results.

The third training regime attempts to reduce the problems of interference using a *pseudopattern* rehearsal strategy (Robins, 1995; Ans and Rousset, 1997; French, 1997). As with all *pseudopattern* approaches, interference is reduced by rehearsing with self-generated data. Usually the data, or *pseudopatterns*, are generated by presenting input to the trained network and recording the output with the *pseudopatterns* then interspersed with the incoming data. We make two changes to this standard approach which we describe later in the chapter.

The paper has the following structure. We begin with a description of Doya's continuous $TD(\lambda)$ algorithm and specify in detail the pendulum swing up task. Three experiments are then reported that describe the use of different approaches to learning using a MLP. Firstly, we show that it is possible to learn a value function using a MLP and simple vanilla gradient descent in this task. Secondly, we describe the results obtained using the vario eta algorithm to train MLP networks in an approach similar to that used by Coulom (2002). Finally, we show that using a *pseudopattern* rehearsal strategy, we can greatly improve learning performance in this instance. We conclude with a summary of results and a discussion of

their implications.

2 Continuous $TD(\lambda)$

For the experiments reported in this chapter we implemented Doya's (2000) continuous $TD(\lambda)$ algorithm to learn a value function. The algorithm is a continuous version of Sutton's (1998) discrete $TD(\lambda)$. In this formulation a function approximator is used to learn a value function using a continuous-time counterpart of the TD error that is derived as follows.

Optimal Value Function for a Discounted Reward Task

Consider a continuous-time deterministic system governed by the following differential equation

$$\dot{x}(t) = f(x(t), u(t)) \tag{1}$$

where $x \in X \subset \Re^n$ is the state and $u \in U \subset \Re^m$ is the action (control input). The immediate reward for a given state and action is

$$r(t) = R(x(t), u(t)) \tag{2}$$

The goal is then to find a policy or control law

$$u(t) = \pi(x(t)) \tag{3}$$

that maximises the discounted future rewards

$$V^{\pi}(x(t)) = \int_t^{\infty} e^{\frac{s-t}{\tau}} R(x(s), u(s)) ds \tag{4}$$

for any initial state $x(t)$. $V^{\pi}(x(t))$ is the value function of the state x under policy π and τ is a time constant for discounting future rewards. The value function for the optimal policy π^* can then be defined as

$$V^*(x(t)) = \max_{t,\infty)} \left[r(x(t), u(t)) + \frac{\partial V^*(x)}{\partial x} f(x(t), u(t)) \right] \tag{5}$$

where $[t, \infty)$ is the time course $u(s) \in U$ of all future controls for $t \leq s < \infty$. The condition for the optimal value function at time t is then given by

$$\frac{1}{\tau} V^*(x(t)) = \max_{u(t) \in U} \left[r(x(t), u(t)) + \frac{\partial V^*(x)}{\partial x} f(x(t), u(t)) \right] \tag{6}$$

which is the discounted Hamilton-Jackobi-Bellman (HJB) equation (Peterson, 1993). The optimal policy is specified by the action that maximises the right-hand side of the HJB equation.

$$u(t) = \pi_*(x(t)) = \arg\max_{u \in U} \left[r(x(t), u(t)) + \frac{\partial V^*(x)}{\partial x} f(x(t), u(t)) \right] \tag{7}$$

2.1 Learning the Value Function

When using function approximators, learning the value function involves changing the adjustable parameters of the function approximator $V(x(t); w)$ in order to minimise a continuous time version of the TD error. TD learning works by attempting to satisfy a consistency condition that is local in time and space. By differentiating definition (4) with respect to time we arrive at consistency condition that should be met when the estimate of the value function is perfect.

$$\dot{V}^{\pi}(x(t)) = \frac{1}{\tau} V^{\pi}(x(t)) - r(t) \tag{8}$$

If the estimate is not perfect then the discrepancy

$$\delta(t) = r(t) - \frac{1}{\tau} V(t) + \dot{V}(t) \tag{9}$$

is used to adjust the estimate. This is the continuous time counterpart of the TD error described by Sutton (1988).

2.2 Exponential Eligibility Traces: The Full $TD(\lambda)$ Algorithm

In principle the value function could be learned by gradient descent on an objective function of the form $E(t) = \frac{1}{2}[\delta(t)]^2$ (Baird, 1993). Resulting in the following update equation:

$$\dot{w}_i = -\eta \frac{\partial E}{\partial w_i} = \eta \delta(t) \left[\frac{1}{\tau} \frac{\partial V(x; w)}{\partial w_i} + \frac{\partial}{\partial w_i} \left(\frac{\partial V(x; w)}{\partial x} \right) \dot{x}(t) \right] \tag{10}$$

where η is the learning rate. However, eligibility traces have been shown to speed learning considerably by allowing updates to be dependent on recent experience as well as the current state of the system. In continuous $TD(\lambda)$ updates are made according to

$$\dot{w}_i = \eta \delta(t) e_i(t), \tag{11}$$

$$\dot{e}_i = -\frac{1}{\kappa} e_i(t) + \frac{\partial V(x(t); w)}{\partial w_i} \tag{12}$$

where e_i is the eligibility trace for parameter w_i, and $0 < \kappa \leq 1$ is the time constant for the eligibility trace.

2.3 A Value Gradient Based Policy

In most reinforcement learning approaches (Sutton and Barto, 1998; Kaelbling et al., 1996), learning consists of an iterative process, called policy iteration, with two distinct phases:

1. In the first phase the value function is estimated for the current fixed policy.
2. In the second phase the policy is updated to make it greedy with respect to the current value function estimate. Meaning the action that maximises value function is chosen as the action to perform in each state.

While there are convergence proofs for this process in the discrete case, the use of function approximators complicates the situation and convergence has only been proved for certain forms of function approximators (Sutton et al., 2000). In practice, there are numerous examples of function approximators successfully being used to approximate and learn a value function.

In Doya's approach a value gradient policy is used where the steepest ascent direction of the value function $\frac{\partial V(x)}{\partial x}^T$ is transformed by the gain matrix $\frac{\partial f(x,u)}{\partial u}^T$ of the system dynamics into a direction in action space, the amplitude of the control action is controlled by an inverse sigmoid function. This results in the following control policy:

$$u = u^{\max} s \left(\frac{1}{c} \frac{\partial f(x,u)}{\partial u}^T \frac{\partial V(x)}{\partial x}^T + \sigma \right) \quad (13)$$

where $u^{\max}$ is the maximum allowable action amplitude, $s()$ is a sigmoid function that saturates as $s(\pm\infty) = \pm 1$, c is the action cost and σ is noise included to encourage exploration. In the limit where $c \to 0$, the control will be a $bang-bang$ policy (Bellman et al., 1956).

$$u = u^{\max} sign \left[\frac{\partial f(x,u)}{\partial u}^T \frac{\partial V(x)}{\partial x}^T \right] \quad (14)$$

3 Experiments

The problem: Pendulum Swing Up with Limited Torque

The control task that we used in our investigation was the pendulum swing up task (Doya, 1996, 2000). The task involves swinging a pendulum up into the vertical position and balancing it there. There are two state variables that describe the system, the angular position (θ) and velocity (ω), and a single control variable (u) that determines the torque that is applied about the fixed end of the pendulum. The dynamics of the system are fully described by the following differential equation.

$$\dot{\omega} = \frac{u + mgl\sin(\theta) + \mu\omega}{ml^2}$$

Where u is the commanded torque, m is the mass of the pendulum, g is gravity, l is the length of the pendulum, θ is the angle of the pendulum, μ is the coefficient of friction and ω is the angular velocity. If the maximum allowable controlled torque u_{max} is less than the maximum load torque mgl then this simple one degree of freedom system provides a non-trivial control problem. In order to perform the task successfully the controller has to swing the pendulum several times in order to build up momentum and also to decelerate the pendulum fast enough to stabilise the pendulum in the upright position. The reward was given as the height of the free end of the pendulum:

$$R(x) = cos(\theta)$$

Each trial was started from an initial state $x(0) = (\theta(0), 0)$, where $\theta(0)$ was selected randomly from the range $[-\pi, \pi]$. Each trial lasted 20 seconds unless the pendulum was over-rotated $(\theta > 5\pi)$, in which case the trial was terminated and a reward of -1 was given for 1 second. In all experiments the following parameter values were used. Maximum torque u_{max}=5, $m = 1$, $g = 9.81$, $l = 1$, $\mu = 0.01$.

3.1 Solution 1: Learning a Value Function Using a MLP Network

Having fixed upon an MLP architecture with which to learn a value function a series of evaluations were performed in order to determine an approximately optimal learning rate μ to use for on-line learning of the value function in an implementation of continuous $TD(\lambda)$. It was immediately obvious that learning would be slow, and long training times were therefore necessary. In order to explore various different learning rates, performance was assessed by measuring the total accumulated reward following 500 episodes of learning. A range of different learning rates were examined and the results are shown in figure 1. Having determined an approximately optimal learning rate, $TD(\lambda)$ was run to convergence to determine whether the approach would actually converge to a solution sufficient to solve the control problem.

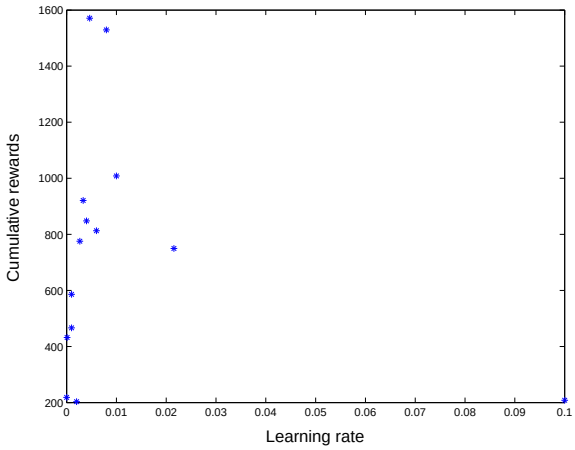


Figure 1: Comparison of different learning rates for a MLP network with 15 hidden units trained to approximate the value function for the task of swinging up a pendulum with limited torque. Performance is measured in terms of sum total reward following 500 episodes with a cumulative reward of 500000 representing the maximum possible. Although the overall performance is quite poor, there is a peak in performance for a learning rate of $\mu \approx 0.008$.

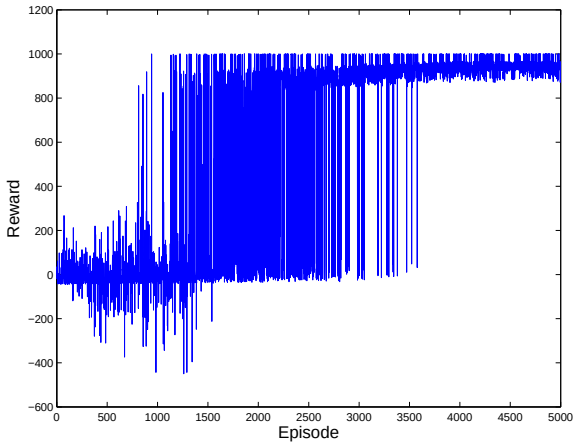


Figure 2: The reward performance of the controller when learning is run to convergence using a MLP network with 15 hidden units and a learning rate of 0.008. Performance is measured in terms of sum total reward during a 20 second episode, with $dt = 0.02$ the maximum reward per episode was 1000.

Experiment 1: Results

Figure 1 shows the performance of the continuous $TD(\lambda)$ algorithm using a 15 hidden unit MLP network and various different learning rates for the pendulum swing up task. The results show that, although the overall performance is quite poor for all learning rates, there is a peak in performance for a learning rate of $\mu \approx 0.008$. Figures 2 and 3 show the performance of the controller when learning is run to convergence using a MLP network with 15 hidden units and a learning rate of 0.008. Performance is measured in terms of sum total reward during a 20 second episode, with $dt = 0.02$ the maximum reward per episode was 1000. Figure 2 shows that it takes around 3500 episodes for the learning process to converge to the correct solution. Inspection of the sum squared TD error [figure 3] also reveals convergence. The landscape of the learned value function is shown in figure 4.

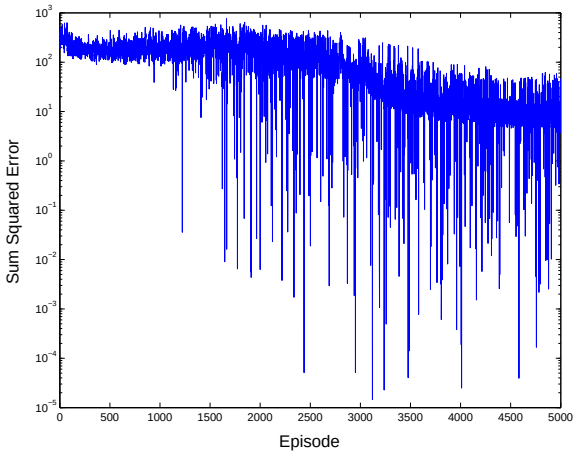


Figure 3: The sum squared TD error performance of the controller when learning is run to convergence using a MLP network with 15 hidden units and a learning rate of 0.008.

3.2 Solution 2: Vario-Eta - Investigating Ill Conditioning of the Learning Problem

In (Coulom, 2002), it was suggested that one of the main problems that needs to be addressed in order

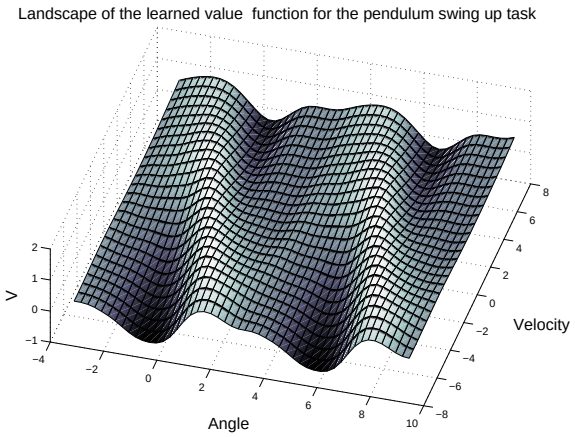


Figure 4: The landscape of the learned value function approximated using a MLP network.

for feedforward neural networks to be employed efficiently for learning a value function, is the problem of ill-conditioning. Ill-conditioning describes a situation where learning is made hard due to the approximation error being far more sensitive to some parameters than others. Therefore if we make parameter updates using gradient descent on the squared approximation error and a global learning rate, some parameters will adapt too fast while others will adapt too slowly. This results in a situation where it is difficult to determine a global learning rate that is suitable for all parameter updates and consequently leads to poor learning performance. The solution to the problem of ill-conditioning involves taking into account the curvature of the error surface and there are many approaches to doing this (Moller, 1993; Bishop, 1995; Levenberg, 1944; Marquardt, 1963; Amari, 1998; Riedmiller and Braun, 1993). The curvature is described by the matrix of second partial derivatives of the squared approximation error with respect to the weights of the network, the so called Hessian matrix. By scaling a global learning rate by the inverse of the Hessian, we can get better conditioning. Thus, we take larger steps along dimensions where the gradient is changing slowly and smaller steps when the gradient is changing more quickly.

Calculation of the Hessian requires $O(W^2)$ operations per pattern to evaluate, (Bishop, 1995), which can become prohibitive for larger networks. The vario-eta algorithm (Neuneier and Zimmermann, 1998), attempts to address this issue by taking a slightly different approach to the problem of ill-conditioning. Individual learning rates can be set by making the observation that weights that exhibit a high degree of variance during learning require a lower learning rate than those that are relatively stable. Therefore, by dividing a global learning rate, μ_{global}, by a by an estimate of a weight's variance, $VAR(w_i)$, a local learning rate $\mu_i = \frac{\mu_{global}}{VAR(w_i)}$, can be defined for each individual weight. It is hoped that improved conditioning of the learning process can be achieved using this approach resulting in faster convergence.

If, as Coulom suggests, ill-conditioning is the main problem that needs to be addressed, then we should expect that using vario-eta will improve performance when compared with simple gradient descent. In the next set of experiments we explored whether using the vario-eta algorithm could speed up learning of a value function approximated using a MLP. A MLP with 15 hidden units was trained using the continuous $TD(\lambda)$ approach with the learning rate for each weight determined by the vario-eta algorithm. Various different global learning rates were investigated in order to determine an approximate optimum value. As in the previous experiment the performance was assessed by measuring the total accumulated reward following 500 episodes of learning. Having determined an approximately optimal learning rate, a long run of 5000 episodes was performed to investigate convergence of the approach.

Experiment 2: Results

Figure 5 shows the results of an exploration of different global learning rates for the vario-eta algorithm when employed for learning a value function with a MLP using continuous $TD(\lambda)$. There is no clear peak as was the case when a single global learning rate used for all weight updates, however, there is some suggestion of a peak around $\mu \approx 0.0001$. The two approximately optimal learning rates, for straightforward gradient descent and vario-eta, produce comparable performance following 500 episodes of learning, suggesting that learning has not been improved significantly. Figure 6 shows the progress of learning using a global learning rate of 0.0001 for 5000 episodes. Performance appears to improve initially but soon becomes unstable and does not appear to converge. Examination of the sum squared TD error in figure 7 shows a similar pattern with a failure to converge. These results are similar to the pattern of results experienced by Coulom in his experiments with feedforward networks. It would appear that employing vario-eta, may actually hinder learning rather than helping, in this instance.

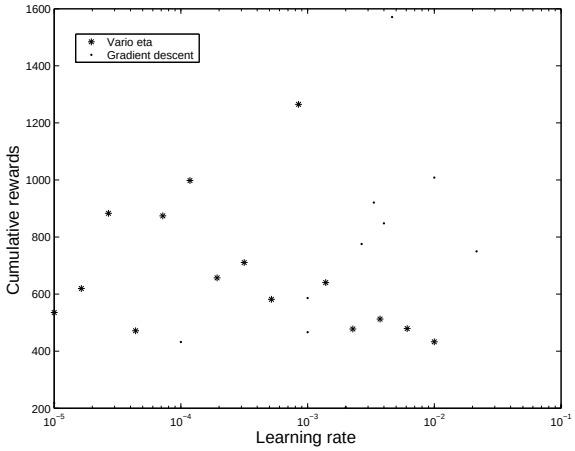


Figure 5: Results of an exploration of different global learning rates for the vario-eta algorithm (blue *) when employed for learning a value function with a MLP using continuous $TD(\lambda)$. Performance using different learning rates and simple gradient descent are included for comparison (red .). Performance is measured in terms of sum total reward following 500 episodes.

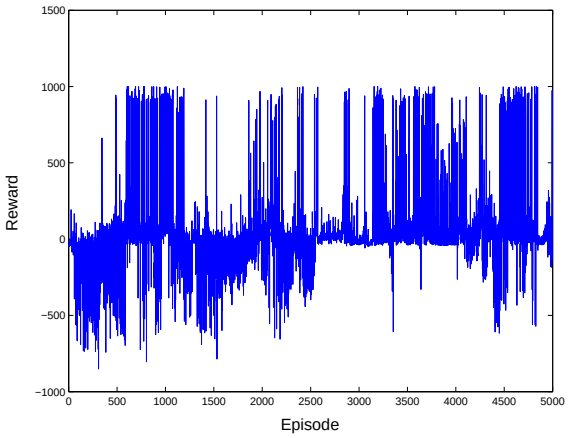


Figure 6: Progress of learning using a MLP network containing 15 hidden units and trained using the vario-eta algorithm with a global learning rate of 0.0001 for 5000 episodes. Performance is measured in terms of sum total reward during a 20 second episode, with $dt = 0.02$ the maximum reward per episode was 1000. Performance appears to improve initially but soon becomes unstable and does not appear to converge.

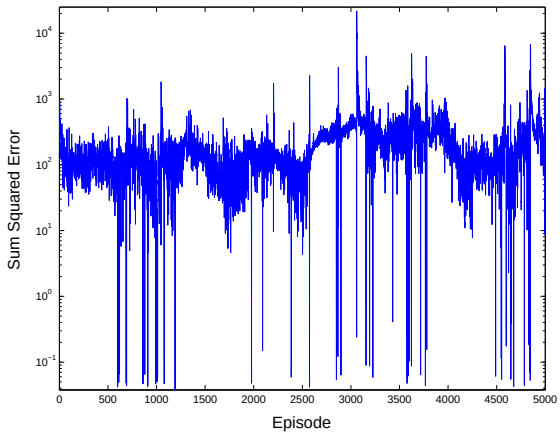


Figure 7: The sum squared TD error plotted against the episode number for a MLP network containing 15 hidden units and trained using the vario-eta algorithm with a global learning rate of 0.0001. Performance is unstable and does not converge.

3.3 Solution 3: A *pseudopattern* rehearsal strategy

If we want *pseudopattern* rehearsal strategies to work effectively, it is important to use *pseudoinputs* that resemble draws from the input distribution. In order to crudely approximate the input distribution we discretised the input space into 20x20 bins and kept a record of when the input was in each of the 400 regions of input space. At each time step the current TD error was used to set the value stored in the currently occupied bin using $BIN(\theta, \omega) = \min(100, \frac{1}{\delta^2})$ and all bins were then decremented using $BIN = 0.997 * BIN$. This update scheme results in a biased approximation of the input distribution, where the bias is toward more recent inputs for which the TD error was small. The values in the bins were normalised to sum to one, so that each bin value represented the probability of choosing to sample from it and samples were then made based on these probabilities.

Twenty five *pseudoinputs* were generated at each iteration using the above sampling procedure with the added restriction that any *pseudoinputs* that were within a small radius of the current training point were rejected. The minimum squared distance between the current input and a *pseudoinput* was 0.03 in the normalised input space, where normalisation involved a linear scaling of the input into the range [-1,+1]. Since the gradient of the squared error with respect to the network weights is zero for the

pseudopatterns, we instead used the Hessian matrix to incorporate the information contained in the *pseudopatterns* into the training process. We used the inverse of the Hessian to scale the gradient information that was used for parameter updates. Thus, providing better conditioning of the problem at the same time as reducing interference.

A MLP with 15 hidden units was trained using the continuous $TD(\lambda)$ approach. Various different global learning rates were investigated in order to determine an approximate optimum value. Since performance was much better than in the previous two experiments the performance was assessed by measuring the total accumulated reward following 100 episodes of learning.

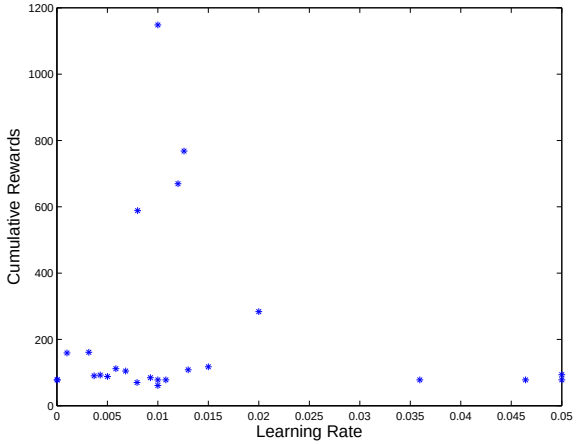


Figure 8: Comparison of different learning rates for a MLP network with 15 hidden units trained using a *pseudopattern* rehearsal strategy to approximate the value function for the task of swinging up a pendulum with limited torque. Performance is measured in terms of sum total reward following 100 episodes with a cumulative reward of 100000 representing the maximum possible. Performance was somewhat inconsistent, but the best performance was achieved for a learning rate of 0.01

Experiment 3: Results

Figure 8 shows an attempt to determine an optimal learning rate for the *pseudopattern* rehearsal strategy. Performance was somewhat inconsistent, but the best performance was achieved for a learning rate of 0.01. In other longer runs, that are not shown here, the learning process always converged eventually with a learning rate of 0.01 and suggests that the initial conditions can have a major effect on performance. In the original work by Doya (2000) using normalised Gaussian networks to approximate the value function, it is easy to initialise the value function to zero everywhere to encourage learning to focus on any *good* surprises that are experienced. In fact, convergence can take much longer if the value function is not initialised in this way. For our *pseudopattern* approach it is less clear how to initialise the learning system best. If we make all of the second layer weights small then the value function will have a value close to zero as we require. However, once we start training, the value function will quickly be pulled away from zero where it is not actively maintained by the *pseudopattern* rehearsal process. Figure 9 shows the course of learning on one of the more successful runs. There is a rapid improvement in performance following the first few successful episodes. This same pattern was observed in most of the runs that were examined, although the time it took to achieve the first success varied. This explains the high variance in performance we saw in our search for an optimal learning rate. Figure 10 shows that the squared TD error is decreasing as learning progresses, consistent with the convergence of the reinforcement learning process.

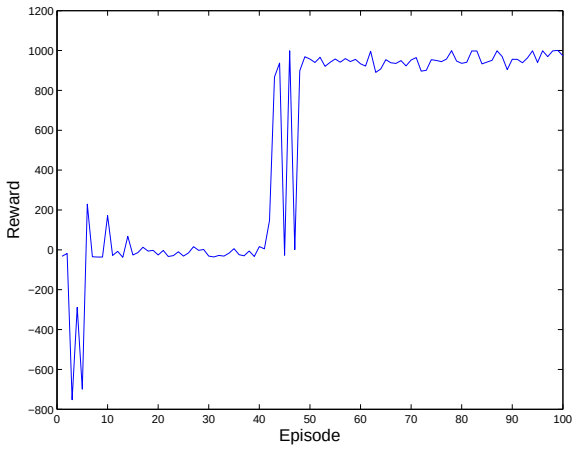


Figure 9: The course of learning using a *pseudopattern* rehearsal strategy and a learning rate of 0.01. There is a rapid improvement in performance following the first few successful episodes and thereafter performance remains close to optimal.

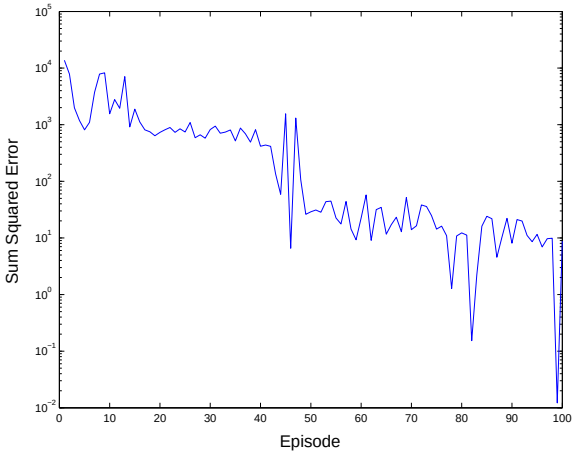


Figure 10: The sum squared TD error plotted against the episode number for a MLP network containing 15 hidden units and trained using a *pseudopattern* rehearsal strategy with a global learning rate of 0.01.

4 Conclusions

Our main results are as follows:

- We were able to learn a compact value function using a MLP.
- Training of the MLP using simple gradient descent is very slow and requires low learning rates.
- Using the vario-eta algorithm causes instability of the learning process.
- But, a *pseudopattern* rehearsal strategy greatly improved performance.

These results show that the continuous $TD(\lambda)$ can be successfully applied using non-linear, distributed function approximators such as MLP's. The final result suggests that interference rather than ill-conditioning presents a greater handicap to the learning of a value function. In fact, interference is likely to be exacerbated by using vario-eta. If we consider what the vario-eta algorithm does, then we can see why. Since it works by scaling individual learning rates by an estimate of their variance, weights which do not have a strong influence on the $current$ prediction, and are consequently changing little, will have their learning rates increased. This will result in larger weight changes for these weights, for only a small decrease in the $current$ prediction error. This is a good recipe for interference. The fact that Coulom achieved some success was possibly due to the fact that following initial trials he made the observation that all weights fell into one of two groups. All hidden to output weights' variances were of a similar magnitude and all remaining weights' variances were also of a different but similar value. He therefore choose to scale his learning rates according to these fixed ratios and not, in fact, use the full vario-eta algorithm.

If we compare the performance using a MLP with the performance achieved by Doya (2000) in the original work, where learning took only 20 or so episodes to converge, it is apparent that for this particular task local linear models are indeed far better suited to the task of learning a value function than a MLP network is. Can we conceive of a situation where this is not the case?

Local linear models can learn fast and are less susceptible to interference but suffer from the curse of dimensionality and deal poorly with irrelevant inputs. MLP networks are compact but require greater amounts of careful training. Therefore it is possible that there exist high dimensional reinforcement learning tasks, where data is plentiful, and speed of learning not critical, for which a MLP network might prove better suited than a local linear model. Coulom's swimmer task, (Coulom, 2002), may be one such problem. With 12 state variables and 5 control variables it would be difficult to cover the entire input space efficiently with local linear models. It is likely that this problem could be finessed by using a constructive incremental approach to local model creation and placement. However it would be interesting to see if it the problems of stability and convergence that Coulom experienced could be removed simply by resorting to standard gradient descent as opposed to employing vario-eta or some variant thereof.

Reinforcement learning in high dimensional continuous systems is always going to be a challenging problem. This suggests that from an engineering perspective we should do all we can to reduce the dimensionality of the problem as much as possible before performing reinforcement learning. For example, if we want an autonomous mobile agent to learn where in its environment good things and bad things are, we could in theory use the raw sensor input as input into a reinforcement learning process. However, more sensible would be to pre-process the raw sensor data to form some spatial, 2D, representation of the environment and then perform reinforcement learning in this lower dimensional space. If we can do this it would seem sensible to always employ local linear models. If not, then we can use MLP networks, and

rehearsal strategies or slow learning rates, to tackle higher-dimensional reinforcement learning problems.

In conclusion, we showed that interference rather than ill-conditioning was a greater handicap to learning, when using a distributed function approximator to learn a value function in a continuous reinforcement learning task. An attempt to improve learning, by addressing the problem of ill-conditioning using vario-eta, resulted the learning process failing to converge. In contrast a *pseudopattern* rehearsal strategy greatly improved learning performance in this task.

References

S Amari. Natural gradient works efficiently in learning. *Neural Computation*, 10, 1998.

C W Anderson. *Strategy learning with multilayer connectionist representations*, pages 103–114. Proceedings of the Fourth International Workshop on Machine Learning. Morgan Kaufmann, Irvine, CA, 1987.

B Ans and S Rousset. Avoiding catastrophic forgetting by coupling two reverberating neural networks. *Academie des Sciences, Sciences de la vie*, 320:989–997, 1997.

L C Baird. *Advantage updating. Technical Report WL-TR-93-1146*. Wright Laboratory, Wright Patterson Air Force Base, OH 45433-7301, USA, 1993.

A R Barron. *Universal approximation bounds for superposition of a sigmoidal function*, pages 930–945. IEEE Transactions on Information Theory. 39(3) edition, 1993.

R Bellman, I Glicksberg, and O Gross. On the bang-bang control problem. *Quarterly of Applied Mathematics*, 14(1):11–18, 1956.

D P Bertsekas and J N Tsitsiklis. *Neuro-Dynamic Programming*. Athena Scientific, 1996.

C M Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, Oxford, 1995.

R Coulom. *PhD thesis: Reinforcement Learning Using Neural Networks, with Applications to Motor Control*. Institut National Polytechnique de Grenoble, Grenoble, France, 2002.

K Doya. Reinforcement learning in continuous time and space. *Neural Computation*, 12(1):219–245, 2000.

K Doya. *Temporal difference learning in continuous time and space*, pages 1073–1079. DS Touretzky, MC Mozer and ME Hasselmo (eds) Advances in Neural Information Processing Systems 8. MIT Press, 1996.

R M French. Pseudo-recurrent connectionist networks: An approach to the "′sensitivity-stability"′ dilemma. *Connection Science*, 9:353–379, 1997.

G J Gordon. *Stable function approximation in dynamic programming*. A Prieditis and S Russel (eds) Machine Learning: Proceedings of the Twelfth International Conference. Morgan Kaufmann, San Franscisco, 1995.

L P Kaelbling, M L Littman, and A W Moore. Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, 4: 237–285, 1996.

K Levenberg. A method for the solution of certain non-linear problems in least squares. *Quarterly Jornal of Applied Mathematics*, II(2): 164–168, 1944.

D W Marquardt. An alogorithm for least squares estimation of non-linear parameters. *Journal of the Society of Industrial and Applied Mathematics*, 11(2):431–441, 1963.

M F Moller. A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks*, 6: 525–533, 1993.

R Neuneier and H-G Zimmermann. *How to train neural networks*. G.B. Orr and K-R Müller (eds) Neural Networks: Tricks of the trade. Springer, 1998.

J K Peterson. *On-line estimation of the optimal value function: HJB-estimates*, pages 319–326. CL Giles and SJ Hanson and JD Cowan (eds) Advances in Neural Information Processing Systems 5. Morgan Kaufmann, San Mateo,CA,USA, 1993.

M Riedmiller and H Braun. *A direct adaptive method for faster back-propagation learning: The RPROP algorithm*. Proceedings of the IEEE International Conference on Neural Networks. 1993.

A Robins. Catastrophic forgetting, rehearsal, and pseudorehearsal. *Connection Science*, 7:123–146, 1995.

R S Sutton. Learning to predict by methods of temporal differences. *Machine Learning*, 3:9–44, 1988.

R S Sutton and A G Barto. *Reinforcement learning: An Introduction*. MIT Press, 1998.

R S Sutton, D McAllester, S Singh, and Y Mansour. *Policy Gradient Methods for Reinforcement Learning with Function Approximation*, pages 1057–1063. Advances in Neural Information Processing Systems 12. MIT Press, 2000.

G Tesauro. Temporal difference learning and TD-G$ammon$. *Communications of the ACM*, 38(3):58–68, 1995.

Short-term memory traces for action bias in human reinforcement learning

Rafal Bogacz
Department of Computer Science
University of Bristol
R.Bogacz@bristol.ac.uk

Samuel M. McClure
Center for the Study of Brain, Mind and Behavior
Princeton University
smcclure@priceton.edu

Jian Li
Department of Neuroscience, Human Neuroimaging Lab
Baylor College of Medicine
jli@hnl.bcm.tmc.edu

Jonathan D. Cohen
Center for the Study of Brain, Mind and Behavior
Princeton University
jdc@priceton.edu

P. Read Montague
Department of Neuroscience, Human Neuroimaging Lab
Baylor College of Medicine
read@bcm.tmc.edu

Abstract

Recent experimental and theoretical work on reinforcement learning sheds light on the neural bases of learning from rewards and punishments. One fundamental problem in reinforcement learning is the credit assignment problem, or, how to properly assign credit to actions that lead to reward or punishment following a delay. Temporal difference learning solves this problem, but its efficiency is significantly improved by the addition of eligibility traces (ET). In essence, ETs function as decaying memories of previous choices that are used to scale synaptic weight changes. It has been shown in theoretical studies that ETs which span a number of actions may improve the performance of reinforcement learning. However, to our knowledge, reinforcement learning models incorporating ETs persisting over a number of actions have not been tested in the behaviour of biological organisms, including humans. This paper reports such a study. We report an experiment in which human subjects performed a sequential economic decision game in which the long-term optimal strategy is different from the strategy that leads to the greatest short-term return. We demonstrate that human subjects' performance on the task is significantly affected by the time between choices in a surprising and seemingly counterintuitive way. However, this behaviour is naturally explained by a temporal difference learning model with ETs persisting across actions. Furthermore, we demonstrate that recent accounts of short-term synaptic plasticity in dopamine neurons may provide a realistic biophysical mechanism for producing ETs that persist on a timescale consistent with behavioural observations.

Embodied learning :
Investigating stable Hebbian learning in a spiking neural network

Daniel Bush* d.bush@sussex.ac.uk
Andrew Philippides andrewop@sussex.ac.uk
Phil Husbands philh@sussex.ac.uk
Michael O'Shea M.O-Shea@sussex.ac.uk

Centre for Computational Neuroscience and Robotics
University of Sussex,
Brighton, BN1 9QG

*Corresponding author

Abstract

The idea that synaptic plasticity holds the key to the neural basis of learning and memory is now widely accepted in neuroscience. The precise mechanism of changes in synaptic strength has, however, remained elusive. Neurobiological research has led to the postulation of many models of plasticity, and among the most contemporary are spike-timing dependent plasticity (STDP) and long-term potentiation (LTP). The STDP model is based on the observation of single, distinct pairs of pre- and post- synaptic spikes, but it is less clear how it evolves dynamically under the input of long trains of spikes, which characterise normal brain activity. This research explores the emergent properties of a spiking artificial neural network which incorporates both STDP and LTP. The direction of future work, which will include the addition of a volume signalling element based on the postulated actions of nitric oxide in neural learning mechanisms, is then outlined.

1 Introduction

The ability of the brain to translate ephemeral experience into enduring memories has long been attributed by neuroscientists to activity-dependent changes in synaptic efficacy. One of the first to suggest a mechanism that could govern this plasticity was Donald Hebb, who hypothesised that 'when an axon of cell A is near enough to excite a cell B, and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place … such that A's efficiency as one of the cells firing B, is increased' (Hebb, 1949). This concept of 'Hebbian' learning has become a mainstay of neural theories of memory, but more precise rules of synaptic change have been difficult to elucidate.

It has become clear, however, that there are certain features which are crucial to a successful model of plasticity (Roberts and Bell, 2002 ; Song, Miller and Abbott, 2000 ; van Rossum, Bi and Turrigiano, 2000). It must generate a stable distribution of synaptic weights, and stimulate competition between inputs to a neuron, in order to account for the processes of activity-dependent development and forgetting, and to maximize the capacity for information storage (Miller, 1996). Pure Hebbian learning cannot achieve this, because those inputs which correlate with post-synaptic firing are repeatedly strengthened, and grow to infinitely high values, while those which do not are persistently weakened. This creates an inherently unstable, bimodal distribution of synaptic weights. Earlier plasticity models have had to resort to a variety of means in order to solve this problem. Often these promoted competition through the use of global signalling mechanisms, such as limiting the sum of strengths of pre-synaptic inputs to a cell, but the biophysical realism of such protocols can be questioned. The exact nature of the additional constraints used can also strongly influence the behaviour of the model (Miller and McKay, 1994).

In considering the neural basis of memory, it is long-lasting alterations in synaptic strength that are of most interest. Experimental evidence for such changes was first found in the hippocampus – a region of the brain long identified with learning – when it was shown that repeated activation of excitatory synapses by high frequency spike trains caused an increase in synaptic strength which lasted for hours, or even days (Lomo and Bliss, 1973). This phenomena - known as long-term potentiation (LTP) - has since been the subject of a great deal of investigation, because it exhibits several features which make it an attractive candidate as a neural learning mechanism. It is synapse specific, vastly increasing the potential storage capacity of individual neurons. It is also associative, in that the repeated stimulation of one set of synapses can simultaneously facilitate LTP at adjacent sets of synapses. This has often been viewed as analogous to the process of classical conditioning.

The processes which trigger LTP are relatively well understood, but experimental limitations have made the biological mechanisms underlying it's expression difficult to clarify (See Malenka and Nicol, 1999, for a review). Contrasting research findings have exacerbated this problem and provoked a great deal of debate. The locus of expression of plasticity is one issue that has been particularly controversial. It is generally accepted that post-synaptic changes, such as an increase in the number or function of AMPA receptors, occur, but debate surrounds the possibility that there are also changes in pre-synaptic operation. Some evidence has suggested that LTP is at least partially mediated by an increase in neurotransmitter release, but it is far from conclusive. If plasticity is expressed at some level on both sides of the synaptic cleft, however, then some signal must pass from post- to pre- synapse, carrying the message that LTP has been induced. Many candidates have been proposed for this role - among them the diffusible, gaseous neuromodulator nitric oxide (Arancio et al., 1996).

The wealth of research into LTP has helped to inform and inspire new plasticity models which are more easily reconcilable with the tenets outlined earlier. The 'BCM' model, named after its creators (Bienenstock, Cooper and Munro, 1982) and based on their consideration of input selectivity in the visual cortex, is a good example. It is Hebbian, but achieves stability through the existence of a 'threshold' firing rate, a crossover point between depression and potentiation which is itself slowly modulated by post-synaptic activity. This makes the strengthening of a synapse more likely when average activity is low, and vice versa, thus generating competition between inputs.

Another contemporary plasticity model, based on the more straightforward empirical observation of distinct pairs of pre- and post- synaptic action potentials (Roberts and Bell, 2002 ; Bi and Poo, 1998), has also generated a great deal of interest. It is known as spike timing dependent plasticity (STDP), because it dictates that the direction and degree of changes in synaptic efficacy are determined by the relative timing of pre- and post-synaptic spiking. Only pre-synaptic spikes which provoke post-synaptic firing within a short temporal window potentiate a synapse, while those which arrive after post-synaptic firing cause depression. Those inputs with shorter latencies or strong mutual correlations are thus favoured, at the expense of others.

The most pertinent feature of STDP is that it implicitly generates competition between synapses,

and experiments with artificial neural networks (ANNs) have shown that this consequently generates inherently stable weight distributions. The shape of the resulting distribution is dependent on the exact nature of the STDP implementation, and the values of parameters used. Some researchers, for example, include the experimental observation that stronger synapses seem to undergo relatively less potentiation than weaker synapses, or an activity dependent scaling mechanism such as that outlined by the BCM model (van Rossum, Bi and Turrigiano, 2000). These features help to generate a weight distribution that more closely resembles the stable, unimodal, and positively skewed distribution found *in vivo* (see *fig 1.2*). Their omission tends to produce a bimodal distribution (Song, Miller and Abbott, 2000 ; Iglesias et al. 2005) more similar to that produced by purely Hebbian learning, but stabilised by innate competition and the inclusion of hard limits on the maximum achievable strength of a synapse (see *fig 1.1*).

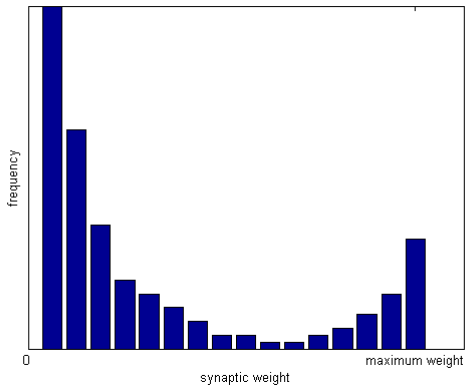


Fig 1.1 – Typical weight distribution generated by STDP

The analysis of STDP is based on isolated pairs of pre- and post- synaptic action potentials, while observations of LTP are mediated by the application of prolonged spike trains more characteristic of normal brain activity. It is not clear how the STDP model causes synaptic weights to develop dynamically with such input, which involves many possible spike pairings. We can presume that both forms of plasticity arise from the same underlying biophysical mechanisms, and some recent work has attempted to reconcile both models within a single theoretical framework (Izhikevich and Desai, 2003). By making a few biologically plausible assumptions, this research has demonstrated that the parameters of STDP can be linked directly with the sliding threshold of the BCM model.

This paper explores the emergent properties of an artificial neural network which implements spike timing dependent plasticity. The form of STDP used is compatible with the BCM model of long-term potentiation, and thus the threshold firing rate can effectively be manipulated. The effects this has on synaptic weight distributions and dynamics are examined. Size-dependent potentiation is introduced, in order to observe the effects that this has on the behaviour of the network. Results obtained from the input of random, uncorrelated spike trains are also compared with those generated by input taken from performance of a simple, embodied, sensorimotor task. The latter will have correlated temporal patterns that are perhaps more representative of firing regimes found *in vivo*, and which STDP has previously been shown to make use of (Izhikevich, Gally and Edelman, 2004).

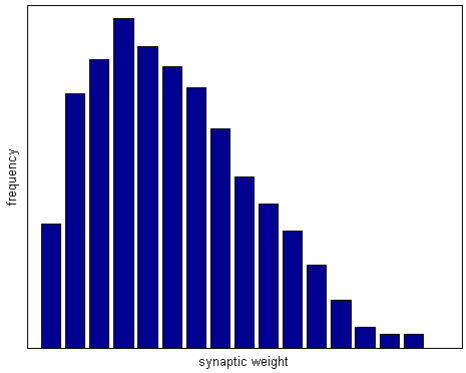


Fig 1.2 – Synaptic weight distribution found in vivo

Previous research findings are replicated in most instances, and some interesting additional properties of the network also noted. Along with a discussion of these preliminary results, the direction of, and motivation for, future work is outlined. As well as elaborating on the interesting findings made so far, this will include the use of additional robotics tasks, in order to further assess how the nature of synaptic input affects resulting weight distributions; and genetic algorithms, to optimise the performance of the network on these tasks. Hopefully this will help to identify which features of the network and plasticity model are most important to simple learning behaviour. A volume signalling element which can dynamically modulate STDP parameters will also be introduced, to investigate the postulated role of a retrograde messenger in LTP. This will build on the work of Gally et al. (1990) and Husbands et al. (1998), who developed connectionist neural networks which incorporated abstract models of a diffusing, neuromodulatory gas.

2 Methods

2.1 Neural Controller

The neural network consists of *20* neurons, which are divided into *9* sensory, *9* intermediate and *2* motor neurons. The network is realistic of the mammalian cortex in that these are *80%* excitatory and *20%* inhibitory, and that each has a randomly chosen axonal delay in the range *[1ms, 20ms]*. Each neuron has *5* randomly assigned post-synaptic connections. Motor neurons have no post-synaptic connections, and sensory neurons have no pre-synaptic connections. Initial synaptic weights are assigned, according to the nature of the test, between *0* and a maximum (for excitatory neurones) or minimum (for inhibitory neurones) weight, which vary between tests. *Fig 2.1* below illustrates the network morphology, with neurons represented as nodes, and typical post-synaptic connections for three neurons shown as straight lines. In the real network, the total number of connections (and thus the total number of synapses) is *90*, which corresponds to *5* post-synaptic connections per neuron, excluding the two motor neurons.

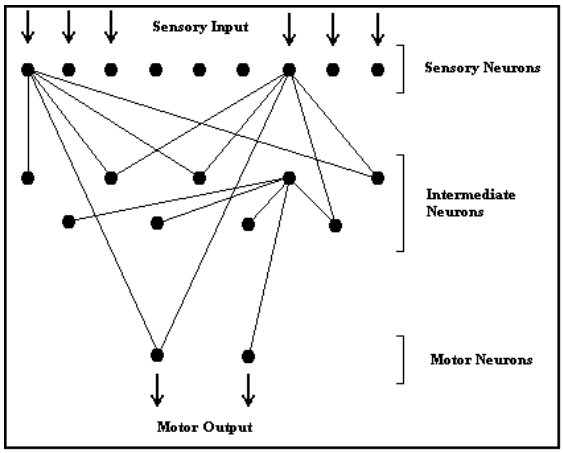


Fig 2.1

The neurons operate using the Izhikevich (2004) spiking model, which dynamically calculates the membrane potential (v) and a membrane recovery variable (u), based on the values of four constants (a,b,c and d)and an applied current (I) according to the equations below.

$$v' = 0.04v^2 + 5v + 140 - u + I$$
$$u' = a(bv - u)$$
$$\text{if } v \geq +30\,\text{mV} \text{ then } \begin{cases} v \leftarrow c \\ u \leftarrow u + d \end{cases}$$

Eqn. 2.1

This model was chosen for three reasons. Firstly, it uses very few floating point operations, and so is computationally advantageous. Secondly, it can exhibit firing patterns of all known types of cortical neurons, by varying the parameters *a, b, c* and *d*. The values used for a standard excitatory neuron are *[0.02,0.2,-65,6]* respectively, and those for an inhibitory neuron are *[0.02,0.25,-65,2]*. It is also one of the most contemporary spiking models available.

In order to introduce neural noise into the system, one neuron is selected at random each time step, and a small current applied to it. A value of *10mA* was used in most tests, although this was varied to assess the effects of neural noise. When distributed randomly over *20* neurons, an applied current of *10mA* produces a spiking rate of approximately *3Hz* per neuron.

2.2 STDP

Mathematically, with $s = t_{post} - t_{pre}$ being the time difference between pre- and post- synaptic spiking, the change in the weight of a synapse (Δw) due to spike timing dependent plasticity can be expressed as:-

$$\Delta w = \begin{cases} A^+ \exp(-s/\tau^+) & \text{if } s > 0 \\ A^- \exp(s/\tau^-) & \text{if } s < 0 \end{cases}$$

Eqn. 2.2

The method of implementing this plasticity is similar to that used by Song et al (2000) and others. Two recording functions (P_+ and P_-) are kept for each synapse. Whenever a spike arrives at a synapse, the relevant P_+ value is reset to the value of the constant A_+, and whenever a post-synaptic spike is fired the relevant P_- values are decreased to the constant A_-. In the absence of any spikes, these functions decay exponentially with the time constants τ_+ and τ_-. Concurrently, whenever a spike arrives at a synapse, P_- is used to decrease the synaptic weight, and whenever the post-synaptic neuron fires, P_+ is used to increase the synaptic weight, as shown below.

$$w_{ij}(t) = w_{ij}(t) + P_+ e^{-kw_{ij}}$$

Eqn. 2.3

Research aimed at reconciling the BCM model of long-term potentiation and STDP (Izhikevich and Desai, 2003) dictates that only nearest neighbour pairs of spikes should be used to direct the plasticity of a synapse. It also allows the calculation of a value for the threshold firing rate, using the formula below.

$$v = -\frac{A_+/\tau_- + A_-/\tau_+}{A_+ + A_-}$$

Eqn 2.4

This in turn means that the expressions $A_+ > |A_-|$ and $|A_- \tau_-| > |A_+ \tau_+|$ should be satisfied, to ensure that the threshold has a positive value at all times. The values of these STDP parameters that were used to generate each set of results are given in section 3.

Size-dependent potentiation was also introduced into the plasticity model in some experiments. Previous research (Bi and Poo, 1998) has shown that the relationship between the level of potentiation and initial synaptic weight is most likely inverse exponential, but a linear relationship was also tested. The formulae governing increases in synaptic weight in these experiments are given below.

$$w_{ij}(t) = w_{ij}(t) + P_+ e^{-kw_{ij}}$$

Eqn. 2.5

2.3 Task

The network was first examined with uncorrelated Poissonian spike trains of varying frequencies as input. In later experiments, a simple robotics task was used to assess the behaviour of the network with input that had more temporal correlation and widely varying spike frequencies. The task chosen was the falling block task, employed previously by Goldenberg et al (2004). An agent of radius *r=15* moves horizontally in an arena which is *400* units wide. The agent has *9* sensory neurons with a range of *205*, which are distributed evenly over a visual angle of $\pi/6$. These sensory neurons each have a randomly determined bias in the range *[0.6:1.0]* which is used to scale an applied current, relative to the distance of any object in their direct line of vision. The agent in it's environment is illustrated by *fig 2.2* below.

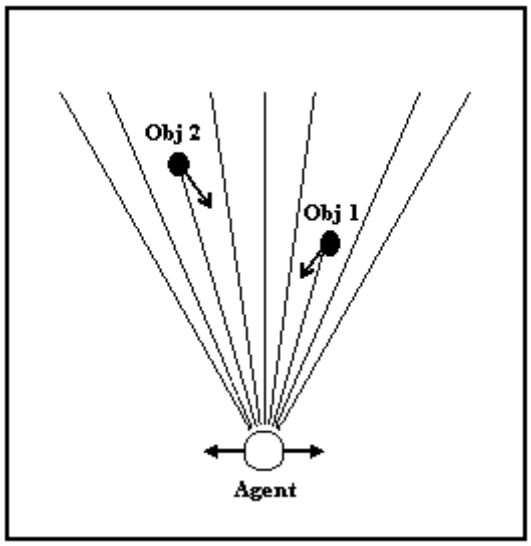


Fig. 2.2

Two blocks of radius *r=13* fall from a height of *198* at randomly assigned angles and from randomly assigned horizontal start positions, constrained only by the criteria that it must be possible for the agent to catch them both. The first object has a random velocity in the range *v=[0.03:0.04]* and the second object in the range *v=[0.01:0.02]*. The agent's horizontal velocity is determined by the sum of the two opposing motors outputs, its maximum velocity being set at *0.05* units/ms. The two motor neurons are leaky integrators, operating according to the equation below, where t^o is the time at which a spike was last received. Each has a randomly assigned gain in the range *[0.01:0.05]* and a decay constant (τ) in the range *[20ms : 40ms]*.

$$v = v\, e^{-(t - t^o)/\tau}$$

Eqn. 2.6

It is important to note that the capacity of the network to learn how to perform this task is not being tested in this paper. The embodiment is needed only to provide realistic sensorimotor input, which has correlated temporal properties that are considered important in assessing the properties of the plasticity model and network. In the future, evolutionary robotics techniques will be used to assess the learning capabilities of the network on this, and other, simple tasks. An outline of the motivations for this is given in section 4.

2.4 Stability

After each *100ms* of experimental time, a histogram of synaptic weights is generated. If the values in each bin (which are of size 1) do not vary by more than ±1 for 10 of the *100ms* steps (i.e. 1 second), then the network is considered to have achieved a stable synaptic weight distribution. In order to test that this criteria was adequate, 30 tests were performed in which the network continued to operate for 100 seconds of simulated time after stability was flagged. In all cases, no further discernible change in the synaptic weight distribution occurred.

3 Results and Discussion

3.1 Manipulation of threshold firing rate

Figure *3.1* represents a standard synaptic weight distribution generated when the network was operated with purely uncorrelated input at a rate of *30Hz*, and the results replicate previous research findings (Song et al., 2000). The values of STDP parameters used in this case correspond to a threshold firing rate of approximately *ν=17Hz*, and intermediate excitatory neuron firing rates had a mean value of approximately *12Hz*. The effects of moving the threshold firing rate (by varying any of the four main STDP parameters) are intuitive, and demonstrated by figures *3.2* (where *ν=350Hz*) and *3.3* (*ν=6.25 Hz*) below. A higher threshold for long-term potentiation allows fewer synapses to reach the maximum possible strength, and a lower threshold has the reverse effect.

However, results suggest that the relationship between weight distribution and STDP parameters is dictated by more complex factors than simply the position of the BCM threshold. Figure *3.4* shows a weight distribution for an identical threshold firing rate as *3.1*, but with different STDP values *(ν=17 Hz)*. The number of synapses which have been potentiated to saturation are fewer, and those which have been persistently depressed larger in frequency. The value of $A_+ \tau_+$ is identical in both cases, but the longer temporal window for potentiation that existed in *3.4* clearly had a lower overall strengthening effect on weight values, compared with the higher degree of synaptic strengthening per spike which was present in the results for *3.1*. Figure *3.5* represents a further manipulation of STDP parameters which again correspond to a threshold rate of approximately *ν=17Hz*. The ratio of $A_+ : A_-$ is equal in these two cases, and the distributions generated are almost identical.

These results suggest that the position of the modification threshold may serve as a guide to the approximate shape of the distribution – or at least, the relative frequency of synapses which have been fully strengthened or weakened – but variations of STDP parameters have more of an influence than merely determining the position of this threshold.

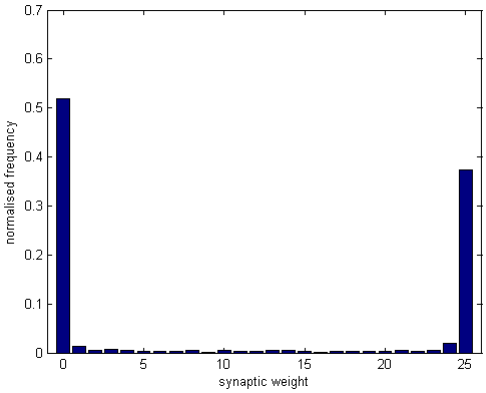


Fig 3.1 - A_+=0.16 ; A_-=-0.1 ; τ_+ = 20ms ; τ_- = 40ms

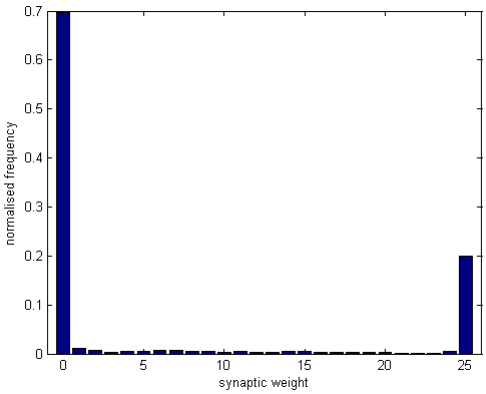


Fig 3.2 - A_+=0.12 ; A_-=-0.1 ; τ_+ = 10ms ; τ_- = 40ms

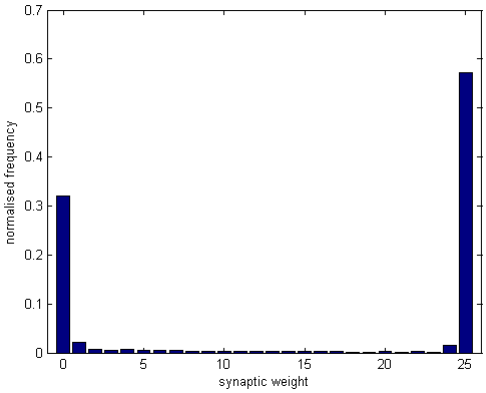


Fig 3.3 - A_+=0.18 ; A_-=-0.1 ; τ_+ = 20ms ; τ_- = 40ms

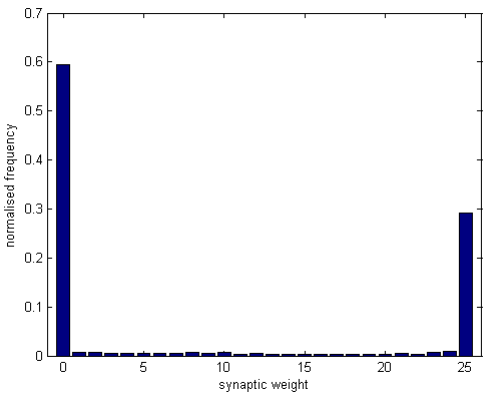


Fig 3.4 - A_+=0.12 ; A_-=-0.1 ; τ_+ = 30ms ; τ_- = 40ms

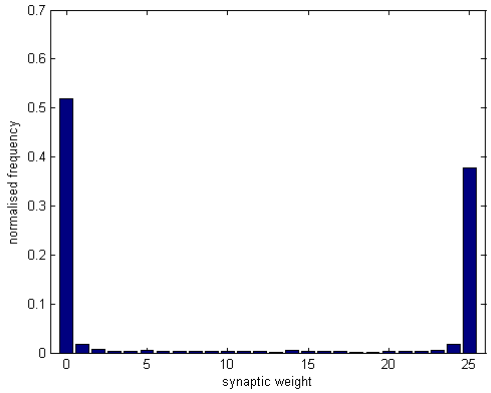


Fig 3.5 - A_+=0.2 ; A_-=-0.125 ; τ_+ = 20ms ; τ_- = 40ms

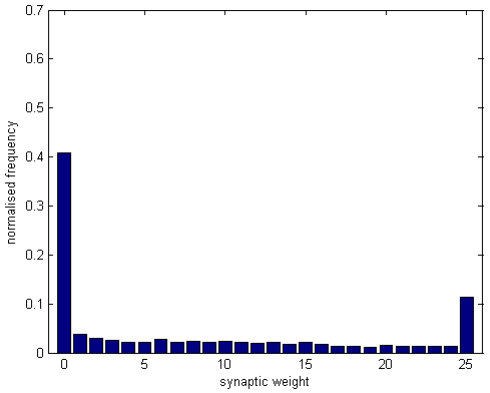


Fig 3.6

3.2 Varying network input

It is useful to make a comparison between the weight distributions arising from uncorrelated input and those generated by input from a closed-loop sensorimotor task. Figures *3.6* and *3.7* illustrate the results from the robotics exercise, when identical parameter values to figures *3.1* and *3.3* respectively were used. It seems that the input in the simple robotics task causes more synapses to adopt intermediate weight values, rather than be pushed to the bounds, and that changes in STDP parameter values have a much smaller effect. Distributions generated by input from the robotics task are generally much more consistent in shape. The effects of manipulating the threshold rate can still be seen, but rather than simply altering the size of the bimodal peaks (as seen in figures *3.1 – 3.3*) it is the frequency and distribution of the intermediate strength synapses that are most affected.

These discrepancies support the intuitive hypothesis that the nature of input to an ANN has a pronounced effect on the evolution of synaptic weights in that network. Much of the previous research in this area has made exclusive use of uncorrelated input, but results found here show that care must be taken in generalising from these findings. The resultant effects of any plasticity model are at least partially defined by the nature of the input it receives.

The particular differences in this case could be explained by the nature of the simple robotics task employed. The activity of the sensory neurons during the task is characterised by relatively short bursts of spiking, interspersed with periods of relative silence, as objects move in and out of the line of sight. As post-synaptic activity is continuously present, the synaptic weights of the sensory neurons which are currently silent will be slowly depressed. This could account for the overall smaller number of very strong synapses, as it is impossible for all sensory neurons to be active simultaneously, and so depression will be perpetually incurred in some of the input neurons.

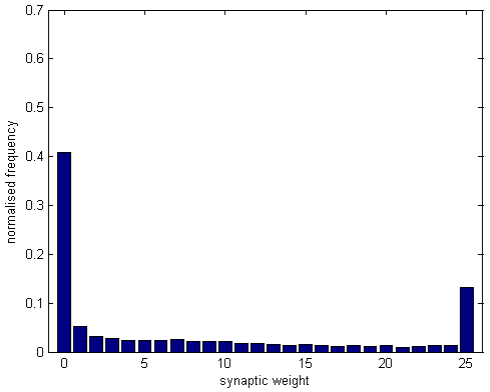


Fig 3.7

3.3 Firing rates

Various firing rates of uncorrelated Poissonian input were used during the course of the investigation. An analysis of the effects on the post-synaptic firing rate (that in the intermediate, excitatory neurons) led to a surprising finding which contradicts previous research (Song, Miller and Abbott, 2000). It was observed that increasing the rate of input action potentials precipitated a small but noticeable decrease in their post-synaptic firing rate, as illustrated by figure *3.8*.

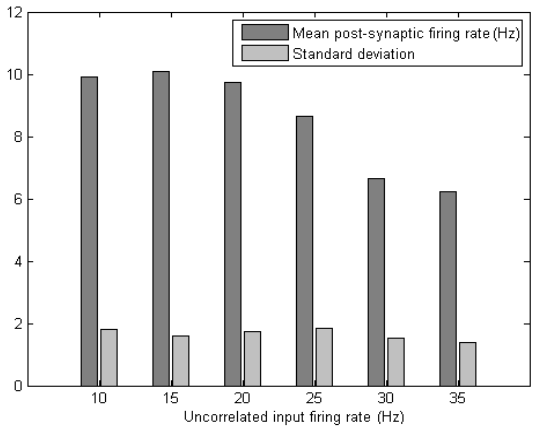


Fig 3.8

It is clear that the key to a good plasticity model, and one of the reasons why STDP is so highly regarded, is that it regulates network output in the face of wide fluctuations in input. However, one would expect this to be merely a damping effect, reducing the natural increase in output firing rate, rather than an inverse relationship such as the one generated by our network. This finding is also at odds with the notion of the sliding modification threshold of the BCM model. All STDP parameters were maintained at identical values when input firing rates were varied, and thus the position of the modification threshold also remained at a fixed point (*v=100Hz* in this case). An increase in input firing rate, one would then assume, would lead to an increase in the number of synapses which were persistently potentiated.

In previous research, however, an increase in input firing rate with fixed STDP parameters has been observed to cause a slow decrease in the number of synapses saturating at the uppermost weight values (Song, Miller and Abbott, 2000). Interestingly, this finding was replicated, and figure *3.9* shows how the relative frequency of synapses saturating at the upper bound varies with the rate of input. One may expect that fewer strong synapses would correlate with lower post-synaptic activity (as further data analysis showed that the number of intermediate strength synapses remained roughly constant), and this is in effect what has been demonstrated by our network. However, the findings in this area, and the reasons why they are inconsistent with previous work, certainly warrant further investigation.

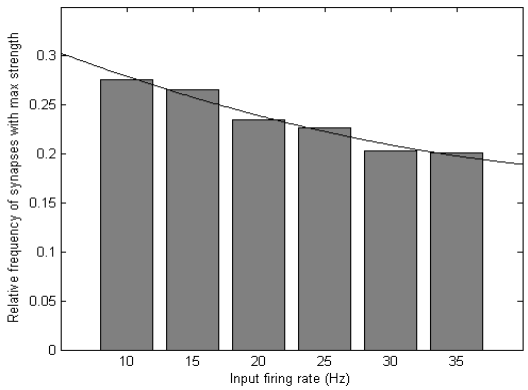


Fig 3.9

3.4 Size-dependent potentiation

The introduction of size-dependent potentiation into the plasticity model has a pronounced effect on synaptic weight distributions. Figures *4a* and *4b* (which were generated using *eqn. 2.5*, with a value of *k=50*) illustrate this, and more closely resemble results found *in vivo*. The peak at *w=0* has been omitted, as these 'silent' synapses are not considered (and cannot be detected) when biological appraisals of weight distributions are made. It is interesting to note, however, that the frequency of synapses found at the lower bound was generally consistent between experiments with and without size-dependent potentiation. This implies that the larger number of synapses adopting intermediate weights was simply a product of the fact that fewer synapses were able to saturate at the upper bounds. Figures *4a* (where *v=100Hz*) and *4b* (where *v=6.25Hz*) also demonstrate that the value of the threshold rate has much less of an effect on the distribution when size dependent potentiation is present, although a small increase in the size of the maximum weight peak can still be clearly seen.

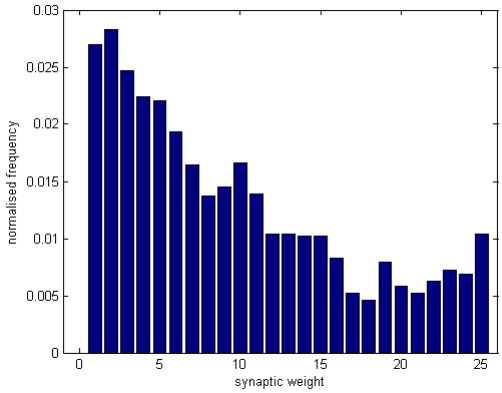


Fig 4a

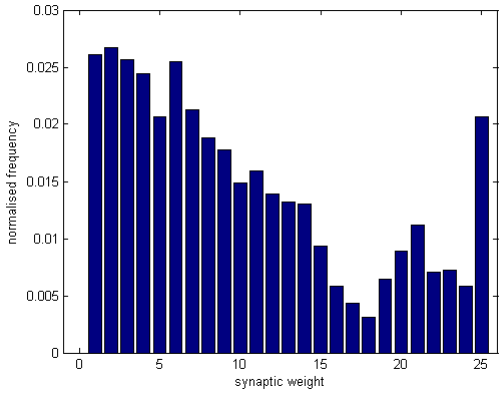


Fig 4b

3.5 Effect of initial weight values

The results obtained with size-dependent potentiation also demonstrated that the initial synaptic weight values have some considerable influence on the stable weight distribution. Figure *5a* illustrates the synaptic weights adopted by the network, using identical STDP parameters to figure *4a*, but with initial synaptic weights all being assigned at the maximum value, rather than uniformly distributed between 0 and w_{max}. The distribution in this case even more closely resembles that found *in vivo*. The higher values of initial weights seem to shift the modal peak of the

weight distribution to a higher value, also increasing average post-synaptic firing rates.

In the consideration of any dynamical system, it would be naïve to assume that initial conditions would not have a pronounced effect on the direction of evolution. In terms of learning, these findings could be explained, at a very abstract level, in that providing the network with any set of initial weights is equivalent to providing some form of 'hard wired' memory. When these initial weights are uniformly distributed, certain neural pathways are favoured over others from the outset. Plasticity will continue to make use of these synaptic connections at the expense of others, the network effectively 'building on what it knows' during development. When all initial weights are set equally, however, one would assume that the network morphology and input alone would dictate the direction of plasticity, as all neural pathways are balanced equally at the outset. These results certainly warrant further investigation, and it would seem sensible to do this by examining the dynamics of individual synapses. This should help to ascertain whether those that are originally assigned low weight values are capable of being significantly strengthened over time, or if their initial weight and the nature of the plasticity model make persistent depression inevitable.

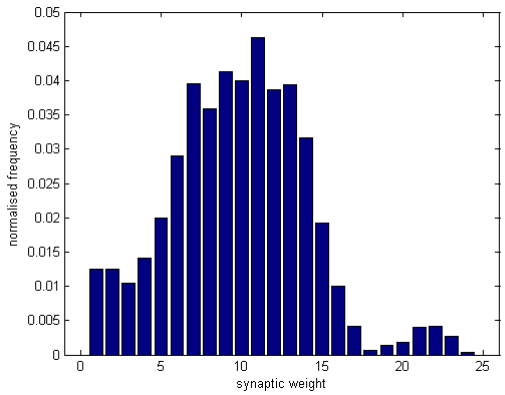


Fig 5a

3.6 Conclusions

The results obtained mostly support previous findings in this area. Manipulation of the BCM threshold firing rate directs synaptic weights in an intuitive manner. The STDP model has a strong regulatory effect on post-synaptic output, though surprisingly it seems to reduce the firing rate in the face of an increased frequency of input. The initial conditions of the network underlying the plasticity model, and the nature of input used, seem to have a pronounced effect on the direction in which it develops, which is to be expected from any dynamical system. Although STDP implicitly generates competition between synapses, the weight distribution it creates is still not representative of that found *in vivo* unless additional experimental observations are included. The use of size-dependent potentiation generates a weight distribution which closely resembles that which occurs naturally in the mammalian cortex. It is left to future work to see whether this, and other phenomena identified by this paper, are beneficial to simple learning behaviour in ANNs.

4 Plan for future work

The next step in this research is to pursue and elaborate on the interesting results which have arisen, and then to extend the scope of the work to allow further insight into the nature and properties of contemporary synaptic plasticity models. In terms of the former, a number of tests have naturally suggested themselves in the course of the research so far. A greater number of embodied robotics tasks should be used to glean any further insight possible into how various forms of input direct synaptic plasticity. The relationship between input and post-synaptic firing rate needs to be clarified, and the implications of this unexpected finding explored. A greater variety of initial synaptic weights should be tested, to further delineate the nature of their effect on the final distribution. There are certain properties of the network which it would also be interesting to experiment with. Timing is clearly critical in this plasticity model, and so the effects of varying or removing axonal delays may generate some interesting results. It is also intended that the morphology of the network be varied. This will include the use of re-entrant connections to more closely model the structure of the hippocampus, which is a region of the brain in which synaptic plasticity is most frequently observed.

It is also considered very important that, in the near future, evolutionary algorithms are used to optimise the performance of our network on the various robotics tasks used. This will build on the work of Floreano and Urzelai (2001), who explored the development of plastic control networks through evolutionary techniques, and help to assess the robustness of various incarnations of this network. It will also hopefully allow the identification of the exact properties of the plasticity model which are exploited to achieve simple learning behaviour. The adaptive abilities of such robots are clearly valuable.

After these examinations, it is felt that the addition of a volume signalling element to the network which reflects the operation of nitric oxide (NO) in the brain would be a useful elaboration. The debate over the function of this gas has provoked a wealth of research, which has made it clear that NO plays some role in neural learning processes, whether or not as a retrograde messenger (see Holscher, 1997 for a review). Experiments with

animals have shown that NO-synthase (NOS) inhibitors impair learning in chicks performing a passive-avoidance task (Holscher and Rose, 1993), and cause amnesia in rats in a water maze task (Chapman et al, 1992). NO has also been identified as a key intracellular messenger in imprinting and odour learning in sheep (Kendrick et al, 1997). In invertebrates, NO has been found to play a critical role in associative learning (Müller, 1996); in the formation of long term proboscis extension response in the honey bee (Hammer, 1997); in tactile learning in the octopus (Robertson et al, 1994); food attraction conditioning in the snail *Helix pomatia* (Teyke, 1996); and the consolidation of memory following appetitive conditioning of *Lymanaea* (Kemenes et al, 2002). In *Aplysia*, it has been implicated in multiple memory processes after learning that a food is inedible (Katzoff et al, 2002). It has also been shown that the gas facilitates LTD in the cerebellum (Schweighofer and Ferriol, 2000) - a neural correlate of motor learning. Simulation studies which model cerebellar learning (Calabresi et al, 1999) have also demonstrated that the inclusion of an abstract gas signalling mechanism enhances learning and improves performance. Studies on the rat hippocampus (Malen and Chapman, 1997) have led to an interesting hypothesis regarding a possible alternative role of NO in the neural correlates of learning. It is possible that NO signalling may mediate the BCM firing threshold, as findings show that NO donors facilitate LTP induction by stimuli that would normally only produce short term potentiation.

With the results obtained in this study, a volume signal could be used to directly modulate STDP variables, dynamically altering the position of the BCM model threshold firing rate, and thus indirectly manipulating individual synaptic weights online. The action of NO as a retrograde messenger, or in any other postulated role that it may have in neural learning processes, can thus be modelled and investigated. This will build on the work of Gally et al. (1990) and Husbands et al. (1998), who developed connectionist neural networks that incorporated abstract models of a diffusing, neuromodulatory gas. Hopefully, the use of evolutionary robotics techniques, in conjunction with a four-dimensional signalling element such as this, will allow the development of a new class of spiking ANN's which can more successfully replicate neural learning mechanisms.

Acknowledgements

The authors would like to thank early reviewers for their input, and for the general help and support of the members of the CCNR at Sussex University.

References

Wickliffe Abraham et al. Heterosynaptic metaplasticity in the hippocampus *in vivo*: A BCM-like modifiable threshold for LTP. *PNAS*, 98 (19): 10924-10929, 2001.

Ottavio Arancio et al. Nitric Oxide acts directly in the pre-synaptic neuron to produce long-term potentiation in cultured hippocampal neurons. *Cell,* 87: 1025 – 1035, 1996.

J.M. Bekkers, G.B. Richerson and C.F. Stevens. Origin of variability in quantal size in cultured hippocampal neurons and hippocampal slices. *PNAS,* 87: 5359-5362, 1990.

Guo-qiang Bi and Mu-ming Poo. Synaptic modifications in cultured hippocampal neurons : dependence on spike timing, synaptic strength and post-synaptic cell type. *Journal of Neuroscience,* 18: 10464 – 10472, 1998.

Elie Bienenstock, Leon Cooper and Paul Munro. Theory for the development of neuron selectivity : Orientation specificity and binocular interaction in the visual cortex. *Journal of Neuroscience,* 2: 32 – 48, 1982.

Chapman et al. Inhibition of nitric oxide synthesis impairs two different forms of learning. *NeuroReport,* 3: 567 – 570, 1992.

Dario Floreano and Joseba Urzelai. Evolution of plastic control networks. *Autonomous Robots,* 11: 311-317, 2001.

Joseph Gally et al. The NO hypothesis: Possible effects of a short-lived, rapidly diffusible signal in the development and function of the nervous system. *PNAS,* 87: 3547-3551, 1990.

Eldan Goldenberg, Jacob Garcowski and Randall Beer. May we have your attention : Analysis of a selective attention task. *Proc. Eighth Int. Conf. Sim. Adap. Behaviour,* 49-56, 2004.

Martin Hammer. The neural basis of associative reward learning in honeybees. *TINS,* 20 (6): 245 – 252, 1997.

Donald Hebb. The Organisation of Behaviour: A Neuropsychological theory. *Wiley, New York*, 1949.

Holscher and Rose. Inhibiting synthesis of the putative retrograde messenger nitric oxide results

in amnesia in a passive avoidance task in the chick. *Brain Research,* 619: 189-194, 1993.

Christian Holscher. Nitric Oxide, the enigmatic neuronal messenger: its role in synaptic plasticity. *TINS,* 20 (7): 298-303, 1997.

P. Husbands, T. Smith, N. Jakobi and M. O'Shea. Better Living Through Chemistry: Evolving GasNets for Robot Control. *Connection Science,* 10 (3&4): 185-210, 1998.

Javier Iglesias et al.. Stimulus-driven unsupervised synaptic pruning in large neural networks. *Proceedings of BV & AI,* LNCS 3704: 59-68, 2005.

Eugene Izhikevich and Niraj Desai. Relating STDP to BCM. *Letters to Neural Computation,* 15: 1511 – 1523, 2003.

Eugene Izhikevich. Which model to use for Cortical spiking neurons? *IEEE Transactions on Neural Networks,* 15 (5): 1063 – 1070, 2004.

Eugene Izhikevich, Joseph Gally and Gerald Edelman. Spike timing dynamics of neuronal groups. *Cerebral Cortex,* 14: 933-944, 2004.

Ayelet Katzoff et al. Nitric oxide is necessary for multiple memory processes after learning that food is inedible in *Aplysia. Journal of Neuroscience,* 22: 9581 – 9594, 2002.

Kemenes et al. Critical Time-Window for NO-cGMP-Dependent Long-Term Memory Formation after One-Trial Appetitive Conditioning. *Journal of Neuroscience,* 22: 1414 – 1425, 2002.

K. M. Kendrick et al. Formation of olfactory memories mediated by nitric oxide. *Letters to Nature,* 388: 670 – 674, 1997.

T. Lomo and T. Bliss. Long-lasting potentiation of synaptic transmission in the dentate area of the anesthetized rabbit following stimulation of the perforant path. *Journal Physioogy,* 232: 331-341, 1973.

Robert Malenka and Roger Nicoll. Long-term potentiation – A decade of progress? *Science,* 285: 1870 – 1874, 1999.

K.D. Miller and D.J. McKay. The role of constraints in Hebbian learning. *Neural Computation,* 6: 100-126, 1994.

K.D. Miller. Synaptic economics: competition and co-operation in synaptic plasticity. *Neuron,* 17: 371-374, 1996.

Uli Müller. Inhibition of nitric oxide synthase impairs a distinct form of long-term memory in the honeybee, *Apis mellifera. Neuron,* 19: 541–549, 1996.

Patrick Roberts and Curtis Bell. Spike timing dependent plasticity in biological systems. *Biological Cybernetics,* 87: 392 – 403, 2002.

J.D. Robertson et al. Nitric oxide is required for tactile learning in Octopus vulgaris. *Proc Biol Sci*. 256: 269 – 273, 1994.

Sen Song, Kenneth Miller and L.F. Abbott. Competitive Hebbian learning through spike timing dependent synaptic plasticity. *Nature Neuroscience,* 3: 919 – 926 , 2000.

T. Teyke. Nitric oxide but not serotonin is involved in aquisition of food-attraction conditioning in the snail *Helix pomatia. Neuroscience Letters* 206: 29-32, 1996.

M.C.W. van Rossum, G.Q. Bi and G.G. Turrigiano. Stable Hebbian learning from spike timing dependent plasticity. *Journal of Neuroscience,* 20 (23): 8812 – 8821, 2000.

M.C.W. van Rossum and G.G. Turrigiano. Correlation based learning from spike timing dependent plasticity. *Neurocomputing,* 38-40: 409-415, 2001.

How are nonlinearly separable discriminations acquired?

Chris Grand and R.C. Honey
Cardiff University, School of Psychology
Tower Building, Park Place
CF10 3AT, Cardiff
GrandC@cardiff.ac.uk and Honey@cardiff.ac.uk

Abstract

We present a series of experiments that investigate what animals learn during a nonlinearly separable discrimination (i.e., a patterning task) in which when two stimuli are presented in isolation (A or B) they are followed by one outcome (e.g., food) and when they are presented together (AB) they are followed by a different outcome (e.g., no food). After acquiring this patterning discrimination rats show greater generalization of fear between A and B than be-tween A and AB. Feed-forward models of both a binary (e.g., Rescorla & Wagner, 1972) and three-layer (e.g., Pearce, 1994) nature provide no obvious account for these results. Alterna-tive models that include reciprocal links will be presented that are better placed to explain some aspects our results.

The influence of motivational and training factors on the con-textual control of biconditional discrimination performance in rats

Josephine E. Haddon and Simon Killcross
Cardiff University, School of Psychology
Tower Building, Park Place
CF10 3AT, Cardiff
HaddonJE@cardiff.ac.uk and KillcrossAS@cardiff.ac.uk

Abstract

It is difficult to explicitly test the influence of motivational factors on decision making processes in humans, and hence it is difficult to assess the extent to which choice behaviour in humans, like rats, is best characterised ac-cording to Action-outcome (A-O) accounts of behaviour (Adams, 1982; Adams & Dickinson, 1981; Balleine & Dickinson, 1991; Dickinson, Balleine, Watt, Gonzalez, & Boakes, 1995). Therefore, it is of interest to investi-gate the role of motivational factors in rats in choice situations similar to those used in common cognitive para-digms in humans. Recently we have presented a task in rats that closely reflects aspects of the response conflict seen in the Stroop-like tasks in humans (Haddon & Killcross, 2005, in press). In this task, following training on two biconditional discrimination tasks, one involving auditory cues, the other visual, rats are presented with audiovisual test compounds the elements of which required either the same (congruent trials) or different (in-congruent trials) lever press responses during training. Test performance demonstrated that animals were able to use (previously incidental) contextual information to guide responding to the ambiguous response information provided by incongruent stimulus compounds. That is, animals responded in accordance with the stimulus ele-ment previously trained in the test context. This novel procedure was adapted to investigate the role of motiva-tional factors on performance in situations of response conflict such that different reinforcers were consistently presented in the two

different contexts (and thus consistently associated with one of the two biconditional dis-criminations, auditory and visual). When the training stimuli are combined at test to produce the audiovisual test compounds, not only are the two types of discrimination in competition, but the stimulus elements are also asso-ciated with different outcomes. Therefore, by manipulating the value of the different outcomes (by employing a reinforcer specific satiety devaluation procedure) we investigated the way in which the motivational variables related to the desirability of a particular outcome influenced the decision to respond when faced with conflicting response information. Furthermore, we can use these findings to establish whether performance on this task is best characterised by A-O approach or a stimulus-response (S-R) approach. We investigated the effect of differ-ential outcome devaluation on the contextual control of responding during incongruent stimulus compounds when both biconditional discrimination tasks were equivalently trained (Experiment 1) and when the bicondi-tional discrimination tasks were differentially trained (Experiment 2), a situation which more closely parallels the asymmetrical response competition observed in the Stroop task. Experiment 1 demonstrated that reinforcer devaluation resulted in a specific disruption of the context-appropriate responding to incongruent stimulus com-pounds, suggesting that the use of contextual control to disambiguate conflicting response information was modulated by the value of the expected reward. Experiment 2 extended this finding, demonstrating that per-formance on this task was influenced both by the training history of the biconditional discrimination stimuli and motivational influences related to the value of the outcome. These findings suggests a combined influence of both S-R and A-O processes in decision making processes and consequently has implications for current models of goal-directed choice behaviour in animals and in human decision making processes that involve motivational and emotional content such as moral reasoning (Greene, et al., 2001; Greene, et al., 2004).

References

Adams, C.D. (1982) Variations in the sensitivity of instrumental responding to reinforcer devaluation. *Quarterly Journal of Experimental Psychology*, Vol. 34B, 77-98.

Adams, C.D. & Dickinson, A. (1981) Instrumental responding following reinforcer devaluation. *Quarterly Journal of Experimental Psychology*, Vol. 33B, 109-122.

Balleine, B. & Dickinson, A. (1991) Instrumental performance following reinforcer devaluation de-pends on incentive learning. *Quarterly Journal of Experimental Psychology*, Vol. 43B, 279-296.

Dickinson, A., Balleine, B., Watt, A., Gonzalez, F. & Boakes, R.A. (1995) Motivational control after extended instrumental training. *Animal Learning and Behavior*, Vol. 23, 197-206.

Greene J.D., Nystrom L.E., Engell A.D., Darley J.M., & Cohen J.D. (2004) The neural bases of cog-nitive conflict and control in moral judgment. *Neuron*, Vol. 44, 389-400.

Greene J.D., Sommerville R.B., Nystrom L.E., Dar-ley J.M., & Cohen J.D. (2001) An fMRI investiga-tion of emotional engagement in moral judgment. *Science*, Vol. 293, 2105-8.

Haddon, J.E. & Killcross, A.S. (2005) Medial pre-frontal cortex lesions abolish contextual control of competing responses. *Journal of Experimental Analysis of Behavior*, Vol. 84 (November), 485-504.

Haddon, J.E. & Killcross, A.S. (in press). Prefrontal cortex lesions disrupt the contextual control of re-sponse conflict. Journal of Neuroscience.

Temporal uncertainty during overshadowing

Dómhnall Jennings[1], Eduardo Alonso[2], Esther Mondragón[3] and Charlotte Bonardi[1]

[1] University of Nottingham, Nottingham, **{djj,cmb}psychology.nottingham.ac.uk**

[2] City University, London, **eduardo@soi.city.ac.uk**

[3] University College London, London, **e.mondragon@ucl.ac.uk**

Abstract

In the natural environment, competing signals for an event of biological significance are likely to convey different temporal information about when that event will occur, and this may influence the ease with which these signals are learned about. An experiment will be described in which a compound of a noise and a light signalled the delivery of a food pellet to rats. This procedure, known as overshadowing, results in less learning to the light than would occur if it had been conditioned in isolation. The temporal information conveyed by the two stimuli was manipulated, in order to investigate the impact of this factor on the overshadowing effect. Specifically, two control groups were trained with a single light conditioned stimulus (CS) of fixed (F) or variable (V) duration, whereas four overshadowing groups received a noise-light compound. For two overshadowing groups the light was of fixed duration (30s), and for two it was variable (with a mean duration of 30s); for one of each of these pairs of groups the noise was fixed and for the other it was variable. This yielded two control groups (F and V) and four overshadowing groups (F(noise)F(light), V(noise)F (light), V(noise)V(light), F(noise)V(light). The dependent measure used in the study was the number of head entries to a food cup made by rats during the course of the stimuli. Date were analysed using a factorial model; the first factor was the distribution, either fixed or variable, of the target (light) while the second factor was whether the overshadowing (noise) stimulus was absent, fixed or variable . There was less responding to the target light when it was variable than when it was fixed, and less overshadowing of the light when the overshadowing noise was variable duration than when it was fixed. These results imply that conditioning is weaker under conditions of temporal uncertainty. Standard models of learning (e.g. Rescorla & Wagner, 1972) do not generally incorporate the effects of time on conditioning and as such they do not predict a difference in conditioning between fixed and variable stimuli. There are, however, models that attempt to explain both conditioning and timing within the same theoretical framework – hybrid models. The theoretical predictions of these models are

considered; it is argued that the best account is in terms of an adaptation of the temporal difference model of classical conditioning (Sutton & Barto, 1990).

The locus of learned predictiveness effects in human learning

M. E. Le Pelley lepelleyME@cf.ac.uk
M. B. Suret suretMB@cf.ac.uk
T. Beesley beesleyT@cf.ac.uk

School of Psychology, Cardiff University
Cardiff CF10 3AT

Abstract

Many previous studies of learned predictiveness effects in animal and human learning indicate an advantage for cues that have previously been experienced as good predictors of outcomes over those that have been poorer predictors. These studies do not, however, reveal whether learned predictiveness exerts its effects at the level of learning or performance (or both). An experiment using human participants and a novel "mutant scientist" paradigm was used to investigate this issue. Results indicated that altering the learned predictiveness of cues after a stage of critical learning influenced responding to those cues, demonstrating that learned predictiveness must exert an influence on performance, in terms of responding to cues.

1 Introduction

It is thought to be relatively well-established in the field of animal conditioning that the amount of processing power devoted to learning about a given conditioned stimulus (CS) is influenced by its past history of predictiveness at an associative level. Establishing a CS as a predictor of a reinforcing event seems to alter the readiness with which that stimulus will engage in later learning (see Le Pelley, 2004, for a recent review). One model of such learned predictiveness effects is that of Mackintosh (1975), which states that the change (Δ) in associative strength of cue A (V_A) on each learning episode is given by:

$$\Delta V_A = S\alpha_A\left(\lambda - V_A\right) \qquad (1)$$

where S is a constant learning-rate parameter, λ is the asymptote of conditioning supportable by the unconditioned stimulus (US) occurring on that trial, and α_A is the associability of cue A. The associability of a cue varies as a function of that cue's experienced predictive ability. Mackintosh proposed that α_A increases if A is a better predictor of the US on trial T than are all other presented cues; α_A decreases if A is a poorer predictor of the current US than are other presented cues.

A number of recent studies have indicated that learned predictiveness processes also exert an influence on human learning (e.g. Bonardi, Graham, Hall, & Mitchell, 2005; Kruschke & Blair, 2000; Le Pelley & McLaren, 2003; Le Pelley, Oakeshott, & McLaren, 2005). These human experiments all indicate that people devote more "processing power" to cues that have previously been established as reliable predictors of outcomes than those that have been established as poor predictors. The current paper represents a first effort to establish the level at which this learned predictiveness exerts its effects on cue processing – at the level of learning (as suggested by the Mackintosh model) or performance.

We will consider the study of Le Pelley and McLaren (2003) in more detail here, as it forms the focus of the current experiment. The basic design of this study is shown in the three left-hand, italicised columns of Table 1. In this table, letters A-Y refer to cues, and O1-O6 refer to outcomes that can be paired with those cues. Thus "AV → O1" indicates that cues A and V were presented together, and were paired with outcome O1.

During the first stage of this experiment, cues A and D were consistently paired with O1, cues B and C were consistently paired with O2, and cues V-Y provided no basis for discrimination between the two outcomes, being paired with O1 and O2 an equal number of times. As such, during Stage 1 cues A-D were the best available predictors of the outcome occurring on each trial and hence, according to the Mackintosh model, should have maintained a high α. The α of cues V-Y meanwhile should have decreased, as these were the poorer predictors of the outcome occurring on each trial.

On each of the Stage 2 trial types shown in Table 1, a good predictor from Stage 1 (A, B, C, or D) was paired with a poor predictor (V, W, X, or Y) with

Table 1: Design of Le Pelley and McLaren (2003), and current experiment

Stage 1	*Stage 2*	*(Test)*	Stage 3		Test
			Grp Consistent	Grp Inconsistent	
AV → O1	*AX → O3*	*(AC)*	AV → O5	AV → O5	AC
AW → O1	*BY → O4*	*(BD)*	AW → O5	AW → O6	BD
BV → O2	*CV → O3*	*(VX)*	BV → O6	BV → O5	VX
BW → O2	*DW → O4*	*(WY)*	BW → O6 **OR**	BW → O6	WY
CX → O2			CX → O6	CX → O6	
CY → O2			CY → O6	CY → O5	
DX → O1			DX → O5	DX → O6	
DY → O1			DY → O5	DY → O5	

Note: Entries in italics show trial types used by Le Pelley and McLaren (2003); whole table shows design of current experiment (except test in column 3, which was used only by Le Pelley & McLaren, 2003).

which it had not been presented in Stage 1, and this novel compound was paired with a novel outcome; compounds AX and CV with O3, and compounds BY and DW with O4.

Following Stage 2, participants were asked to rate how likely each of outcomes O3 and O4 was to follow various cue compounds. Following Dickinson, Shanks and Evenden (1984), these ratings provided an index of the strength of the cue–outcome associations developed over the course of training.

The Mackintosh model predicts that, at the end of Stage 1, cues A-D (good predictors in Stage 1) will have higher associabilities than cues V-Y (poor predictors in Stage 1). According to this model, this will promote more rapid learning of associations between these good predictors and the Stage 2 outcomes than between the poor predictors and Stage 2 outcomes. Therefore participants should develop strong associations from A and C to O3, strong associations from B and D to O4, weak associations from V and X to O3, and weak associations from W and Y to O4. In line with these predictions, participants rated compound AC as a strong predictor of O3 and compound BD as a strong predictor of O4, while VX and WY were perceived to be weak predictors of O3 and O4 respectively. This was exactly the pattern of results observed by Le Pelley & McLaren (2003).

This study clearly indicates a difference in the processing afforded to cues A-D and cue V-Y, and that this difference arises as a result of the difference in their experienced predictiveness during Stage 1. The question now becomes one of exactly where this learned predictiveness exerts its influence. The Mackintosh model as presented above (and as presented originally by Mackintosh, 1975) states clearly that associability influences *learning*, determining how rapidly a cue undergoes changes in associative strength. If two cues with different α values are presented simultaneously and reinforced, the cue with the higher α will develop a stronger association to the outcome than will the cue with the lower α. In this conceptualisation of the model, responding to cue A, R_A, is simply a function of that cue's associative strength, i.e.:

$$R_A = kV_A \qquad (2)$$

where k is a constant.

There exists an alternative view of the locus of learned predictiveness, however. Mackintosh (1975) raised the possibility that learned predictiveness might also influence *performance*, in terms of responding to a cue. That is, as opposed to Equation 2, Mackintosh tentatively suggested that responding might also be a function of α, i.e.:

$$R_A = k\alpha_A V_A \qquad (3)$$

In the absence of compelling experimental evidence to support the idea that α influences performance as well as learning, however, Mackintosh remained agnostic on this issue[1].

This raises the issue of how best to interpret the findings of Le Pelley & McLaren (2003). In the discussion above (and in the original paper) we appealed to a model implicating α in learning only. That is, cues that were experienced as good predictors in Stage 1 engaged the learning process more strongly in Stage 2 than those that were experienced as poorer predictors. Responding to the good predictors would then be greater on test, as these cues would have higher associative strengths. It is, how-

[1] Mackintosh (1975) did cite evidence from Wagner, Logan, Haberlandt and Price's (1968) study of the relative validity effect in support of the idea that α can influence performance. However, it is possible to explain this data using a model that makes no appeal to learned predictiveness effects at all (e.g. Rescorla & Wagner, 1972), and hence this evidence is not persuasive.

ever, also possible to account for these data using a model that implicates α in performance only, without influencing learning at all. Suppose that, as before, cues A-D develop higher α values than cues V-Y during Stage 1 (by virtue of the fact that the former cues are consistently paired with the same outcomes). If α does not influence learning, then during Stage 2 all cues will form equally strong associations to the outcomes with which they are paired. If α influences responding as in Equation 3, then responding to the good predictors from Stage 1 on test will be greater than that to the poor predictors. For example, for cues A and X at the time of test:

If $\alpha_A > \alpha_X$ and $V_A = V_X$
Then, by Equation 3, $R_A > R_X$

Thus it is theoretically possible to account for these data using a model that makes no recourse to α in the learning mechanism.

These predictions were tested by computational simulation, using two different models. In the first model α influenced learning only, with performance being directly proportional to associative strength. This model, then, effectively combined Equations 1 and 2, although it was modified slightly to allow for the multiple-outcome design of this experiment. In the second model, α exerted no influence on learning, but did influence performance. This model effectively combined the learning equation:

$$\Delta V_A = S\left(\lambda - V_A\right) \qquad (4)$$

with performance as specified in Equation 3[2]. Again the model was modified slightly to deal with a multiple-outcome design. It was found that both models could reproduce faithfully the results observed by Le Pelley and McLaren (2003). The exact formulation and parameterisation of these models is unimportant for the current discussion. The simulations merely provide an existence proof that both approaches are able to explain the advantage for good predictors over poor predictors observed in this experiment. It would also, of course, be possible to have α influence both learning *and* performance (i.e. combine Equations 1 and 3), with the resultant model also able to account for these results.

The idea that learned predictiveness might influence performance as well as learning has been taken up explicitly by Kruschke (1996; 2001) in his ADIT and EXIT models. Despite these strong claims regarding the locus at which α exerts its effects, to the best of our knowledge there currently exists no conclusive empirical evidence bearing on this issue. All of the empirical effects of learned predictiveness at present described in the literature, in both animal and human studies, could be explained by a model implicating α in learning only, in performance only, or in both learning and performance. The experiment described in this paper represents a first attempt to decide between these alternative views in a study of human learning.

The design of our experiment is shown in Table 1 (whole table apart from the third column, shown in parentheses). Stages 1 and 2 are as for Le Pelley and McLaren (2003). Immediately following Stage 2, participants receive a third stage of training, in which cue compounds are paired with novel outcomes (O5 and O6). For participants in Group Consistent, cues A-D are once again good predictors of outcomes in this stage (A and D are consistently paired with O5; B and C are consistently paired with O6), while cues V-Y are poor predictors (being paired with O5 and O6 an equal number of times). For these participants, then, predictiveness in Stage 3 is consistent with what was learnt in Stage 1. For participants in Group Inconsistent, on the other hand, cues A-D are poor predictors, whereas cues V-Y are good predictors in Stage 3. For this group, predictiveness in Stage 3 is inconsistent with what was learnt in Stage 1. Following Stage 3, participants are asked to rate how likely it is that each of the Stage 2 outcomes (O3 and O4) would follow compounds AC, BD, VX, WY. This is the same test as used by Le Pelley and McLaren (2003). The question of interest is what effect Stage 3 training has on the pattern of responding to these compounds.

For both groups, at the outset of Stage 2 the α values for cues A-D should be higher than those for V-Y as a result of the higher predictiveness of the former cues during Stage 1. If α influences learning, then this will promote more rapid learning of associations between A-D and the Stage 2 outcomes than associations between V-Y and the same outcomes. The α values of the cues should diverge in the two groups during Stage 3. In Group Consistent cues A-D will maintain a high α during Stage 3, and cues V-Y will maintain a low α. In Group Inconsistent, on the other hand, we might expect the α of V-Y to rise over Stage 3, and the α of A-D to fall (reflecting the reversed predictiveness of these cues). Nevertheless, if α affects learning *only*, then there is no way for these subsequent changes in α to influence participants' reports of the relationship between test cues and Stage 2 outcomes (they will not directly influence the associations formed in Stage 2

[2] We are not wedded to a particular formalization of either of these models. For example, the models could be instantiated with an aggregated error term ($\lambda - \Sigma V$) in the learning equation instead of a separable error term ($\lambda - V_A$), which would render them capable of explaining more examples of cue competition in learning (see Le Pelley, 2004). Our aim is to contrast models involving α in either learning or performance components, and we have chosen the simplest models that illustrate these points.

as training in Stage 3 is with different outcomes to those used in Stage 2). In other words, if responding is purely a function of associative strength, then there is no way that subsequent changes in α can influence this responding. Therefore, if α influences learning only then we would expect similar results in Group Consistent and Group Inconsistent, with both groups providing higher ratings for compounds AC and BD than compounds VX and WY.

Suppose instead that α affects performance only. In that case, all cues will develop equally strong associations to whichever of outcome O3 or O4 they are paired with during Stage 2. However, changes in α during Stage 3 will influence the extent to which these associations are expressed in ratings made on test. In Group Consistent, we would expect stronger responding to compounds made up of cues A-D (as these cues maintain high α during Stage 3) than to compounds made up of cues V-Y (which maintain low α during Stage 3). In Group Inconsistent, however, the pattern of α values at the time of test will be quite different. To the extent that the α of cues V-Y ends Stage 3 higher than that of cues A-D, we would expect that participants would give higher ratings for compounds VX and WY than for AC and BD. That is, the reversal of the α values of the component cues during Stage 3 should interfere with the effect observed by Le Pelley and McLaren (2003), reducing (and possibly reversing) the advantage for AC/BD over VX/WY.

This selective influence on the pattern of responding to the different compounds will only be observed if α influences performance. As noted earlier, it is possible that α influences both learning and performance. On this view we would still expect Inconsistent training to reduce the advantage for AC/BD over VX/WY. However, we might expect the interfering effect of Stage 3 training to be slightly less, as AC/BD will retain the advantage of having higher associative strengths, but will lose out in terms of having lower α values on test.

This experiment used a novel "mutant scientist" paradigm. Participants played a scientist who specialises in creating mutants. They were told that mutants are created by combining certain chemicals with a special "goo" substance. Thus the letters A-Y `in Table 1 were represented by different chemicals, and outcomes O1-O6 by different types of mutants that could be created.

2 Method

2.1 Participants, Apparatus and Materials Thirty-eight Cardiff University undergraduates participated in exchange for course credit. Participants were randomly assigned to groups, with 19 in each of Group Consistent and Group Inconsistent. Participants were tested individually, with stimuli presented on a 17-inch computer monitor, and all responses made via the mouse. The eight chemical names were Bizancrine, Daktyre, Halorite, Kluphane, Nelomine, Ontone, Quezalin, and Yestimox. These were randomly and independently assigned to the letters A-Y in the experimental design shown in Table 1 for each participant. The six mutant names were Draguts, Goygle, Jominoid, Necromon, Rargon and Snarlig, which were again randomly assigned to outcomes O1 to O6 for each participant. Pictures of the different mutants were obtained from the web, with pictures being randomly assigned to mutant names for each participant.

2.2 Procedure At the outset of the experiment, participants read the following on-screen instructions:

> "In this experiment you take on the role of a scientist who specialises in creating mutants. A mutant is created by combining different chemicals with a special blue 'goo' substance. When certain chemicals are combined with the blue goo a mutant is born. Different chemicals can produce different types of mutants, but some chemicals might have no effect at all.
>
> You have just been given a newly-discovered set of chemicals to experiment with. In an attempt to discover which chemicals result in the creation of the different types of mutants, you arrange a series of trials. On each trial, the chemicals to be used are displayed at the top of the screen. On all trials you use two different chemicals at the same time.
>
> Your aim is to predict which mutant will be created when you mix the chemicals with the goo.
>
> On each trial, click on the mutant that you think will be created when the chemicals displayed at the top of the screen react with the goo. When you have made your decision, click the OK button. The computer will then tell you whether your prediction was correct or incorrect. A blue box will appear, indicating the mutant that was actually created on that trial. If you make an incorrect prediction the computer will beep.
>
> Since nobody currently knows the effect of these new chemicals (or in fact whether any of them are capable of creating mutants at all) you will start out guessing, but with the aid of the feedback your predictions should start to become more accurate.
>
> Your reaction times are not important: you may take as long as you like on each trial."

A typical Stage 1 trial is shown in Figure 1. On each trial the message "The following chemicals are used:" appeared at the top of the screen, above the names of two chemicals, followed by the message "What sort of mutant do you think will be created? Click on the mutant that you predict, then click OK". Below this were pictures of two mutants, along with their corresponding names. Participants entered their predictions by clicking on one of these pictures, and then clicking an OK button. Immediate feedback was then provided: a blue box highlighted the correct answer for that trial. If participants had made a correct prediction, the word "Correct" appeared in place of the "What sort of mutant…" question; if they had made an incorrect prediction, the word "Wrong" appeared and the computer beeped.

Stage 1 comprised 14 blocks, with each of the eight trial types occurring once per block. Trial order within a block was randomized, with the constraint that there could be no immediate repetitions across blocks. For each trial type the order of presentation of the chemicals (left/right) was counterbalanced across blocks. For example, for trial

type AV→O1, there would be seven presentations with chemical A to the left of chemical V, and seven presentations with V to the left of A (the order of these presentations was randomized). The two mutants presented on each Stage 1 trial were always O1 and O2. For each trial type, the order of presentation of mutants (left/right) was counterbalanced across blocks. So for trial type AV→1, there would be seven presentations with mutant O1 to the left of mutant O2, and seven presentations with mutant O2 to the left of mutant O1 (again in random order).

After Stage 1, the following message appeared:

> "The next phase of your research involves using a slightly different type of goo. This red goo creates new types of mutants. Some of the chemicals that you mix with this red goo are the same as those that you used earlier. Once again your aim is to predict what type of mutant will be created when each combination of chemicals is mixed with the red goo, by clicking on the appropriate mutant and then clicking OK. Feedback will be provided, and should allow your predictions to become more accurate."

The form of each Stage 2 trial was the same as that for Stage 1, except that (i) the goo pictured on each trial was red, rather than the blue used in Stage 1, and (ii) the two mutants pictured on each trial represented O3 and O4. There were six blocks in Stage 2, with each of the four trial types appearing once per block. Counterbalancing and randomization of trial order, chemical presentation order and mutant presentation order were as for Stage 1.

The message appearing at the end of Stage 2 was the same as that appearing at the end of Stage 1, except that now participants were told that, in the following stage, they would be using a yellow goo. The form of each Stage 3 trial was the same as for the previous stages, except that (i) the goo pictured on each trial was yellow, and (ii) the two mutants on each trial represented O5 and O6. Stage 3 comprised 10 blocks, with counterbalancing and randomisation as in previous stages.

After Stage 3, the following message appeared:

> "Your work in creating mutants is starting to be recognised and you are becoming an esteemed professional in the field! However, some people are a little sceptical about your scientific understanding of the experiments you are conducting. As a test you are asked to make decisions for certain individual chemicals.
>
> You should rate how likely you think it is that each type of mutant will be created when THIS CHEMICAL ALONE is mixed with the red goo, on a scale from 0 to 10. A rating of 0 means that mixing this chemical alone with the red goo is VERY UNLIKELY to create that type of mutant, while a rating of 10 means that mixing this chemical alone with the red goo is VERY LIKELY to create that type of mutant. You may use any value from 0 to 10 to indicate your opini on.
>
> When you have entered your rating, click OK to continue. Note that you will end up rating each chemical twice, giving one rating for each type of mutant. You will not receive feedback on these ratings."

Each of the four test compounds shown in Table 1 was presented in random order for rating. On each test trial, the message "The chemicals used are:" appeared above the names of two chemicals, which were pictured being poured onto the red goo that had been used during Stage 2. Below that came the message "How likely is it that the following mutant will be created?", along with a picture and name of one of the Stage 2 mutants (O3 or O4). Participants entered their rating by clicking one of 11 radio buttons labelled from 0 to 10, the leftmost being 0 (labelled "Chemicals very unlikely to create this mutant"), and the rightmost being 10 ("Chemicals very likely to create this mutant"). Participants rated the ability of a given pair of chemicals to create one type of mutant (e.g. O3), and on the succeeding trial rated the ability of that same compound to create the other type of Stage 2 mutant (O4 in this case). Whether participants rated mutant O3 or O4 first was consistent across all test compounds, and was determined randomly for each participant.

3 Results and Discussion

Figure 2 shows mean percent correct of participants' predictions during each block of the three training stages (chance = 50% correct). Learning is evident in all stages. Over Stages 1 and 2, performance is very similar in Groups Consistent and Inconsistent (which receive identical training during this time). T-tests conducted on the overall percent correct (averaged across all blocks) for each participant in each stage found no difference in performance for Stage 1 or Stage 2, both ts < 1. During Stage 3, however, learning is considerably more rapid in Group Consistent than in Group Inconsistent. A t-test on the overall performance for each group revealed a significant difference, $t(36) = 2.27$, $p < .05$. The difference between the groups decreases as training continues, however, with no significant difference in performance on the final block of Stage 3, $t < 1$.

This difference in performance of the two groups during Stage 3 is predicted by both views of the locus of α, in learning or performance. According to both models, cues that were better predictors during Stage 1 (A-D) will begin Stage 3 with higher αs than those that were poorer predictors (V-Y). On the learning-based view of α, this will tend to favour cues A-D over cues V-Y in the learning process

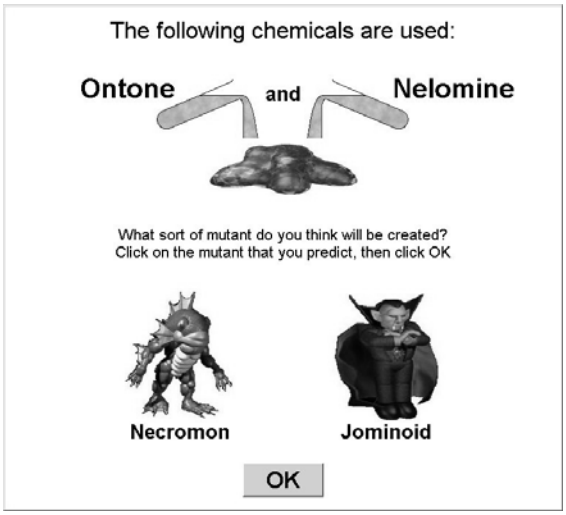


Figure 1: Screenshot of a typical Stage 1 trial.

during Stage 3. This will produce faster learning in Group Consistent (for whom A-D are the cues that predict the correct answer and so must be learnt about for performance to improve) than in Group Inconsistent (for whom cues A-D are irrelevant, and instead V-Y must be learnt about). On the performance-based view of α, responding to cues A-D will tend to be amplified relative to that for cues V-Y. This will enhance performance in Group Consistent, as A-D are the cues that will eventually come to control responding (by virtue of consistent pairings with the same outcomes during Stage 3). Performance will be relatively impaired in Group Inconsistent, however, as cues V-Y (which will be only weakly responded to by virtue of their low αs) are those that must ultimately come to control responding. As such, the observation of a performance difference during Stage 3 training cannot decide between learning- and performance-based views of α.

The difference in performance of the two groups during Stage 3 relates to studies of intradimensional and extradimensional shift learning in animals and humans (e.g. Schwartz, Schwartz, & Teas, 1971; Shepp & Schrier, 1969; Whitney & White, 1993). The difference is that in these more traditional designs it is the cues that change between pre-shift and post-shift discriminations while the outcomes remain the same: in the current design, the cues remain the same while the outcomes change.

The results of main interest from this study concern the ratings given to compounds during the test phase. Participants provided two ratings for each compound: one for how strongly that compound predicted O3, the other for how strongly it predicted O4. Following Le Pelley and McLaren (2003; see also Le Pelley, Oakeshott, & McLaren, 2005) we used these ratings to calculate difference scores for each compound. This was done by taking the rating for each compound with respect to the outcome (O3 or O4) with which its constituent cues were paired in Stage 2, and subtracting from that the rating for the same compound with respect to the outcome with which its cues were not paired in Stage 2. For example, the difference score for AC is given by the rating for AC with respect to O3 minus the rating for AC with respect to O4, because A and C were paired with outcome O3 during Stage 2. Likewise, the difference score for BD is given by BD's rating for O4 minus its rating for O3, because B and D were paired with O4 during Stage 2. These difference scores index the differential predictiveness of compounds with respect to Stage 2 outcomes – that is, the extent to which a compound predicts the outcome with which it was paired more than it predicts the outcome with which it was not paired. High difference scores indicate strong, selective performance, while a difference score of zero indicates no selective performance. The advantage of using difference scores over raw rating data is that the former are free from influences of generalization that would otherwise render the results uninterpretable (see Le Pelley, Oakeshott, Wills, & McLaren, 2005).

Finally, we averaged difference scores for compounds AC and BD (which both consist of cues that were good predictors during Stage 1), and for compounds VX and WY (which both consist of cues that were poor predictors during Stage 1). These mean difference scores are shown in Figure 3.

In Group Consistent, responding to compounds made up of good predictors from Stage 1 (AC/BD) is considerably better than that for compounds made up of poor predictors from Stage 1 (VX/WY). The pattern of responding in Group Inconsistent is quite different, however. In this latter group, responding to AC/BD is, if anything, poorer than that for VX/WY. These data were initially analysed using ANOVA with one between-subjects factor of Group (Consistent vs Inconsistent) and one within-subjects

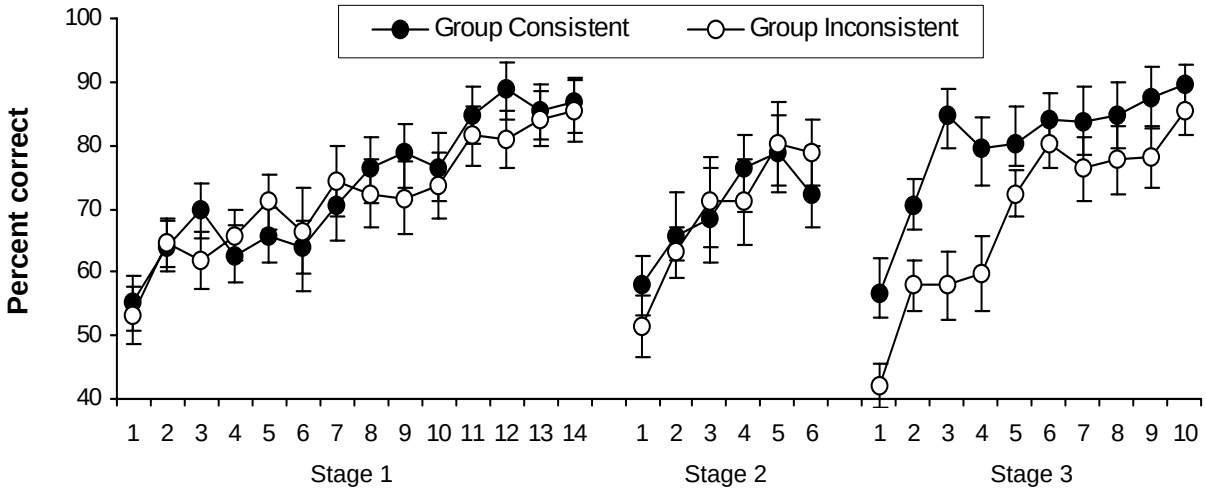


Figure 2: Percent correct (averaged over all trial types) across blocks during the 3 training stages.

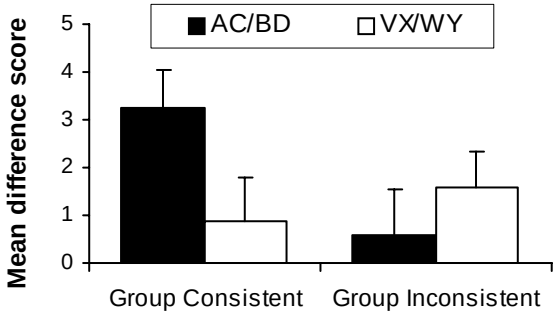


Figure 3: Mean difference score for the test compounds of Groups Consistent and Inconsistent.

factor of Compounds (AC/BD vs VX/WY). This revealed that there was no main effect of Group, $F < 1$, or Compounds, F (1, 36) = 1.33, $p = .26$. The interaction of these two factors, however, was highly significant, F (1, 36) = 8.06, $p < .01$, indicating a significant difference in the pattern of performance in the two groups.

Pre-planned t-tests were used to analyse this interaction further. Paired t-tests revealed that, in Group Consistent, compounds AC/BD received difference scores that were significantly higher than those of VX/WY, t (18) = 3.20, $p < .01$, while in Group Inconsistent the apparent reversal in performance to these compounds (with mean difference score for VX/WY higher than that for AC/BD) failed to reach significance, t (18) = 1.08, $p = .30$.

Group Consistent showed a clear learned predictiveness effect in line with that observed by Le Pelley and McLaren (2003), with better performance to compounds made up of cues that were good predictors during Stages 1 and 3. This confirms that α is exerting effects in this study, but does not tell us whether it is exerting these effects in learning or performance. The Stage 3 training received by Group Inconsistent, during which cues that had previously been experienced as predictive (A-D) were now found to be nonpredictive, and *vice versa*, exerted a selective influence on the pattern of results. In this latter group, there was no longer any advantage for compounds AC/BD over VX/WY on test. Thus it would seem that changes in α of cues *after* the critical learning phase during Stage 2 are sufficient to alter responding to those cues. These results therefore lie beyond a model in which α only affects learning. Instead they demand that α is able to influence performance, in terms of responding to cues.

Before we can be completely confident in this conclusion, however, we must rule out two alternatives. One possibility is that the Stage 3 training of Group Inconsistent (which, as evidenced by Figure 2, is somewhat harder than that for Group Consistent) caused them to "give up" at this task, with performance on test falling to floor levels such that no advantage for AC/BD over VX/WY could be observed. Two aspects of the data contradict this suggestion. The first is the failure to find a significant main effect of Group in the ANOVA, indicating that overall level of performance of the two groups on test was comparable. The second is that performance to VX/WY is, if anything, better in Group Inconsistent than in Group Consistent. One-sample t-tests of the VX/WY score against a hypothesised mean of zero (indicating no learning) reveal a significant effect for Group Inconsistent, t (18) = 1.58, $p < .05$, but no significant effect for Group Consistent, $t < 1$. Thus we have evidence for appropriate responding to VX/WY in Group Inconsistent, but not in Group Consistent (although direct comparison of the two groups on responding to VX/WY fails to reach significance, $t < 1$). If the failure to observe a difference in responding to AC/BD and VX/WY in Group Inconsistent were the result of a generally lower level of responding than in Group Consistent, then we would expect performance on VX/WY to be nonsignificant, given that it is nonsignificant in Group Consistent. That this is not observed is a sign that the difference between these two groups is more selective, in that responding to AC/BD is impaired in Group Inconsistent compared to Group Consistent, but responding to VX/WY is, if anything, enhanced in the former group.

A second alternative is that the difference between groups reflects some form of associative interference arising from Stage 3 training. Up to this point we have assumed that the only influence of Stage 3 training on responding to the Stage 2 relationships will be in terms of changes in α exerting an influence on the extent to which Stage 2 associations are expressed. It is possible, however, that Stage 3 training will exert a more direct effect in terms of retroactive interference: associations developed during Stage 3 could interfere with retrieval of information learnt in Stage 2 at the time of test. However, it is hard to see how this could produce the results that we observed. During Stage 3, cues that are good predictors will doubtless develop stronger associations to their respective outcomes than will cues that are poorer predictors. In Group Inconsistent, this would produce greater interference for cues V-Y than for cues A-D, with the prediction that (compared to Group Consistent), we should see a greater impairment in responding to VX/WY than to AC/BD. This is, of course, the opposite of the results observed.

Our results therefore seem to demand that α is able to influence performance as suggested in Equation 3, providing the first evidence to bear specifically on the locus of learned predictiveness effects. As acknowledged earlier, it remains to be seen whether α also influences learning. The fact that the level of responding to AC/BD and VX/WY did not reverse significantly in Group Inconsistent might be

seen as indicating that there is a persistent advantage for AC/BD in terms of higher associative strengths developed as a result of the higher αs of these cues during Stage 2. The influence of α on responding may then be insufficient to completely reverse this advantage. However, it is also possible that changes in α during Stage 3 are sufficiently slow that reversal on the basis of responding is not seen even if all cues have equal associations to the Stage 2 outcomes. As it stands, then, a model incorporating α at the response level only is able to incorporate all of our current data, and that of all other studies of learned predictiveness effects that have taken as support of the Mackintosh (1975) model in both human and animal learning. Whether α also influences learning remains as an issue for future research.

Acknowledgements

This work was supported by ESRC grant RES-000-23-0983 to M. E. Le Pelley.

References

Bonardi, C., Graham, S., Hall, G., & Mitchell, C. (2005). Acquired distinctiveness and equivalence in human discrimination learning: Evidence for an attentional process. *Psychonomic Bulletin & Review, 12*, 88-92.

Dickinson, A., Shanks, D. R., & Evenden, J. L. (1984). Judgement of act-outcome contingency: The role of selective attribution. *Quarterly Journal of Experimental Psychology, 36A*, 29-50.

Kruschke, J. K. (1996). Dimensional relevance shifts in category learning. *Connection Science, 8*, 225-247.

Kruschke, J. K. (2001). Towards a unified model of attention in associative learning. *Journal of Mathematical Psychology, 45*, 812-863.

Kruschke, J. K., & Blair, N. J. (2000). Blocking and backward blocking involve learned inattention. *Psychonomic Bulletin & Review, 7*, 636-645.

Le Pelley, M. E. (2004). The role of associative history in models of associative learning: A selective review and a hybrid model. *Quarterly Journal of Experimental Psychology, 57B*, 193-243.

Le Pelley, M. E., & McLaren, I. P. L. (2003). Learned associability and associative change in human causal learning. *Quarterly Journal of Experimental Psychology, 56B*, 68-79.

Le Pelley, M. E., Oakeshott, S. M., & McLaren, I. P. L. (2005). Blocking and unblocking in human causal learning. *Journal of Experimental Psychology: Animal Behavior Processes, 31*, 56-70.

Le Pelley, M. E., Oakeshott, S. M., Wills, A. J., & McLaren, I. P. L. (2005). The outcome-specificity of learned predictiveness effects: Parallels between human causal learning and animal conditioning. *Journal of Experimental Psychology: Animal Behavior Processes, 31*, 226-236.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review, 82*, 276-298.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64-99). New York: Appleton-Century-Crofts.

Schwartz, R. M., Schwartz, M., & Teas, R. C. (1971). Optional intradimensional and extradimensional shifts in the rat. *Journal of Comparative and Physiological Psychology, 77*, 470-475.

Shepp, B. E., & Schrier, A. M. (1969). Consecutive intradimensional and extradimensional shifts in monkeys. *Journal of Comparative and Physiological Psychology, 67*, 199-203.

Wagner, A. R., Logan, F. A., Haberlandt, K., & Price, T. (1968). Stimulus selection in animal discrimination learning. *Journal of Experimental Psychology, 76*, 171-180.

Whitney, L., & White, K. G. (1993). Dimensional shift and the transfer of attention. *Quarterly Journal of Experimental Psychology, 46B*, 225-252.

A Hybrid Cognitive-Associative Model to Simulate Human Learning in the Serial Reaction Time Paradigm

Rainer Spiegel[*]
[*]Institute of Medical Psychology
Ludwig-Maximilians-University
Goethestrasse 31/1
80336 Munich, Germany
rainer.spiegel@campus.lmu.de

I.P.L. McLaren[†]
[†]Department of Experimental Psychology
University of Cambridge
Downing Site
Cambridge, CB2 3EB, UK
iplm2@cus.cam.ac.uk

Abstract

We present a computational model to simulate the findings of a series of experiments using the Serial Reaction Time paradigm on the problem devised by Maskara and Noetzel (1993). In contrast to other hybrid architectures, the model presented here simulates the experimental findings rather closely, although the predictions made by the model are counter-intuitive with respect to variants of the problem. The general finding is less counter-intuitive and can be predicted by the model as well: shorter and less numerous sequences can be better represented cognitively, whilst associative learning drives performance on longer and more numerous sequences.

1 Introduction

In this paper we have two main aims: To present evidence that humans are able to make use of both cognitive and associative learning strategies based on different task conditions, and to show that this requires a hybrid architecture where cognitive and associative parts of the model interact. This model differs from other cognitive architectures that exist for similar problems, but were not applicable to fully simulate human performance in our task (e.g. Anderson, 1993; Hofstadter, 1995; Marshall, 1999; Mitchell, 1993; Slusarz & Sun, 2001; Sun et al. 2001).

Our approach will be to take a sequence learning problem, devised by Maskara and Noetzel (1993), that can be learned associatively (e.g. Spiegel & McLaren, in press) and simulated with an associative model such as the Simple Recurrent Network (SRN, Elman, 1990). The SRN not only predicted human learning in this task, but also the way people generalised to variants of the problem. In this paper we will present further variations to the task parameters and show conditions where the SRN still learns the problem, but does not generalise the way people do. Briefly, we will also discuss conditions when learning breaks down and, as a consequence, no generalisation is observed. In order to capture learning as well as generalisation in this task paradigm, a new model was created. It is based on 2 initial experiments and was cross-validated with 5 further experiments. Whilst the model will be discussed in detail, not all experimental data will be presented here, as the results supporting associative performance have already been published (Spiegel & McLaren, in press). The model will be easier understood with the knowledge of the experimental paradigm and the SRN, which is a sub-component in our hybrid cognitive-associative architecture.

Consequently, we will start by briefly describing the SRN and its predictions for Maskara and Noetzel's sequence learning problem, followed by the experimental paradigm to test human performance.

The SRN belongs to the family of associative learning models, because apart from its general learning algorithm, it does not contain any pre-implemented rules or strategies that tell it how to solve particular problems. As a consequence, it has often been applied to model different kinds of human learning without awareness, sometimes termed implicit learning, e.g. (Cleeremans, 1993). The architecture of the SRN can be found in Figure 1.

The SRN receives sequential information in the form of vectors containing binary numbers. A sequential element is presented to the SRN at the input units and the purpose of the SRN is to predict the next input (i.e. the next sequential element) at the output units. The present input is processed by the hidden units. These capture it along with a blended

representation of all previous sequential steps coming from the context units. The hidden units then drive the output unit activations. At the output units, the desired activations of output units are contrasted with the current activations of output units. This discrepancy is applied to tell the SRN which weight values to update through an associative learning algorithm called backpropagation (Rumelhart et al. 1986).

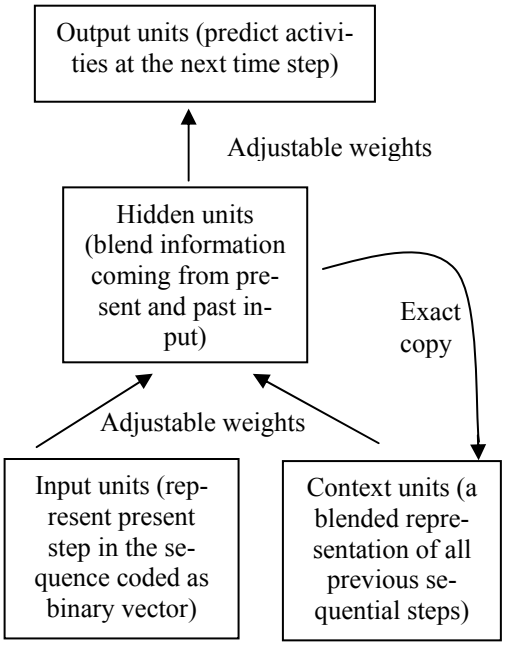


Figure 1: Jeffrey Elman's Simple Recurrent Network (SRN)

2 Experiments

In all the experiments reported in this paper, human participants were trained on a serial reaction time task (Nissen & Bullemer, 1987), where they were asked to press keys on their keyboard corresponding to sequences of signals on a computer screen. Learning was assessed via changes in reaction times and accuracy, and also indexed by measuring generalisation to novel sequences. At no time were the participants told that there was structure in the sequences they were exposed to. The participants' task was simply to respond as rapidly as possible to the signal (a circle filling) by pressing the appropriate key. The screen turning black at the end of each sequence was the only segmentation occurring. In the model, setting context units to zero marked the end of a sequence. Hence, this was an incidental learning paradigm as far as our human subjects were concerned, in that they believed themselves to be engaged in nothing more than a choice reaction time task. In the model tests, changes in unit activations were taken to be equivalent to reaction time and accuracy changes in humans. We chose the simple recurrent network (SRN) as our associative model, because it is a widely recognised associative model of sequence learning (Cleeremans, 1993; Marcus et al. 1999; Pinker, 1999) that had been applied to simulate other types of learning without awareness before (Cleeremans, 1993). Human subjects were trained on sequences using three screen locations. For ease of exposition, the screen locations can be denoted by the first three letters of the alphabet, with each letter standing for a different screen location. The experimental sequences we applied had the following structure:

1st sequence type:
A*B*(varying numbers of Cs)*B*A

2nd sequence type:
A*BB*(varying numbers of Cs)*BB*A

Humans, as well as the associative model, were expected to learn that the number of B flashes was always the same before and after the varying numbers of C flashes. In all the experiments (apart from where noted) 18 Experimental and 18 Control participants / models were trained on two consecutive sessions. The Experimental groups received 100 percent of their training on the sequences displayed above, whose grammar we denote as: AB(C+)BA and ABB(C+)BBA (each letter corresponding to a particular circle on the screen). Because the experiment was counterbalanced across participants, all three letters could correspond to any of the three circles, depending on the counterbalance condition. In what we call *consistent* sequences, there were always as many Bs before as after the Cs. Thus, although the number of Cs is variable, the Experimental group was expected to learn to anticipate the final A in the first sequence type once the B followed the Cs and the final B in the second sequence type once the preceding B followed the Cs. In contrast, the Control group just received 50 percent of their training on the *consistent* sequences, and the other 50 percent of training on the following *inconsistent* sequences: AB(C+)BB and ABB(C+)BAA. Hence, for the Control group, the rule that there are as many Bs before as after the Cs does not hold true. As a result, the Control group was not expected to learn to anticipate the final location in the first sequence type and the location preceding the final location in the second sequence type. The training phase was followed by a testing phase, in which

both Experimental and Control groups received an equal number of consistent and inconsistent sequences. If the Experimental group had learned the rule inherent in the consistent sequences, they should perform significantly better than controls on the respective locations in the consistent sequences than in the inconsistent sequences (which they had not received throughout the training phase).

When the SRN was run on this problem it soon became clear that, first of all, it was able to solve it. This finding stood in contrast to earlier predictions by Maskara and Noetzel (1993). In other words, the network was able to predict the final A in a single B sequence, and the second B after the intervening C elements in a double B sequence. In all our simulations the same numbers of Experimental and Control SRNs (each with a different random seed) as humans were trained to facilitate direct comparison between human and model learning. All SRNs were trained with 6 hidden units, a learning rate of 0.1, no momentum term and 40,000 training trials (= total number of sequence presentations during training). We chose these parameters because this is the minimum number of hidden units and training trials that the SRNs needed to converge on this problem, and more hidden units or training trials did not lead to better generalisation to the novel sequences. Otherwise our simulations were entirely standard and for more information about the chosen parameters, see Rumelhart et al. (1986).

The only way the following experiments and simulations varied was with respect to the intervening number of C elements. In Experiment 1, we start by training and testing participants on 1 and 3 intervening C elements. In addition, we will test their generalisation performance to 2 C elements. In Experiment 2, we train and test participants on 1, 3, 5 and 7 C elements and will check generalisation to 2, 4 and 6 C elements.

2.1 Experiment 1

2.1.1 Method

Participants

Participants (aged 18-40 years) were randomly assigned to two groups, Experimental and Control. There were 15 participants in each group, who were paid for their participation.

Apparatus

The experiment was programmed in the Future Basic II programming language, was run on a Macintosh computer and took place in a quiet room. Its light was dimmed to a level that had been indicated as convenient by other participants in a pilot test. The participants' distance from the screen was approximately 80cm, which was roughly at eye level. The screen's diagonal was 30cm in size. White outlines of circles were arranged in a triangular formation on a black background (Figure 2). The circles flashed by filling with white colour one at a time. Circles were two centimetres in diameter and the centres of the circles on the bottom of the triangle were approximately 5.5cm apart. The centre of the upper circle was approximately 4cm apart from the centres of the two lower circles. We chose this discrepancy in distance because pilot work had indicated that an equilateral triangular shape would probably confuse participants in terms of the horizontal assignment of keys. The horizontal assignment consisted of the adjacent "v", "b" and "n" keys on each participant's keyboard, with the v-key corresponding to the lower left circle, the b-key to the upper middle circle and the n-key to the lower right circle. The three outlines were located in the middle of the computer screen. A triangular layout was chosen instead of other possibilities. This was to avoid any particular stimulus appearing in the foveal area, because this might have had the effect of biasing participants on where to look (see Lewicki et al. 1987; Nelson & Loftus, 1980).

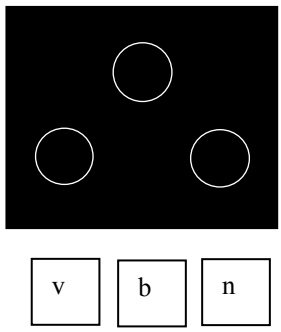


Figure 2: The three circle outlines with their three corresponding keys from a conventional computer keyboard underneath the display (note that in the experiment, the entire background of the screen is black whilst the circles appear in the middle of the screen). Participants were instructed to use index, middle and ring fingers of their preferred hand.

Procedure

During the experiment, the entire screen background was black, with the outlines of three white circles in the middle. Each circle was assigned a particular key on the keyboard, and participants were instructed to press the respective key as quickly and accurately as possible once a circle filled with white; only one circle at a time filled

with white colour. After each key press, the three outlines reappeared for 180ms until the next circle filled. An auditory signal sounded if the subject had pressed the wrong key. After the last response of a sequence the screen turned black for 600ms until the outlines reappeared so that the next sequence could begin.

The Experimental groups received 100 percent of their training on *consistent* sequences: AB(C+)BA and ABB(C+)BBA (with each letter corresponding to a particular circle on the screen or key on the keyboard respectively). All three letters could correspond to any of the three circles, depending on the counterbalance condition. The Control group only received 50 percent of their training on the *consistent* sequences, and the other 50 percent of training on the following *inconsistent* sequences: AB(C+)BB and ABB(C+)BAA. The experiment consisted of 4 training blocks. The Experimental groups received 18 randomly selected presentations of each consistent sequence in each block, whilst the Control groups received 9 presentations of all consistent and 9 presentations of all inconsistent sequences in each block.

The training was followed by 2 test blocks in which both Experimental and Control groups received 6 presentations of all sequences in each block (i.e. consistent and inconsistent sequences, including those with 2 Cs to test generalisation). In both training and testing phase, Experimental and Control groups had received an equal number of consistent and inconsistent sequences. So if the Experimental group had learned the rule inherent in the consistent sequences, they should perform significantly better on the respective locations in the consistent sequences than in the inconsistent ones (which they had not received during training). In the Control group, we do not expect a significant performance difference between consistent and inconsistent sequences as they had experienced both during the training phase. After the last trial on the second day, an interview was carried out exploring to what extent the participant was able to verbalise the sequential structure or fragments of it. We also investigated whether s/he had become aware of particular rules or the number of C flashes during training and/or testing.

2.1.2 Results

For the trained sequences, the tests using the SRN simulations resulted in the difference *Consistent minus Inconsistent* activities being greater in the Experimental group simulation than in the Control group simulation. Activities increase with training in the SRN, whereas reaction times and errors should decrease (= faster and greater accuracy) with training in humans. As will be seen in the context of the experimental results, the SRN did not show generalisation to the sequences with 2 Cs.

The human experiments measured the *average reaction time* and *accuracy* differences between inconsistent and consistent sequences in the testing phase on the locations where consistent and inconsistent sequences diverged (i.e. on the final letter in the first sequence type and on the letter preceding the final letter in the second sequence type, which are the same locations where activity differences were measured in the SRN). Both reaction times and accuracy were assessed to rule out the possibility of a speed-accuracy trade-off. We expected the Experimental group to be slower (= higher reaction times) and more likely to make errors on these locations in the inconsistent sequences, i.e. *Inconsistent minus Consistent Reaction times* and *Inconsistent minus Consistent Error numbers* should be greater in the Experimental group than in the Control group. The results for reaction times, errors and the SRN simulations are displayed in Figure 3. In the analyses that follow we report tests on the reaction time followed by accuracy data.

Participants showed learning on the trained sequences with 1 and 3 Cs, $F_{1,28}$=9.01, p<.01 (individual comparisons for both 1 and 3 Cs: p<.05). Testing generalisation to 2 Cs revealed a significant result as well, $F_{1,28}$=2.83, p=.05. The results for the accuracy data complement those obtained with the reaction time measure, as there was a significant result for the trained sequences, $F_{1,28}$=4.44, p<.05 (with both individual comparisons pointing in the expected direction and the post-hoc test of the 3 C case revealing a significant result p<.05). The generalisation test for the 2 C case also showed a trend in the expected direction. Whilst the SRN simulations had also predicted learning on the trained sequences, $F_{1,28}$=8.89, p<.01 (individual comparisons for both 1 and 3 Cs: p<.05), they did not generalise to the novel sequences containing 2 Cs, $F_{1,28}$=.44, p=.52.

2.1.3 Discussion

Whilst the SRN performs well in simulating learning performance on the trained sequences, it does not model the generalisation to sequences with 2 Cs. Note that this is not a peculiarity of the SRN. As has been shown in Spiegel and McLaren (in press) as well as Spiegel and McLaren (2003), any associative model that successfully extracts the statistical regularities of the trained sequences would perform this way. Note that 2 Cs were never experienced in

training. Thus, the SRN is able to exploit this fact by developing a flip-flop representation that is tuned to 1 and 3 Cs (more details on this representation can be found in Spiegel et al. (2002)). Based on these results, one might be tempted to hypothesise that human performance on this task was not entirely associative. The possible reasons for human performance in this task will be discussed in the light of the other experiments.

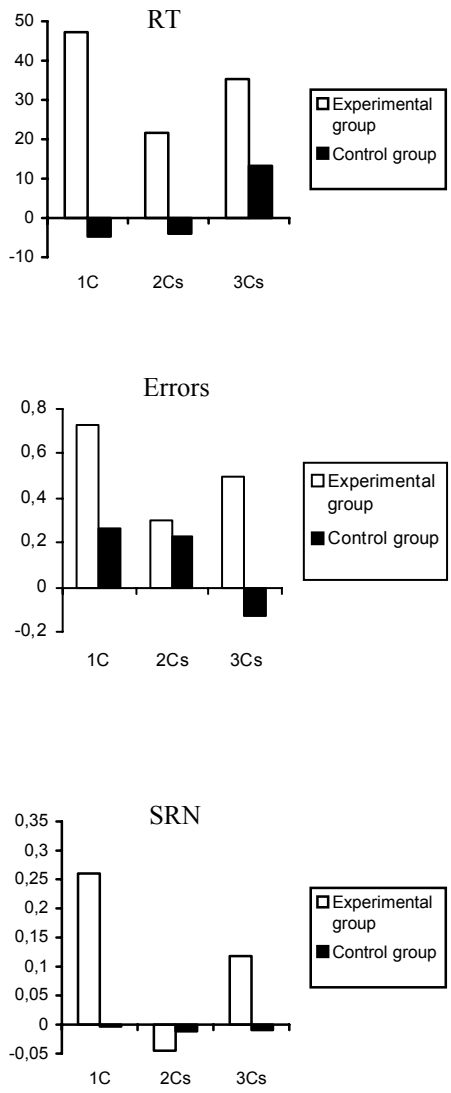


Figure 3: Results for both trained and novel sequences in Experiment 1. White/black bars stand for Experimental/Control conditions. (RT) Average Reaction-time-differences in humans in milliseconds (RT_inconsistent minus RT_consistent). (Errors) Average Error-differences in humans (Errors_inconsistent minus Errors_consistent). (SRN) Average Activity-differences in the SRNs (Activity_consistent minus Activity_inconsistent).

2.2 Experiment 2

In Experiment 2, experimental participants were trained on 1, 3, 5 and 7 Cs and their generalization to 2, 4 and 6 Cs was tested. The method was the same except where noted.

2.2.1 Method

Participants

There were 12 per group aged 16 to 38 years. Since we are dealing with more numerous and, at least in part, longer sequences, it was not possible to train and test the participants on the same day. Pilot work had indicated that this would have made our participants too tired. Consequently, we chose a 2 session experiment with 4 training blocks on the first day, two training blocks on the second day and 3 test blocks on the second day as well. During training, the Experimental group received 10 randomly selected presentations of each consistent sequence per block, whilst the Control group received 5 presentations of all consistent and 5 presentations of all inconsistent sequences. In each of the 3 test blocks, both Experimental and Control group received 3 presentations of all sequences including the novel ones to test generalisation performance.

2.2.2 Results

There was overall learning of the trained sequences (1, 3, 5 and 7 Cs) when considering the results of the error differences $F_{1,22}$=3.44, p=.0385. Though all of the individual post-hoc comparisons pointed in the expected direction, none of them reached significance when applying the somewhat conservative Bonferoni procedure (without the Bonferoni procedure the 5 C case would have reached significance at p<.05). The reaction time differences pointed in the same direction, but did not reach significance, $F_{1,22}$=1.48, p=.12.

There was no overall sign of generalisation to the novel sequences (2, 4, 6 Cs), neither when considering the error differences $F_{1,22}$=.82, p=.19, nor when taking into account the reaction time differences $F_{1,22}$=.12, p=.37. None of the post-hoc comparisons reached significance either.

In this experiment, the SRN models the learning of the trained sequences ($F_{1,22}$=8.27, p<.01) and the absence of generalisation ($F_{1,22}$=.1, p=.38) to the novel sequences rather closely (the discussion of Experiment 1 provides a possible reason why). Similar to the human data, none of the individual post-hoc comparisons reached significance apart from one: the trained 1 C case ($F_{1,22}$=7.92, p<.01). The results are displayed in Figure 4.

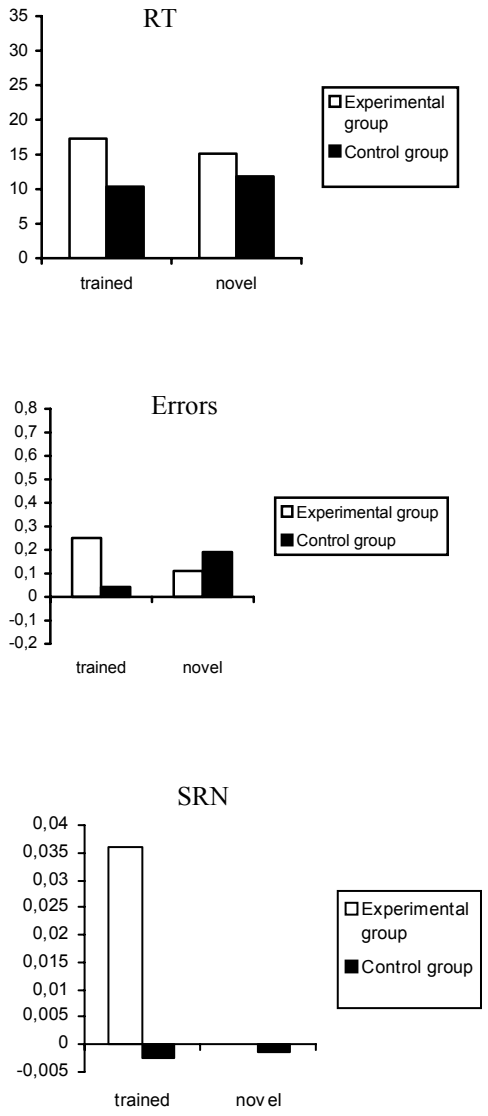


Figure 4: Results for both trained and novel sequences in Experiment 2. White/black bars stand for Experimental/Control conditions. (RT) Average Reaction-time-differences in humans in milliseconds (RT_inconsistent minus RT_consistent). (Errors) Average Error-differences in humans (Errors_inconsistent minus Errors_consistent). (SRN) Average Activity-differences in the SRNs (Activity_consistent minus Activity_inconsistent).

2.2.3 Discussion

The results of this experiment, where training was on longer and more numerous sequences, stand broadly in line with the associative model, whilst the results of the first experiment could not be explained by the associative model. It would be dangerous, though, to interpret the findings of the first experiment as non-associative and the ones of the second experiment as associative based on just two experiments and its simulations. Consequently, we aimed for further experiments by testing variants of this task. We believe that it is a much more critical test to develop specific hypotheses about the nature of the learning in a particular task, to make predictions based on these hypotheses and to finally test these predictions. This could be done by creating a computational model based on our hypotheses and let the model predict the results of future experiments. Those predictions could then be cross-validated with future experimental data. In the next section we will briefly discuss already existing computational models and make our case for designing a new model, which will be discussed subsequently.

3 The hybrid cognitive-associative model

In order to generate hypotheses about the nature of learning in our task, we had interviewed participants after the first two experiments. We asked whether they believed the signals had appeared in random or sequential order. In case they had realised a particular sequential structure, we asked them to describe this order or to show it on the screen. We are aware that post-test interviews are not considered a particular strong technique by implicit learning theorists, as there might exist a discrepancy between the learning mechanism during the task and what is actually verbalised after the task. In order to generate hypotheses that can subsequently be implemented in a computational model and tested in independent future experiments, however, we nevertheless regard them a useful addition. This belief is further supported by the fact that none of our participants in Experiment 2 (where we hypothesised associative learning) was able to verbalise the underlying rule. Though most subjects had realised that there was some kind of structure, much of the verbalised information was wrong. Because their learning is in line with the associative model, one could hypothesise that they did not find a particularly efficient cognitive strategy that would help them on this task. So their learning might have been driven by acquiring the statistical regularities of the sequences associatively, without any particular insight in the underlying structure of the sequential problem. In Experiment 1, where learning was on shorter and less numerous sequences, one might hypothesise that this problem should be easier to represent cognitively. Indeed, participants verbalised a lot more correct information about the sequences. In the Experimental group, where the previously mentioned

rule was present throughout all training trials, 2 out of 14 participants (14.29 percent) verbalised this rule right away[1]. They became aware of the symmetries of the number of As and Bs across the Cs, i.e. the fact that there were always as many As and Bs before and after the Cs. They realised that the sequences were analogous to each other with respect to the rule and they considered the rule more important than the intervening numbers of C. There were 4 participants (28.57 percent) who had realised repetitions such as BB or CCC, but not the underlying rule. 7 subjects (50 percent) said that the sequences flashed in the order ABCBA, but they were unaware of the number of repetitions on particular locations. The amount of verbalised information is not the only reason why one might assume that at least some people had formed an efficient cognitive representation throughout Experiment 1. When looking at the individual results of those 2 subjects who had verbalised the rule correctly, it becomes clear that they showed the largest effects in terms of both learning and generalisation. Note that the probability of picking the 2 subjects with the largest effects at random from a sample of 15 is 0.00952, i.e. $p<.01$. Consequently, there is a case for taking the correlation between verbalisation of the rule during the interview and the ability to generalise in the experiment seriously. Eliminating the data generated by these 2 participants from the statistical analysis still revealed learning of the trained sequences by the remaining subjects, but no longer significant generalisation to the 2 C sequences. The result of the remaining participants would be broadly in line with an associative explanation, but note that there was still a strong descriptive trend towards generalisation, which was absent in the SRN. So one could hypothesise that the verbalised information from the remaining subjects who provided incomplete yet correct information about the sequential structure had influenced generalisation performance as well. Something similar could be said about Experiment 2, where subjects provided a non-significant trend toward generalisation and verbalised some (though often incorrect) aspects about the sequential structure. In contrast, the purely associative model did not even show a trend towards generalisation (see Figure 4).

Our hypotheses were that humans are able to make use of both cognitive and associative mechanisms when performing on this task. When sequences are shorter and less numerous, the problem appears somewhat easier to represent and they are more likely to make use of cognitive learning strategies along with associative mechanisms. When sequences are longer and more numerous, the problem will be harder to represent cognitively so that participants are less likely to make use of cognitive strategies and more prone to associative learning.

The question is how to implement these hypotheses into a computational model that can then be tested empirically. To enhance clarity, we will give a brief overview before we explain the model in detail. The model consists of two components, an associative one and a cognitive one. Because the SRN seems to be a good associative model for part of the human performance on this task, the SRN will act as the associative subcomponent. Based on the data we gathered in the experiments and post-test interviews, we assume several aspects for the activation of cognitive mechanisms: We assume a threshold of activity until a stable cognitive representation is formed. As long as this threshold is not reached, the model will be driven by the associative subcomponent alone. There is a higher probability that this threshold is reached when the model is trained on less numerous sequences. For less numerous sequences it will take, on average, a shorter time until the same sequence recurs on the screen. There will also be less sequences competing for activation with each other. Along with many theories on learning, memory and cognition, we argue that activity decays if it is not refreshed. So the interplay of activation and decay will be in favour of activation when trained on less numerous sequences, because the recurring sequence presentations will have the consequence that the threshold is passed earlier. Hence, a cognitive representation will be formed more easily in this case. With more numerous sequences, on the other hand, there will be a stronger weight on decay. This will prevent a cognitive representation to be formed as easily as for fewer sequences. Apart from the number of sequences during training, length also plays a role, but things are more complicated in this regard. Both human experiments and the simulations with the SRN had shown that shorter sequences are not always learned more easily than longer sequences. In addition, sequence length is already part of the SRN-subcomponent and, as will be seen later, this associative subcomponent interacts with the cognitive part. Consequently, we did not implement another parameter controlling sequence length into the cognitive part of the model.

Having provided a brief overview, we will describe the model in greater detail now. Being one of the most cited papers in psychology (Marcus, 2001, p. 25), the associative learning mechanisms of the

[1] Although there were 15 participants in the Experimental group, only 14 agreed to take part in the interview. The interview result of the 15th participant would have been less relevant anyway, as the results of this participant did not indicate learning.

SRN (Elman, 1990; Rumelhart et al. 1986) are well-known in the Experimental Psychology community (and were briefly described above). Of particular interest will be the cognitive part of the model and how it interacts with the SRN-subcomponent. Similar to the SRN, the cognitive part computes an activity for every step in the sequence. This holds true for each sequence that occurs during training. Say there are 4 sequences, ABCBA, ABBCBBA, ABCCCBA, ABBCCCBBA. In this case each letter is represented by a different, position-specific activity, e.g. all final letters A would hold a different activity value depending on how many times the particular sequences had been presented during training and in which order these sequences had been presented. If one sequence is presented and then it takes 4 sequences until this sequence recurs, the particular activity of the final letter in our example will hold a different value than the activity of the final letter in another sequence that is presented twice in a row. If a sequence is presented twice in a row, the activity of this letter will not be influenced by decay, whilst decay has an influence in the other sequence. However, the cognitive part does much more than that. As the interviews had revealed, people are able to form analogies between sequences if these share an underlying structure. In the sequences of our experiments, there is a high overlap between sequences. As we hypothesised based on the interviews where participants verbalised common structure that even led to the formation of a rule in a minority of subjects, those analogies help people form a cognitive representation. Consequently, the cognitive part of the model computes more than just activities of the individual sequential elements. There are relations between the letters, e.g. the As and Bs in those sequences are symmetric across the Cs. Some elements repeat, such as the BBs and the CCCs. As a result, additional activities are computed for every symmetry in the model, more activities are computed for every repetition in the model etc.

Now we have symmetries and repetitions in all 4 sequences and the question is how they can be seen as analogous to each other. Similarly, in other learning tasks there might be hundreds of different sequences and the question could be how they can be related to each other. This is only possible if the person spots which sequences share a common structure. The common structure apart from the previously mentioned symmetries and repetitions is the order of screen flashes, which was ABCBA in all 4 sequences. Combining the order of screen flashes with symmetries and repetitions helps seeing analogies in terms of symmetries and repetitions between the sequences, because it frees the individual sequences from their position-specific activation values. This will be clarified with an example. All 4 sequences' final letter is an A, but this letter is located on sequence position 5 in ABCBA, on sequence position 7 in ABBCBBA and ABCCCBA and on sequence position 9 in ABBCCCBBA. When taking into account the order of screen flashes (ABCBA) and neglecting repetitions, all 4 would be on position 5. Likewise, when considering the BBs in sequences ABBCBBA and ABBCCCBBA, there is a BB on positions 2 and 3 in both sequences, but another BB on positions 5 and 6 in the first example and on positions 7 and 8 in the second example. When taking into account the order of screen flashes, however, the BBs would be on positions 2 and 4 in both sequence types. This makes it easier to spot the common structure. In fact, none of the participants in our experiments had counted the exact positions. Rather, they had often verbalised the order of screen flashes and explained where repetitions and symmetries were located with respect to this order. So repetitions and symmetries are relative with respect to the order of flashes. When being given a new yet overlapping sequence that had not been experienced during training before, such as ABBCCBBA, the model would be able to perform generalisation when considering repetitions and symmetries with respect to the order of flashes rather than with respect to their absolute position in the sequence. We already know that the SRN would not generalise to this sequence (the reasons were stated above). The question is how the cognitive part of the model manages this problem. In order to generalise, it is vital that the cognitive part is able to recognise the analogy between the novel sequence and the trained sequences. This is only possible if at least one of the trained sequences is represented cognitively, i.e. if the activities for the previously mentioned symmetries, repetitions and the activities representing order of screen locations all pass a particular threshold value.

When simulating human performance with a computational model, we consider it necessary that this threshold value is a constant value that is not changed from simulation to simulation. Changing parameter values between experiments would make it easier to simulate experimental data. To give our model the hardest possible test, we chose an a-priori threshold value of 0.95 that was subsequently applied to all our simulations.

When implementing the model's mechanisms that helped to find symmetries, repetitions and the order of screen flashes, we drew a lot of inspiration from the codelet-type approach proposed by FARG,

the Fluid Analogies Research Group (Hofstadter, 1995; Marshall, 1999; Mitchell, 1993). In their approach, each codelet represents a small piece of code. Combining all codelets in a joint effort will come up with a structure for a previously unstructured problem. One single codelet just searches for all symmetries, another codelet searches for all repetitions, etc. The FARG approach is not directly transferable to our task (e.g. their aim was to model perception rather than learning and hence their models neither related to cognitive nor to associative learning). Computationally, however, codelets of this type inspired us how to develop part of the program code for the cognitive component in our model. Codelets look through all sequences and find all symmetries, repetitions, etc., regardless of the sequential task's size. How these codelets are programmed is probably of little interest to the Experimental Psychology community, but detailed instructions can be found in the previously mentioned papers of the Fluid Analogies Research Group and in Spiegel (2002). The learning taking place based on the interplay of activation and decay will be of greater interest here. Therefore, the learning function of the cognitive part will be considered next. This part is inspired by findings reported in the literature on human memory, interference and decay, e.g. Baddeley (1990). We argue that the cognitive part is necessary not only because the associative model does not fully account for our experimental findings, but also because curves combining memory, interference and decay do not have the shape of those in connectionist models (Baddeley, 1990). These frequently follow linear or logistic activation functions. When trying to remember new information, for example, there is a lot of decay, partly caused by interference with other information and partly caused by the fact that new information is not consolidated in memory yet. Once information gets consolidated in memory through rehearsal, decay becomes less. This does not mean, however, that this information cannot be lost, i.e. decay can still cause this information to be forgotten. Once previously well-established and then forgotten information is refreshed a single time, however, it can become very active again. Refreshing this information a single time can re-establish the memory trace. Once the trace is re-established, decay has a different impact on this memory trace than it used to have when this trace was newly established. This is based on the fact that a memory trace that had been activated many times in the past decays more slowly than a newly established trace. As argued in Baddeley (1990), combining these differential parameters (that all seem to play a role in cognitive processing) will require the consideration of other equations and computational models than the previous ones. Figure 5 displays one attempt of such an equation (above) and its resulting curve (below).

$$a_{i+1} = 1 - \left[\frac{1}{1+\beta\left(e^{\left(\left(\frac{\beta}{1-a_i} \right)^{\frac{1}{2}} \right)} \right)} \right] + \lambda \left[a_i - \frac{a_i}{1+\eta\left(\frac{1}{\varsigma} \right)} \right]$$

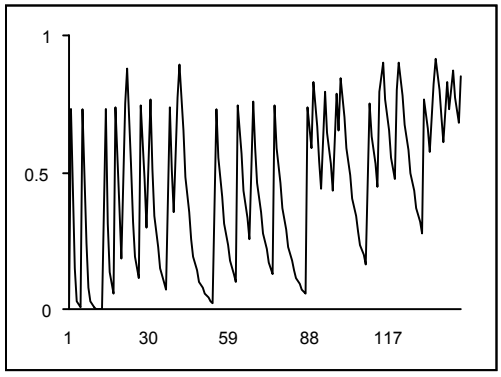


Figure 5: The equation of the cognitive part's activation function (above) and its resulting curve (below), with the Y-axis displaying activity and the X-axis displaying the trials with activation (curve going up) and decay (curve going down).

In the equation, α_{i+1} stands for the activity that resulted from either excitation or decay of the memory trace α_i. The equation further contains the exponential, i.e. Euler's number *e* (2.718), the number of previous excitations η of this particular memory trace and the number of competing sequences ς that are experienced during the training phase. The parameters β and λ take on values of 1 (excitation) or 0 (decay), with β being 1 during excitation and 0 during decay and λ being 0 during excitation and 1 during decay. Thus, the activity of a memory trace ranges between values of 0 and 1. The power of *e* would be undefined for an activity value of $\alpha_i = 1$, which is prevented by subtracting 0.001 from 1 (resulting in an activity value of 0.999 in this case). As can be seen at the start of training, a newly established memory trace will decay fast when it is not

activated again. The fact whether the trace is activated again is based on the probability of its particular stimulus appearing once more at the next sequence presentation. If the probability is low, decay will be catastrophic as it takes, on average, a longer time until this particular stimulus is presented again. Another reason why decay will be large is that the number of previous activations of this stimulus will be small at the start of training. The more often a stimulus has been presented in the past, the less influential the decay. As a result, the activation function will be applied anytime a memory trace gets activated or decays. Eventually, several parts of the sequences will be represented cognitively because their activity values pass the previously mentioned threshold of 0.95. All information that lies above threshold is permitted to look for sequences with analogical structure.

As previously mentioned when referring to the interviews, not all aspects were verbalised the same number of times. Symmetries, for instance, were only verbalised by 2 participants (14.29 percent), repetitions by 4 participants (28.57 percent) and the order of screen flashes by 7 participants (50 percent). So people were more likely to become aware of some aspects (e.g. repetitions) than of other aspects (e.g. symmetries). In order to take into account these proportions, appropriate probabilities were assigned to those aspects, e.g. each symmetry got activated with the previously mentioned activation function in 14.29 percent of the cases, each repetition got activated in 28.57 percent of the cases. In terms of the order of screen flashes, the activation function was applied in 50 percent of the cases a flash changed the screen location (e.g. from A to B, from B to C, from C to B or from B to A). Another reason for assigning these probabilities is that people do not typically become immediately aware of every symmetry, repetition or change of screen location. Since codelets work by recognising all of these in 100 percent of the cases, a model based on codelets alone, i.e. without taking into account proportions, would be an inappropriate model of human performance on this task.

Having described the activation function, the question is how the sequences with activity values above threshold can be combined with each other. Let us imagine we train the model on the four sequences ABCBA, ABCCCBA, ABBCBBA and ABBCCCBBA. Let us further assume that these sequences are represented cognitively in terms of the previously explained symmetries, repetitions and order of screen locations, i.e. activities representing these aspects all have passed threshold values. A new sequence with overlapping sequential structure such as ABCCBA can be seen as analogous to these trained sequences. The activity values of this new sequence are, of course, unknown. Now an example will be provided how the activity values of the trained sequences are combined in order to get activity values for the new sequence. Up to the step where only ABCC is presented to the model, it would assume that this is the trained sequence ABCCCBA. If one C less had been presented (ABC), there would be 2 candidates among the trained sequences: ABCBA and ABCCCBA. Where things get interesting is when the sequence ABCCB is presented to the model, as there are no candidates among the trained sequences that have this particular structure. So let us assume an incomplete sequence with the structure ABCCB is presented to the model. What does this new sequence have in common with the trained ABCBA, ABCCCBA, ABBCBBA and ABBCCCBBA sequences? It has symmetries between the Bs across the Cs. What do all the trained (old) sequences have in common? They not only have symmetries between the Bs across the Cs, but also carry symmetries between the As across the Bs and Cs. Therefore, all four trained sequences predict a symmetry between the As to follow a symmetry between the Bs (irrespective of the fact whether there is a symmetry between the individual Bs or the BBs). The incomplete sequence ABCCB has an A as its first letter. As a consequence, the addition of another A would be the only possible answer in terms of analogy-making to the other sequences. It would not be possible to add another B, because all Bs are symmetric across the Cs in the trained sequences and a sequence ABCCBB would partly violate this symmetry. But how would the model settle on a solution to predict the letter A in the sequence ABCCB_? It would make analogies to all four trained sequences, by calculating numerical values for all of them. Consider the first sequence ABCBA. Until the penultimate letter in its sequential order, this sequence is equivalent to the sequential order of screen flashes in the incomplete ABCCB sequence, which has the order ABCB (repetitions on letters are ignored when expressing the order of screen locations. What has to be kept in mind is that all trained sequences in our example have ABCBA as the order of screen locations). The first numerical value would therefore be the activity of the penultimate letter (shown in *Italic*) in this ABC*B*A sequence (recall that all these letters have activities greater than the threshold .95). What follows this screen location is a switch to screen location A, which also has an activity greater than .95. In addition, this screen location carries another important aspect. There is the previously

mentioned symmetry between the As across the Bs and the C (*A*BCB*A*). This symmetry is also expressed in terms of a numerical value greater than .95. Both the first and the second numerical value are then multiplied with each other.

The same would be done with the next analogous sequence, i.e. the ABCCCBA sequence. It would again take the activity of the penultimate letter in its order of screen locations ABC*B*A and multiply this activity with the other activity, which is again the symmetry between the As. It would then do the same for the ABBCBBA sequence and the ABBCCCBBA sequences. Note the fact that two Bs appear before and after the Cs does not affect the analogies. As a result, we have four numerical values, each representing an analogy to one of the four trained sequences. We will later see how these analogies help the novel and incomplete sequence ABCCB to come to the solution ABCCBA. Each of those 4 activities, which will be termed analogy values α_j (with j ranging from 1 to 4) consists of the product of two activities (first numerical value representing screen location times second numerical value representing symmetry). There are further examples, e.g. the new sequence ABBCCB-? could draw analogies to the 2 sequences ABBCBBA/ABBCCCBBA (in this example j ranges from 1 to 2), but now the 2 analogy values α_1 and α_2 would be calculated by the product of three activities (first numerical value representing screen location ABC*B*A, second numerical value representing repetitions ABBC*BB*A/ABBCCC*BB*A, third numerical value representing symmetries A*B*BCB*B*A/A*B*BCCCB*B*A). Formally, α_j can be described using equation 1.

Equation 1:

$$\alpha_j = \phi \cdot \prod_m \nu_m$$

The activity of the penultimate letter in the order of screen locations ABC*B*A is denoted by the symbol ϕ. The numerical values (e.g. activities representing symmetries or repetitions) are denoted by the ν. In our example, there were never more than 2 ν values (symmetries or repetitions), hence the subscript *m* was 2 at maximum.

Having shown how the numerical analogy values are calculated, we have yet to show how they are combined with each other in order to predict the next step in the sequence. This will also involve a demonstration how the cognitive part of the model interacts with the associative SRN. How both parts of the model interact with one another and jointly produce the output is specified in Equation 3. First, however, we need to explain how the SRN produces the output (Equation 2).

Equation 2:

$$out_i = \frac{1}{1 + e^{-\left(\sum_h \omega_{ih} o_h + \zeta_i\right)}}$$

Equation 2 represents the logistic activation function for a particular output unit in the classical backpropagation algorithm, which is applied in the SRN. This function is applied to all output activities of the SRN and ensures values within the range 0 to 1. The Σ-sign is the sum over the weights (ω_{ih}) leading to this particular output unit from each hidden unit o_h plus the bias ζ_i on this particular output unit. How this squashing function was derived can be found in Rumelhart et al. (1986). The interaction between the cognitive part and the associative SRN is nothing more than an elaboration of Equation 2 (which was the associative part alone) and is described in Equation 3. The first part of the numerator in Equation 3 is entirely the same as Equation 2. The value N in the denominator indicates how many analogies in other sequences could be found. Take one of the examples from above. We had assumed four sequences to be analogous to the novel (and incomplete) sequence ABCCB. Since there are four sequences, N would take the value 4. If there had not been any analogies (= no cognitive representation, e.g. due to activities below threshold), N would have taken a value of 0 and the equation would reduce to that of an ordinary SRN (= entirely associative outcomes). In our example, however, the cognitive part of the model was strong enough to perceive analogies to other sequences (= the cognitive part had indicated values above threshold to form analogies). As a result, all the analogies will be incorporated in the prediction of the following sequential element. This is done in the second part of the numerator in Equation 3. How the analogy value α_j is calculated has already been explained in Equation 1. This value (in our case the values for 4 analogies, hence j=1 to j=4) will be combined with the related outputs of the SRN (hence o_{jh} instead of o_h). These are the SRN-generated outputs of the 4 sequences to which analogies were formed. The reason why α_j stands in the numerator is that normal backpropagation has the number 1 in the numerator (see equation 2) and multiplying the value α_j with 1 equals α_j.

Equation 3:

$$out_i = \frac{\dfrac{1}{1+e^{-\left(\sum_h \omega_{ih} o_h + \zeta_i\right)}} + \sum_j \dfrac{\alpha_j}{1+e^{-\left(\sum_h \omega_{ih} o_{jh} + \zeta_i\right)}}}{N+1}$$

The Σ-sign over the analogies j indicates that all analogies having been formed by the cognitive part are taken into consideration, and in order to ensure that the output of the model is bounded between 0 and 1, the number of analogies appears in the denominator. Here, the number of analogies is represented by N, which carries a value of zero in case no analogies are formed (= the output reduces to the one of an SRN, as α_j would also be zero in this case).

4 Simulations of Experiments 1 and 2

Because the model was inspired by the first two experiments, the first test will be to check whether it successfully simulates learning and generalisation in both experiments, i.e. will the cognitive part of the model play a sufficiently large role to simulate generalisation in Experiment 1? Likewise, will the associative part of the model dominate in Experiment 2? The description of these tests will be very brief, as successful simulation in these cases would not mean much. A more powerful test will certainly be to cross-validate the model with novel experiments (i.e. ones that have not been applied to design the model). These will be discussed in the next section.

When considering the trained sequences with 1 and 3 Cs, the analysis of variance on the simulation results for Experiment 1 revealed a significant overall effect, $F_{1,28}$=14.45, p<.001 (individual comparisons for both 1 and 3 Cs: p<.01). The critical comparison, i.e. whether the model generalises to the novel sequences with 2 Cs, produced a significant difference between Experimental and Control groups, $F_{1,28}$=7.78, p<.01. According to the overall findings, the model simulated human performance in Experiment 1 better than a purely associative model such as the SRN, which had not shown generalisation to the 2 C case.

Referring to the simulation of Experiment 2, there was significant learning of the trained sequences, $F_{1,22}$=7.23, p<.01, but no generalisation to the novel sequences with 2, 4 and 6 Cs: $F_{1,22}$=.16, p=.35 (with all individual comparisons p>.1). The simulation stands in line with the human data as well as the associative model. When inspecting the previously mentioned activity values of the cognitive part, it became clear that they lay below threshold to an extent that made it impossible to form a correct cognitive representation of the complete structure. As had been described earlier, this seemed similar to the representation of our participants. The simulation results of both experiments are displayed in Figure 6.

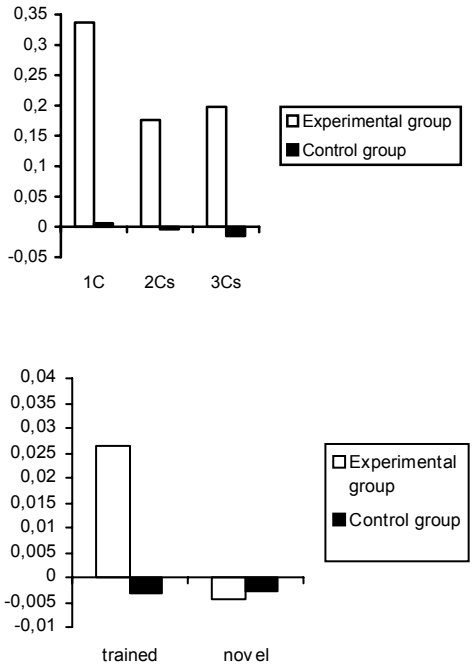


Figure 6: Average activity differences between Experimental and Control group in Experiment 1 (above) and Experiment 2 (below). The hybrid model simulated both learning of 1 and 3 C sequences and generalisation to 2 C sequences in Experiment 1, and learning of trained sequences (1, 3, 5 and 7 Cs) with no generalisation to novel sequences (2, 4 or 6 Cs) in Experiment 2.

5 Further experimental data and simulations

5.1 Experiment 3

Thus far, both experiments contained 1 and 3 Cs during training. The question arises how people would perform if we trained them on 2 and 4 Cs and tested their generalisation performance to 3 Cs.

5.1.1 Method

The experimental procedures of Experiment 3 were the same as in the first experiment, with the only exception that a total of 36 subjects aged 16 to 39 took part in this experiment. We also chose 36 participants for this and all further experiments. In this respect, Experiments 1 and 2 helped us to find out about a good sample size. The reason we went for a slightly larger sample size was based on exploratory data analyses, e.g. among other parameters, these analyses had indicated that distributions represented the bell shape much better with slightly larger samples.

5.1.2 Results

In order to test our hybrid model, we made predictions first. The cognitive model predicted overall learning of the trained sequences, $F_{1,34}$=15.88, p<.001, with post-hoc tests (Bonferoni) on individual 2 and 4 C sequences being significant at p<.05. The model also predicted a significant effect for the 3 C sequences, $F_{1,34}$=15.54, p<.001 (Figure 7b). This finding stands in contrast to the purely associative SRN, that had predicted learning of the trained sequences, $F_{1,34}$=10.99, p<.01, but no generalisation to 3 Cs, $F_{1,34}$=1.52, p=.11, where the Control group even had a slightly larger activity difference than the Experimental group (Figure 7a).

In the human experiment, we turn to the reaction times first. There was a significant result for the trained sequences, $F_{1,34}$=7.58, p<.01. Considering the individual comparisons, the 4 Cs case revealed a significant result (p<.01), taking into account the Bonferoni specifications, whilst the 2 and 3 C sequences showed non-significant results pointing in the same direction (Figure 7c). Considering the error differences next, the trained sequences revealed once more a significant difference between Experimental and Control group, $F_{1,34}$=4.53, p<.05. In terms of the individual comparisons, the 2 Cs case revealed a significant result according to the Bonferoni specifications, $F_{1,34}$=4.72, p=.019. The error differences for the other trained sequences with 4 Cs were not significant, but pointed in the expected direction as well. Referring to the novel sequences with 3 Cs, there was a significant effect of generalisation performance, $F_{1,34}$=5.62, p=.012. The error data are displayed in Figure 7d. To summarise the results, there was evidence of learning the trained sequences with 2 and 4 Cs as well as generalising to the novel 3 Cs sequences.

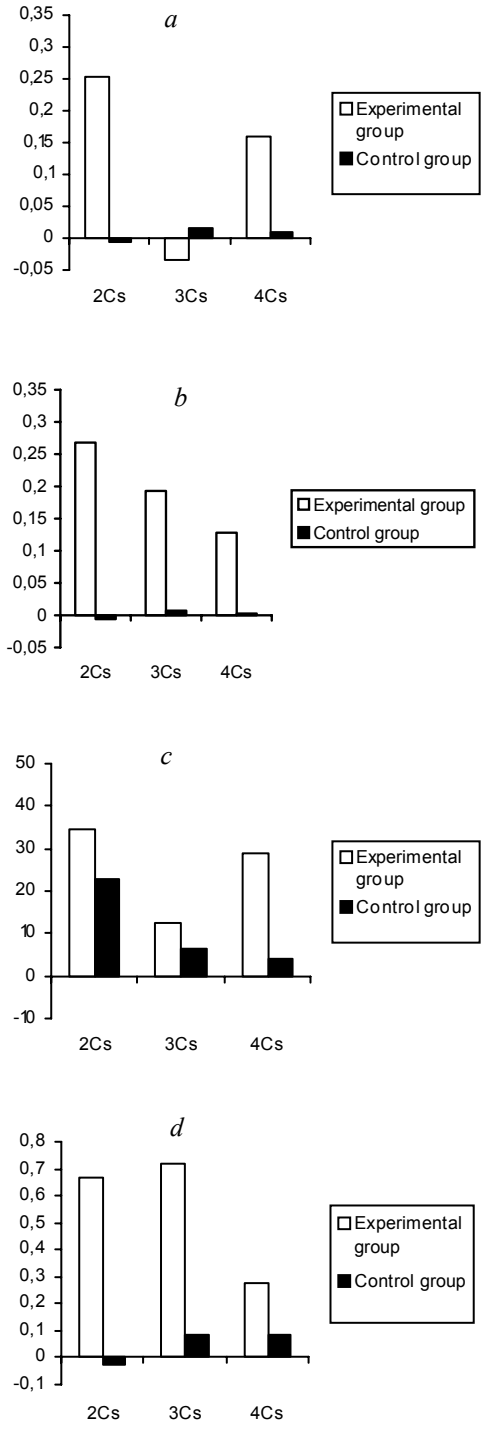


Figure 7: Experiment 3. Training was on 2 and 4 Cs. Generalisation to 3 Cs was tested. Making use of average activity differences, predictions by the SRN (Figure 7a) are contrasted with predictions by the Hybrid Model (Figure 7b). In the human experiment, average reaction time differences (Figure 7c) are displayed along with average error differences (Figure 7d).

The evidence for this consisted of all reaction time and error differences pointing in the same direction and having significant results in at least one of the dependent variables. Both generalisation and learning were predicted by the hybrid model, whilst the SRN had only predicted learning. After having had a closer look at the hybrid model by inspecting the previously mentioned activity values of the cognitive part, it became clear that they lay above threshold. This allowed the model to draw analogies between the novel 3 Cs sequences and the trained 2 and 4 Cs sequences.

5.1.3 Further tests of the model's predictions

Following Experiment 3, we applied the model to predict the results of several other experiments. In both Experiments 4 and 5, we trained the model on 1, 3 and 5 Cs. Generalisation to 2 and 4 Cs was tested in Experiment 4, whilst generalisation to 6 and 7 Cs was tested in Experiment 5. In Experiment 6, training was on 1, 3, 4 and 6 Cs, whilst generalisation to 2 and 5 Cs was tested. Experiment 7 was similar to Experiment 2 in so far as training was on 1, 3, 5 and 7 Cs, but generalisation was tested to 2 and 6 Cs only. Experiment 7 acted as a Control for Experiment 6 (i.e. same sample size, overall generalisation was tested based on a blend of 2 sequence types (2 and 6 Cs) instead of 3 (2, 4 and 6 Cs)). To anticipate a bit, there was evidence for associative performance in all these experiments. The associative sub-component of the model, i.e. the SRN, succeeded in predicting both learning and generalisation in all these experiments. The simulation data of the SRN and the experimental results are described in a separate manuscript on associative sequence learning in humans (Spiegel & McLaren, in press). Because the paper presented here is on the hybrid cognitive-associative model, the simulation results of the hybrid model will be demonstrated (they are not contained in Spiegel and McLaren (in press), where only associative simulations with the SRN are demonstrated).

In Experiment 4, the hybrid model predicts significant learning of the trained sequences with 1, 3 and 5 Cs, $F_{1,34}$=6.43, p<.01. There was no overall effect of generalisation to the novel sequences with 2 and 4 Cs, $F_{1,34}$=.15, p=.35. Likewise, none of the individual comparisons revealed a significant result towards generalisation, neither the 2 Cs case, $F_{1,34}$=.03, p=.43, nor the 4 Cs case, $F_{1,34}$=.87, p=.18. These simulation results were confirmed with experimental data. When inspecting the previously mentioned activity values of the cognitive part, it became clear that they lay below threshold, which prevents analogy-making to the successfully learnt trained sequences. As the interviews with participants in this experiment later revealed, people had not verbalised any analogies either (even after prompting them about possible analogies).

In Experiment 5, the model once more predicted learning of the trained 1, 3, 5 Cs sequences, $F_{1,34}$=8.56, p<.01. This was the first time, however, that generalisation outside the training range (which was 1 to 5 Cs) was tested. Thus far, generalisation was only tested to numbers of C that were larger than the smallest number of C elements (1) during training and smaller than the largest number of C elements during training. The model predicts generalisation to the novel sequences with 7 Cs, $F_{1,34}$=4.13, p=.025, and the absence of generalisation to the novel sequences with 6 Cs, $F_{1,34}$=2.29, p=.07. The 6 C case comes close to showing a significant result in the expected direction. In contrast, a purely associative SRN shows no trend in the expected direction for novel sequences with 6 Cs. Having analysed the activities in the cognitive part of the hybrid model, it became clear that the positive trend had been caused by a small number of activities passing threshold values, hence opening the door for analogy-making. This number, however, was not large enough to cause significant generalisation due to the cognitive part of the model (even though we had performed one-tailed tests). Although the model simulates other experimental data better than in this case, its predictions are in line with the experimental data.

In Experiment 6, training was on 1, 3, 4 and 6 Cs, whilst generalisation was assessed on sequences with 2 and 5 Cs. The model predicted significant learning of the trained sequences, $F_{1,34}$=27.08, p<.001. There was significant overall generalisation to the novel sequences, $F_{1,34}$=8.63, p<.01. Among the individual comparisons, both the 2 Cs case, $F_{1,34}$=5.44, p=.013 and the 5 Cs case, $F_{1,34}$=9.94, p=.0015 revealed significant generalisation performance. Providing training on both odd and even numbers of C revealed generalisation to novel odd (5) and novel even (2) numbers. These predictions were confirmed in the human experiments. Had the previously mentioned threshold values been passed, the results would have been similar, as analogy-making to successfully learned (trained) sequences would have resulted in successful generalisation as well. In this experiment, however, the activities of the cognitive part of the model essentially stayed below threshold.

As in Experiment 2, training in Experiment 7 was on 1, 3, 5 and 7 Cs (odd numbers only). The generalisation test was on sequences with 2 and 6

Cs. The model predicted successful learning of the trained sequences, $F_{1,34}$=8.73, p<.01 and the absence of generalisation to 2 and 6 Cs, $F_{1,34}$=.17, p=.34. Neither the 2 C case, $F_{1,34}$=.23, p=.32, nor the 6 C case, $F_{1,34}$=.01, p=.47 revealed evidence of generalisation. These predictions are in line with the associative perspective and the experimental data (Spiegel & McLaren, in press). They are counter-intuitive, because generalisation is dependent on the intervening number of C elements. These intervening elements were irrelevant to application of the rule governing the sequences, i.e. the rule that B flashes are the same before and after the C flashes. Yet, the intervening flashes heavily impact on the model's predictions and on human performance. If the participants had only been exposed to sequences with an odd number of Cs in training, they learned the trained sequences, but were unable to generalise to novel sequences with an even number of C elements. If there had been both odd and even numbers of Cs during training, they learned the trained sequences and generalised to novel sequences with both odd and even numbers.

Thus far, we had only presented experiments where the model and participants succeeded in learning the problem. There was no presentation of an experiment where the model or people failed to learn the problem in the first place. Such a case does exist. When training the model on 2, 4, 6 and 8 Cs and testing generalisation to 3, 5 and 7 Cs, neither the SRN, nor our hybrid model, nor participants showed any signs of learning or generalization. This finding is even robust to parameter variations, as giving the model more power through longer training or raising the number of hidden units in the SRN-subcomponent showed no improvement. The data of these findings are already published in Spiegel and McLaren (2003).

6 General Discussion

The generalisation present in Experiment 6 and absent in Experiment 7 is in complete agreement with the predictions of our model, which were predominantly driven by the SRN-subcomponent. In the human case, however, it may perhaps be explained by the observation that the trained sequences with 5 and 7 Cs were slightly longer than the ones with 4 and 6 Cs, or perhaps because generalisation to the novel (but shorter) sequence with 5 Cs is easier than that to the novel sequence with 6 Cs. Equally, it could be that in Experiment 6 there was more training on shorter and thus less complex sequences (if we classify sequences with 1, 3 and 4 Cs as short ones and 5, 6 and 7 Cs as long ones). These potential confounds are controlled by Experiments 4 and 5, where training was on 1, 3 and 5 Cs. These sequences experienced during training were, on average, shorter than for the 1, 3, 4 and 6 Cs problem. Furthermore, the use of 2 and 4 C element generalisation tests in Experiment 4 meant that the novel test sequences were also shorter than those used for Experiment 6. Because we still found no generalisation to even numbers of C elements in Experiment 4, we were able to exclude the potential confound that the length or the amount of training on shorter sequences was responsible for generalisation to even numbers in Experiment 6 and the absence of generalisation in Experiment 7. Indeed, the striking feature of Experiment 7's results is how similar they are to Experiment 4, despite the changes in design between the two experiments. Also noteworthy is the fact that in both experiments, the pattern of results obtained with human participants is closely modelled by our hybrid model, which was driven by the associative sub-component in these experiments. We think that the contrasting generalisation shown by participants trained on the 1, 3, 5, 7 Cs and 1, 3, 4, 6 Cs problems is due to the different associatively-generated representations formed as a consequence of training. The results of Experiment 5 are of particular interest, because they show that humans as well as the model do not generalise to the 6 C case, but do generalise to 7 Cs after training on 1, 3 and 5 Cs sequences. It must be acknowledged that finding generalisation to the sequence that is more distant from the trained sequences in the absence of any generalisation to a closer sequence is quite remarkable. Taken in combination with the results from Experiment 6, these findings establish that generalisation of some kind can be obtained with either training sequence. This makes an explanation of the results based on non-transferable or weaker learning occurring with one training sequence rather than the other untenable. Instead, the pattern of generalisation obtained is simply that predicted by the model throughout. This also applies to those tasks where training was on shorter and less numerous sequences. When the cognitive part of the model played a large role by producing activity values above threshold, the model generalised to 3 Cs sequences after having been trained on 2 and 4 Cs. We believe that this combination of predictions and its confirmation by the human experiments are quite remarkable, for the model's parameters were not changed a single time. Instead, the model's implementation that was originally derived from the results produced in Experiments 1 and 2, led to successful predictions in the following experiments. It

not only predicted successful learning and generalisation, but also learning in the absence of generalisation as well as the total breakdown of learning. We believe that our model therefore contributes to a better understanding of how cognitive and associative processes interact in this type of sequence learning problem. Certainly, there need to be more experiments to better understand how these cognitive and associative processes interact in detail. In spite of its predictions, which were correct in so far as the model's results were successfully cross-validated with further empirical tests, it might turn out that the model needs to be modified based on future data. This is particularly important with respect to tests of the cognitive sub-component of the model. With the exception of Experiment 3, more experiments were predicted by the associative sub-component of the model rather than by an above-threshold influence of the cognitive part. We nevertheless believe that it is necessary to introduce this model to the Experimental Psychology Community at this stage, for its results have now been successfully applied to a total of 8 experiments (which implies over 500 hours of testing human subjects). Moreover, it provides an alternative to present hybrid computational models. Take ACT-R (e.g. Anderson, 1993) or Clarion (Slusarz & Sun, 2001; Sun et al. 2001), which are powerful approaches in Cognitive Science and have been adapted to psychology experiments. Apart from learning, the ACT-R architecture has been applied to problems as diverse as driving and flying behaviour, graphical user interfaces, programming, video games, just to name a few. The Clarion architecture has also been applied to the management of a sugar production factory or a minefield navigation task. The variety of successful applications of these models had the consequence that they were held relatively general. Their cognitive component in the form of a symbolic production system requires telling the model the rules it has to adhere to. These two models thus act on a higher level, being more general and thus applicable to a wider range of problems rather than being more specific to a particular learning task. Because our aim was to understand a particular learning task, we aimed for a model that acts on a lower level. We therefore did not tell the model any particular rules, nor did we adjust the model's parameters once we had specified parameters following the first 2 experiments. In this particular class of serial reaction time experiments, we believe that our model has the advantage that it might contribute to a more detailed analysis of how cognitive and associative processes interact, for it was designed with this particular problem in mind, whilst the other models were adjusted to these tasks after they had been designed with other tasks in mind. Our particular experiments, however, might not be of general interest across several scientific disciplines, which is where the strength of ACT-R and Clarion lies. Nevertheless, our sequence learning task has found some resonance in the psychology community (e.g. Russell, 2004, pp. 347-349). The reason we developed a new model in the first place was because taking one of these existing models, telling these models the rules they have to adhere to and making parameter adjustments would have been a less stringent test criterion.

We do not believe that our model's cognitive mechanism is entirely correct (we even believe that parts of it will be falsified in the long run). However, we believe that it is important to implement models that are based on present knowledge of experimental psychology, and that can be critically evaluated on present and future data. Future knowledge from psychology experiments will probably contribute to modifications or entirely new models. Our model is not neurally plausible either, as neuroscientific findings on those cognitive mechanisms are still lacking and the model's associative subcomponent relies on backpropagation, which contradicts current neurobiological knowledge. On the other hand, the first author of this paper compared the associative sub-component with a class of models that are considered neurally plausible according to Koerding and Wolpert (2004), and found no difference in terms of the overall results apart from the duration it takes the models to converge on this learning task (Spiegel, 2002).

Referring to psychological plausibility, the error correcting training algorithm present in the associative sub-component is psychologically plausible to some extent, as participants receive feedback after every response, indicating whether they have pressed the right or the wrong key. In summary, this paper mainly deals with a psychologically inspired model that will hopefully be applied to simulate the interaction of cognitive and associative processes in future serial reaction time experiments. When naming our model, we have chosen the name SARAH, an acronym for Sequential Adaptive Recurrent Analogy Hacker (the original, neutral connotation of the word hacker is implied here, i.e. someone who enjoys perfecting an algorithm).

References

Anderson, J.R. *Rules of the Mind*. Hillsdale, N.J.: Erlbaum, 1993.

Baddeley, A. *Human Memory: Theory and Practice*. Hove, UK: Erlbaum, 1990.

Cleeremans, A. *Mechanisms of implicit learning. Connectionist models of sequence processing*. Cambridge, MA: MIT-Press, 1993.

Elman, J.L. Finding structure in time. *Cognitive Science,* 14: 179-211, 1990.

Hofstadter, D. *Fluid Concepts and Creative Analogies.* New York: Basic Books, 1995.

Koerding, K.P. and Wolpert, D. Bayesian Integration in Sensorimotor Learning. *Nature,* 427: 244-247, 2004.

Lewicki, P., Czyzewska, M. and Hoffman, H. Unconscious Acquisition of Complex Procedural Knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 13: 523-530, 1987.

Marcus, G.F. *The Algebraic Mind. Integrating Connectionism and Cognitive Science*. Cambridge, MA: MIT-Press, 2001.

Marcus, G.F., Vijayan, S., Rao, S.B. and Vishton, P.M. Rule learning by seven-month-old infants. *Science,* 283: 77-80, 1999.

Marshall, J.B. Metacat: A Self-Watching Cognitive Architecture for Analogy-Making and High-Level Perception. PhD Thesis. Department of Computer Science and Cognitive Science Program. Indiana University, 1999.

Maskara, A. and Noetzel, W. Sequence recognition with recurrent neural networks. *Connection Science,* 5: 139-152, 1993.

Mitchell, M. *Analogy-Making as Perception: A computer model.* Cambridge, MA: MIT-Press, 1993.

Nelson, W.W. and Loftus, G.R. The Functional Visual Field During Picture Viewing. *Journal of Experimental Psychology: Human Learning and Memory,* 6: 391-399, 1980.

Nissen, M.J. and Bullemer, P. Attentional requirements of learning: evidence from performance measures. *Cognitive Psychology,* 19: 1-32, 1987.

Pinker, S. Enhanced: Out of the minds of babes. *Science,* 283: 40-41, 1999.

Rumelhart, D.E., Hinton, G.E. and Williams, R.J. Learning representations by backpropagating errors. *Nature,* 323: 533-536, 1986.

Russell, J. *What is Language Development? Rationalist, empiricist, and pragmatist approaches to the acquisition of syntax*. Oxford: Oxford University Press, 2004.

Slusarz, P. and Sun, R. The Interaction of Explicit and Implicit Learning: An Integrated Model. In J.D. Moore & K. Stenning (Eds.): *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science* Society, 980-985, 2001.

Spiegel, R. Human and Machine Learning of Spatio-Temporal Sequences: An Experimental and Computational Investigation. PhD-Thesis, University of Cambridge, UK, 2002.

Spiegel, R. and McLaren, I.P.L. Abstract and associatively-based representations in human sequence learning. *Phil. Trans. R. Soc. Lond. B,* 358: 1277-1283, 2003.

Spiegel, R. and McLaren, I.P.L. Computational Modeling of Human Performance in a Sequence Learning Experiment. *International Joint INNS/IEEE Conference on Neural Networks (IJCNN)*, 1: 212-217, 2003.

Spiegel, R. and McLaren, I.P.L. Associative Sequence Learning in Humans. To appear in the Journal of Experimental Psychology: Animal Behavior Processes, in press.

Spiegel, R., Suret, M., Le Pelley, M.E. and McLaren, I.P.L. Analyzing State Dynamics in a Recurrent Neural Network. *IEEE Proceedings of the World Congress on Computational Intelligence* (WCCI-IJCNN), 834-83, 2002.

Sun, R., Merrill, E. and Peterson, T. From implicit skills to explicit knowledge: a bottom-up model of skill learning. *Cognitive Science,* 25: 203-244, 2001.

Algorithms for cue competition in predictive learning: Suggestions from EEG and eye-tracking data

Andy J. Wills*, Aureliu Lavric*, Gareth S. Croft*, Tim L. Hodgson*
*School of Psychology
University of Exeter.
a.j.wills@ex.ac.uk

Abstract

Determining the extent to which a cue predicts an outcome is one of the most fundamental forms of learning. Some algorithms (e.g. Hebbian learning) assume that the rate of learning is solely determined by cue-outcome contiguity, but empirical work in humans and other animals indicates that *cue competition* can also affect learning rate (e.g. the phenomenon of blocking, Kamin, 1969, where presentation of a novel cue in compound with a cue that is already known to predict the outcome retards the formation of an association between the novel cue and the outcome). Aspects of cue competition are captured by a number of associative algorithms, including the Widrow-Hoff (1960) algorithm (a.k.a. LMS rule, delta rule, Rescorla-Wagner (1972) model), and by some reasoning-based (propositional) accounts (e.g. De Houwer et al., 2005). Given the wide range of models that predict cue competition, it seemed productive to ask what processes are employed by one highly successful predictive learning system – the human brain. In the current work, we investigated a hypothesis, suggested by the Mackintosh (1975) associative algorithm, that cue competition is a result of attention being selectively directed to cues that are known good predictors of the outcome. Specifically, we employed measurements of the EEG and of eye movements as our indices of attention in a cue competition experiment with adult humans. Support was found for our hypothesis, with features of the event-related EEG that have previously been implicated in selective attention being modulated in the expected manner. Looking time measures from eye-tracking, generally believed to be an index of overt attention, were also modulated in the expected direction. The EEG data implicates a relatively fast process (~120ms), possibly located in areas of the brain that deal with vision and object recognition (inferior occipital and temporal regions). The timing and location of these effects makes it unlikely they are the direct product of a high-level reasoning-based process, although the possibility remains that they are the top-down product of previous high-level reasoning.

References

De Houwer, J., Vandorpe, S., & Beckers, T. (2005). On the role of controlled cognitive processes in human associative learning. In A. J. Wills (Ed.), *New directions in human associative learning* (pp. 41-64). London: LEA.

Kamin, L. J. (1969). 'Attention-like' processes in classical conditioning. *Miami symposium on the prediction of behavior: aversive stimulation*. M. R. Jones, University of Miami Press.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, 82, 276-298.

Rescorla, R. A. and A. R. Wagner (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning II: Current research*. A. H. Black and W. F. Prokasy. New York, Appleton-Century-Crofts: 64 - 99.

Widrow, B. and M. E. Hoff (1960). Adaptive switching circuits. *IRE WESCON Convention.*

A Reinforcement Learning Agent with Associative Perception

Zhanna V. Zatuchna*
*School of Computing Sciences,
University of East Anglia,
Norwich, NR4 7TJ, England
zhanna.zatuchna@gmail.com

Anthony J. Bagnall†
†School of Computing Sciences,
University of East Anglia,
Norwich, NR4 7TJ, England
ajb@cmp.uea.ac.uk

Abstract

One of the most perspective ideas of further development of Reinforcement Learning (RL) research involves using associative learning models to improve performance of reinforcement learning agents. Learning Classifier Systems (LCS) have proved to be one of the most successful classes of RL methods that have been applied to maze environments. However, so far LCS have shown their effectiveness for small sized and simple maze environment tasks only. We try to overcome the limits by tying up the connection between LCS performance and principles of established psychological phenomena, those of associative learning in particular. We bring together the ideas of imprinting, laws of organization and stimulus generalization to create a basis for introducing an associative perception and recognition to the LCS framework. As a result, we develop the Associative Perception Learning Model, a new concept for modelling the learning process in autonomous learning agents. The model has been implemented as AgentP, a new LCS with Associative Perception and its performance has been evaluated on existing and new maze problems.

1 Introduction

The ability to adjust behaviour through learning process in complex environments is inherent in all living creatures (Dayan and Balleine, 2002). Using sequential interactions, an organism acquires knowledge about its environment and about the effects of its behaviour on it. All research in the Reinforcement Learning (Sutton and Barto, 1998; Dayan, 2001) field are based on this principle. However, how exactly the learning is achieved, depends on a particular learning model. The importance of designing of a thought-through learning model based on an appropriate research methodology has been discussed by many authors (Sutton, 1991; Bryson, 2004; Nehmzow, 2006).

Learning Classifier Systems (Holland and Reitman, 1978) is a reinforcement learning class of machine learning techniques created on a mixture of psychological and biological ideas. LCS are a promising direction of machine learning research that have been experiencing great increase of interest in the recent years (Bull and Kovacs, 2005). However, as LCS research has advanced, the connection between the algorithm used and the biological inspiration has weakened.

In recent years LCS proved to be one of the most promising classes of RL methods that have been applied to maze environments, an essential kind of the reinforcement learning problem. However, so far LCS have shown their effectiveness for small sized and simple maze environment tasks only, and there is a need for improving of their performance.

One of the most perspective ideas for further development of RL research is incorporating associative learning models (Hall, 1991; Dickinson and Balleine, 1993; Alonso and Mondragon, 2004) into the Reinforcement Learning framework. As an approach to the problem we present the Associative Perception Learning (APL) Model, a new concept for modelling the learning process in autonomous learning agents. The APL model has been implemented as a Learning Classifier System with Associative Perception and evaluated on a set on maze environments with promising results.

The rest of the paper is structured as fol-

lows. Section 2 defines mazes as a reinforcement learning problem. Section 3 introduces Learning Classifier Systems and performs a short review of their performance in maze environments. Section 4 gives a brief background into the principles of imprinting, laws of organization and stimulus generalization and explains how these principles are incorporated in the Associative Perception Learning Model. In Section 5 we analyze how the model, implemented in AgentP, a new LCS with Associative Perception, performs in action. Finally, conclusions are provided.

2 Maze Environments as a Reinforcement Learning Task

Learning from interaction is a fundamental idea underlying nearly all theories of learning and intelligence (Sutton and Barto, 1998; Dayan, 2001) and is essential in designing and building autonomous software systems for real-life applications (Alonso and Mondragon, 2004). Reinforcement Learning attempts to formalize the problem of learning from interaction. The learner and decision-maker is called the *agent* (Sutton and Barto, 1998). The agent interacts with the environment and the latter provides it with a signal which comprises information about the agent's surrounding. The agent and the environment interact continually; the agent selects actions, the environment responds to those actions by providing an occasional *reward* and presents new situations to the agent. In other words, the agent try to learn how to solve a problem or perform a certain task through *trial and error* interactions. It knows nothing about the task it has to learn, and it is only interested in maximizing the reward it receives. There are many different research approaches to emulating the reinforcement learning process in artificial cognitive systems, from neural nets (Kamo et al., 2002; Niv et al., 2002) to Learning Classifier Systems (Holland and Reitman, 1978; Wilson, 1995; Stolzmann, 2000; Bull and Hurst, 2001).

The phenomenon of spatial learning has been extensively studied in psychology (Prados and Redhead, 2002; Pearce et al., 2004). The maze problem serves as a formalized spatial learning model on the one side, and as a reinforcement learning task on the other. They are usually represented as grid-like two-dimensional areas that may contain different objects of any quantity and various quality. Figure 1 presents an example of a maze environment, a virtual equivalent of the classic T-maze used in psychological experiments to assess a rat's ability to remember spatial information.

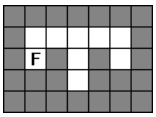


Figure 1: T-maze.

Maze environments serve as a simplified virtual model of the real world, and a learning agent in a maze environments is an example of an intelligent system modelling an animal trying to reach its objectives. The relative simplicity of mazes allows us to control the process of learning and trace the behaviour of the learning agent at every stage. At the same time the idea of maze environments includes a virtually unlimited number of graduated complexity levels, enabling researchers to use as simple or as complex environments as they need. These two factors make maze environments a good research paradigm for many navigation-based problems (Nehmzow, 1995) of Artificial Intelligence, from domestic appliance robots and autopilots for the automotive industry to network routing agents and autonomous walking robots for space research.

The process of learning begins when the agent is randomly placed in the maze on an empty cell. At each step it performs an action by attempting to move to an adjacent cell. The agent is allowed to move in all directions, but only through empty cells. The task is to learn how to reach food as fast as possible from any square. Once an external reward is received (usually by reaching a food cell), the agent's position is randomly reset and the task repeated. The agent uses a learning algorithm to form a policy to minimize the steps taken to food based on its ability to perceive the environment and the rewards received.

Usually a learning agent in maze environment is able to perceive not the whole picture of the environment, but the surrounding cells only. It may make mazes harder to solve because of the *aliasing problem*: some cells look exactly the same, but are in different places and demand different actions. Presence of aliasing cells may lead to a non-optimal behaviour in the maze

and decrease the agent's performance. Aliasing mazes represent a task of increased difficulty for learning agents, and so far the problem of learning in aliasing environments has not been resolved.

More detailed description of the maze problem and characteristics of maze complexity can be found in (Bagnall and Zatuchna, 2005).

3 Learning Classifier Systems

The maze problem has been widely used in machine learning research (Cassandra et al., 1994) to assess the performance and learning abilities of adaptive agents. The majority of RL techniques applied to mazes, as well as those with the most promising performance, belong to the class of Learning Classifier Systems (Holland and Reitman, 1978). Learning Classifier Systems are rule-based reinforcement learning systems, where an agent learns to perform a certain task by modifying its rule-based knowledge about the world through interacting with an unknown environment. LCS have proved their ability to solve optimally simpler mazes (Wilson, 1995; Stolzmann, 2000; Bull and Hurst, 2001) and some more intricate environments (Lanzi and Wilson, 1999; Métivier and Lattaud, 2002). However, so far the problem of learning in complex aliasing maze environments has not been resolved.

Another problem of the LCS performance in maze environments concerns computational resources. Traditionally Learning Classifier Systems (Holland and Reitman, 1978) include a *generalization* mechanism that randomly omits some part of the environment information before memorizing. The main purpose of the process is to find the main underlying regulations of the maze, evolve generalized knowledge of the environment and make the algorithm more scalable. However, LCS that have been used on mazes have tended to have the number of rules two orders of magnitude larger than the number of cells available for the learning agent in the maze. For example, LCS needed 6000 rules to solve maze Woods102 (Lanzi and Wilson, 1999) (Fig. 2, left) and 2800 rules to solve maze E1 (Métivier and Lattaud, 2002) (Fig. 2, right).

The larger sets of rules demand not only large amount of memory, but also significantly larger amount of learning time. For example,

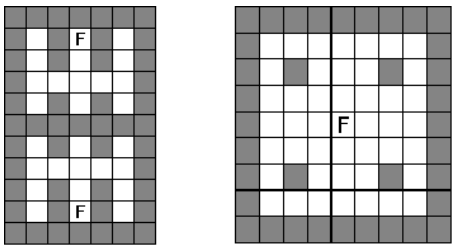


Figure 2: Woods102 (left); E1 (right).

Bull reported his system needed 10,000 trials for Woods1 maze environment (Bull and Hurst, 2002) (Fig. 3, left), and 25,000 trials for small Woods101 environment (Bull and Hurst, 2003) (Fig. 3, right). Thus, the present approach to generalization in LCS, when a learning agent tries to evolve the optimally generalized rules from the beginning, before it has learnt the environment, seems to be damaging for the learning process and may actually make it more difficult.

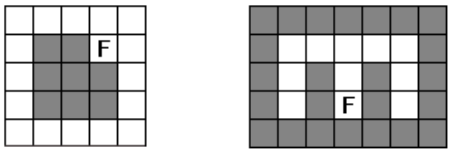


Figure 3: Woods1 (left); Woods101 (right).

4 The Associative Perception Learning Model

As we discussed in previous sections, there are two major problems associated with the performance of LCS learning agents in maze environments. First, to our best knowledge so far none of them has proved its ability to solve complex aliasing mazes. Second, the published results of the experiments in maze environments suggest that the architecture of most LCS agents may require prohibitively large rule sets to solve larger mazes.

Trying to overcome the problems and improve the performance of Learning Classifier Systems in maze environments we reviewed the major psychological approaches to explaining learned behaviour in humans and animals and developed the Associative Perception Learning Model, a new concept for modelling the learning process in autonomous learning agents that ap-

proaches the problem of learning through mechanism of associative perception and recognition in a complex environment. The APL model is based on the following psychological principles:

Imprinting. Imprinting is an especially rapid and relatively irreversible learning process first observed and described by Konrad Lorenz (1935). In the process of imprinting, distinctive attributes of external objects are memorized by an individual and become connected with his behavioural reactions. The imprinting phenomenon has been extensively studied in psychology and biology (Honey and Bolhuis, 1997; Enquist et al., 2002). According to the principle, the learning agent in the Associative Perception Learning model absorbs the environment signals as they are perceived, without any changes or generalization.

Laws of Organization. Gestalt theory emphasizing higher-order cognitive processes was created early in the XX century (Wertheimer, 1938). The focus of the theory was the idea of grouping, which occurs when characteristics of stimuli cause an individual to structure or interpret a visual field or problem as a global construct. The rules of interpretation may take several forms, such as grouping by proximity, similarity, closure, etc. These factors were called the laws of organization and explained in the context of perception and problem-solving. In the APL model environment signals, received sequentially, are grouped according to the rules of interpretation and perceived as a single indecomposable image, employing the associative way of learning.

Stimulus Generalization. The ability to generalize plays an important role in both natural and artificial cognitive systems (Ghirlanda and Enquist, 2003) and has been extensively studied in the reinforcement learning research (Sutton, 1996; Balkenius and Winberg, 2004). The behavioural phenomenon termed stimulus generalization was first described and interpreted by Pavlov (1927) and later extensively studied by Skinner (1953). According to their research, an individual that has learnt a certain behaviour, responds in a similar manner to stimuli that are similar to the one on which he was trained. In terms of maze learning, stimulus generalization would mean creating a post-learning generalization mechanism that allows to transfer the experience obtained in a certain maze section to another maze area with similar attributes. Experiments with rats demonstrating the effect in action were recently performed by Pearce et al. (2004).

4.1 Image Creation

The Associative Perception Learning Model operates perceptive images where the sensory input perceived at the initial position of the agent is connected with both the sensory input perceived at the result position and the agent's action. I.e., the model links the input at time t, S^t, the action taken a and the next input S^{t+1} together, creating a single image (Figure 4).

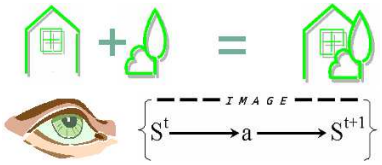


Figure 4: Image creation in the Associative Perception Learning Model.

The image is perceived non only as a time vector reflecting the cause-and-effect relations but also a united information structure describing the attributes of its components through their interconnections.

4.2 Image Connection

The associative perception mechanism employs a *sliding image formation* principle (Fig. 5). Each representation of an environment state S^t is associated with two others, the previous S^{t-1} and the following one S^{t+1}. Thus, the formation of images occurs under sliding state-to-state attention. As a result, the system is always able to keep track of the connections between images and place a perceived state in the context of the surroundings. In the case of reinforcement learning, it allows the system to share any received reward among the associated images.

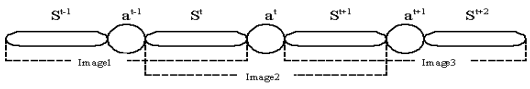


Figure 5: Sliding image formation.

4.3 Differentiation

In the real life, we can distinguish two similar objects, such as houses, based on their surround-

ings. Other buildings or trees growing beside become associated with the original object and give us a clue to differentiate it correctly. Influence of associative processes on differentiation has been extensively studied by Murphy et al. (2004). To disambiguate aliasing cells, the Associative Perception Learning Model uses the same principle. To decide whether two images are the same, it matches them against each other element by element. If at least one of the elements of the first image does not match the corresponding element of the second image, the rest of the elements are considered to be non-matching also, even if they replicate each other perfectly. To draw the conclusion about the image match, all elements of both images have to be consistent.

Figure 6 illustrates the process: in the course of differentiation, memorized images are in turn compared to the current perceptive image. For the performed action a the current initial state S^t_{cur} is compared with the memorized initial state S^t_{imp}, and the current result state S^{t+1}_{cur} is compared with the memorized result state S^{t+1}_{imp}. If one state matches and the other does not, the matching state is considered to be aliasing and marked with a distinguishing sign to allow it to be correctly disambiguated in the future.

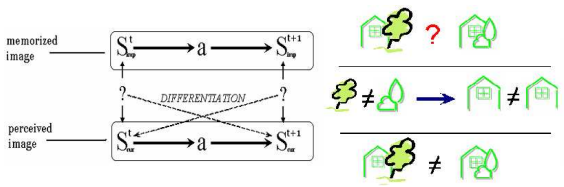


Figure 6: Differentiation based on association.

4.4 Generalization

A generalization mechanism would be required for extensive tasks that include dozens of thousands states. Generalization in the APL model is based on the post-learning principle and is applied when the agent has been showing a stable performance in the maze for a certain period of time. The mechanism attempts to find general patterns in the images that performed best in the past. The found patterns are then memorized and used as a reference set of the first choice. If there is no suitable pattern found, the system operates in the regular learning mode. Thus, being placed in a vast maze environment, the system would start from exploring the immediate surrounding, extract valuable information from its present images into a more compact representation and then use the extraction at the next stages of learning. This mechanism allows the agent to extrapolate the knowledge obtained in one area of the environment to the others areas of the same environment, or, alternatively, transfer the knowledge obtained in a certain environment to other similar environments.

5 Experiments and Results

The Associative Perception Learning Model described in Section 4 has been implemented as a Learning Classifier System with Associative Perception, called AgentP. Perceptive images of the APL model have been supplemented with a reward prediction coefficient to form the set of behaviour rules in AgentP.

AgentP has been tested on the maze environments discussed in Section 3 and shown the optimal performance results with significantly smaller amount of rules in comparison to the other LCS. For example, AgentP needs only 82 rules to solve Woods102 (compare to 6000 rules reported by Lanzi and Wilson (1999)) and only 240 rules to solve E1 (2800 rules reported by Métivier and Lattaud (2002)). Overall, the algorithm performs on minimal memory requirements: the number of rules required for solving a maze is usually a value of the same order as the size of this maze.

AgentP also outperforms the other LCS in terms of learning time. It needs around 25 trials to solve aliasing Woods101 environment (compare to 25,000 trials reported by Bull and Hurst (2003)) and 18 trials on average to solve Woods1 (10,000 trials reported by (Bull and Hurst, 2002)).

We also tested AgentP on several other mazes used in LCS/RL research before (31 mazes in total). The results of these tests indicate that in the majority of cases (94%) AgentP has been able able to solve them optimally. The results were compared to other agents where appropriate and the comparison conformed (Zatuchna, 2004; Zatuchna and Bagnall, 2005b) that AgentP solves the mazes in less time and with less memory than similar agents.

To perform more thorough tests of the agent architecture we have have evaluated AgentP on a large set of new mazes. Figure 7 (left) shows AliasIIMaze17, one of the 80 new mazes used

in the experiment. The maze is toroidal and contains one hundred cells in total, including 15 aliasing squares. Aliasing cells which produce the same sensory input are marked with the same numbers. AgentP has reached the optimal performance of 2.72 steps to food after 723 trials in average (all experiments consisted of 50 runs in a row). AliasIIMaze20 (Figure 7, right) includes 19 aliasing squares that produce 9 different sensory inputs. The maze has a longer average distance to food, and has been successfully solved by AgentP (7.31 steps to food) after 484 trials in average.

On the whole, AgentP has solved 72 of 80 available mazes (optimal performance on 90% of environments). The full collection of mazes, specification of their parameters as well as a complete set of correctness, convergence and memory statistics for AgentP is available at (Zatuchna and Bagnall, 2005a).

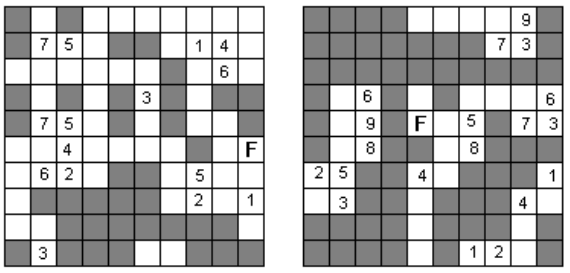


Figure 7: AliasIIMaze17 (left); AliasIIMaze20 (right).

The architecture of AgentP discussed in this paper represents a simulation of the learning process in its early stages only and does not involve the post-generalization mechanism described in the original APL model. The idea of this stage of the research is to rectify whether the model of Associative Perception Learning incorporated into the LCS framework is able to sharpen the learning ability of Learning Classifier Systems in maze environments and create a system capable of differentiating of complex aliasing patterns. To verify the viability of the idea of post-generalization for maze learning and test the learning advantages of the APL model in full, more advanced experiments are required.

6 Conclusions

In this paper we presented the Associative Perception Learning Model, a new concept for modelling the learning process in autonomous learning agents. The model approaches the problem of learning through mechanism of associative perception and recognition in a complex environment. The model operates perceptive images where information about the environment state at the initial position of the agent is connected with both information about the environment state at the result position and the agent's action. The system employs a refined differentiation mechanism, that allows to provide more precise and accurate recognition of the environment information. The Associative Perception Learning Model includes the sliding image formation principle, that allows the system keep track of the connections between images.

The model has been implemented as AgentP, a new LCS with Associative Perception. Its performance has been evaluated on existing and new maze problems. AgentP has been able to show the optimal performance on 90% of the mazes The results of the experiments show that AgentP is capable to solve the majority of the mazes optimally, and for existing mazes it does it in less time and with less memory than other LCS-based agents. It allows us to suggest that the presented learning model is a promising design in the area of learning agents for maze environments and supports the idea of benefits that can bring psychologically justified algorithms for autonomous learning agents in general.

References

Eduardo Alonso and Esther Mondragon. Agency, learning and animal-based reinforcement learning. In M. Rovatsos M. Nickles and G. Weiss, editors, *Agents and Computational Autonomy: Potentil, Risks and Solutions*, pages 1–6. Springer-Verlag, 2004.

Anthony J. Bagnall and Zhanna V. Zatuchna. On the classification of maze problems. In Larry Bull and Tim Kovacs, editors, *Foundations of Learning Classifier Systems*, pages 307–316. Springer, 2005.

Christian Balkenius and Stefan Winberg. Cognitive modeling with context sensitive reinforcement learning. In *Proceedings of AILS '04*, 2004.

Joanna J. Bryson. Modularity and specialized learning: Reexamining behavior-based artificial intelligence. In Jochen Triesch and Tony Jebara, editors, *The Proceedings of The Third International Conference on Development and Learning (ICDL'04): Developing Social Brains*, pages 309–316, 2004.

Larry Bull and Jacob Hurst. ZCS: Theory and practice. Technical Report 01-001, UWE Learning Classifier Systems Group, 2001.

Larry Bull and Jacob Hurst. ZCS redux. *Evol. Comput.*, 10(2):185–205, 2002.

Larry Bull and Jacob Hurst. A neural Learning Classifier System with self-adaptive constructivism. Technical report, University of the West of England, 2003.

Larry Bull and Tim Kovacs, editors. *Foundations of Learning Classifier Systems*. Springer, 2005.

Anthony R. Cassandra, Leslie Pack Kaelbling, and Michael L. Littman. Acting optimally in partially observable stochastic domains. In *Proceedings of the Twelfth National Conference on Artificial Intelligence (AAAI-94)*, volume 2, pages 1023–1028. MIT Press, 1994.

Peter Dayan. Reinforcement learning. In C. Gallistel, editor, *Steven's Handbook of Experimental Psychology*. Wiley, 2001.

Peter Dayan and Bernard Balleine. Reward, motivation and reinforcement learning. *Neuron*, (36):285–298, 2002.

Anthony Dickinson and Bernard Balleine. Actions and responses: The dual psychology of behaviour. In M.W. Brewer N. Eilan, R.McCarthy, editor, *Spatial Representation*, pages 277–293. Oxford: Basil Blackwell, 1993.

Magnus Enquist, Stefano Ghirlanda, Daniel Lundqvist, and Carl-Adam Wachtmeister. An ethological theory of attractiveness. In G. Rhodes and L. A. Zebrowitz, editors, *Facial Attractiveness: Evolutionary, Cognitive, and Social Perspectives*. London: Ablex, 2002.

Stefano Ghirlanda and Magnus Enquist. A century of generalization. *Animal Behaviour*, (66): 15–36, 2003.

Geoffrey Hall. *Perceptual and Associative Learning*. Oxford University Press, 1991.

John H. Holland and Judith S. Reitman. Cognitive systems based on adaptive algorithms. In D. A. Waterman and F. Hayes-Roth, editors, *Pattern-directed Inference Systems*. New York: Academic Press, 1978.

Rob C. Honey and Johan J. Bolhuis. Imprinting, conditioning, and within-event learning. *Quarterly Journal of Experimental Psychology*, (50B):97–110., 1997.

Masashi Kamo, Stefano Ghirlanda, and Magnus Enquist. The evolution of signal form: Effects of learned vs. inherited recognition. In *Proceedings of the Royal Society of London*, volume B269, pages 1765–1771, 2002.

Pier Luca Lanzi and Stewart W. Wilson. Optimal Classifier System performance in non-Markov environments. Technical Report 99.36, Dipartimento di Elettronica e Informazione - Politecnico di Milano, 1999.

Konrad Lorenz. Der kumpan in der umwelt des vogels. *Journal of Ornithology*, pages 137–215, 1935.

Marc Métivier and Claude Lattaud. Anticipatory Classifier System using behavioral sequences in non-Markov environments. In *IWLCS*, pages 143–162, 2002.

Robin A. Murphy, Esther Mondragona, Victoria A. Murphy, and Nathalie Fouquet. Serial order of conditional stimuli as a discriminative cue for pavlovian conditioning. *Behavioural Processes*, 2004.

Ulrich Nehmzow. *Scientific Methods in Mobile Robotics - Quantitative Analysis of Agent Behaviour*. Springer, 2006.

Ulrich Nehmzow. Animal and robot navigation. *Robotics and Autonomous Systems*, 7(1-2):71–81, 1995.

Yael Niv, Daphna Joel, Isaac Meilijson, and Eytan Ruppin. Evolution of reinforcement learning in uncertain environments: A simple explanation for complex foraging behaviors. *Adaptive Behavior*, 10(1):5–24, 2002.

Ivan P. Pavlov. *Conditioned Reflexes.* London: Oxford University Press., 1927.

John M. Pearce, Mark A. Good, Peter M. Jones, and Anthony McGregor. Transfer of spatial

behaviour between different environments: Implications for theories of spatial learning and for the role of the hippocampus in spatial learning. *Journal of Experimental Psychology: Animal Behavior Processes*, (30):135–147, 2004.

Jose Prados and Ed Redhead. Preexposure effects in spatial learning: From gestaltic to associative and attentional cognitive maps. *Psicologica*, 23(Special Issue on Spatial Learning): 59–78, 2002.

Burrhus F. Skinner. *Science and Human Behavior.* New York: Macmillan., 1953.

Wolfgang Stolzmann. An introduction to Anticipatory Classifier Systems. In Wolfgang Stolzmann Pier Luca Lanzi and Stewart W. Wilson, editors, *Learning Classifier Systems. From Foundations to Applications*, pages 175–194. Springer-Verlag, 2000.

Richand S. Sutton. Generalization in reinforcement learning: Successful examples using sparse coding. In *Advances in Neural Infromation Processing Systems: Proceedings of the 1995 Conference*, page 10381044. Cambridge, MA: MIT Press., 1996.

Richard S. Sutton. Reinforcement learning architectures for animats. In J.-A. Mayer and S. W. Wilson, editors, *From Animals to Animats: Proceedings of the First International Conference on Simulation of Adaptive Behaviour.*, pages 288–296. Cambridge, MA: MIT Press, 1991.

Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 1998.

Max Wertheimer. Laws of organization in perceptual forms. In *A Source Book of Gestalt Psychology*, pages 71–88. London, Routledge and Kegan Paul, 1938.

Stewart W. Wilson. Classifier fitness based on accuracy. *Evolutionary Computation*, 3(2):149–175, 1995.

Zhanna V. Zatuchna. AgentP model: Learning Classifier System with associative perception. In Xin Yao et al., editor, *Proceedings of the Parallel Problem Solving from Nature Conference (PPSN)*, volume 3242 of *Lecture Notes in Computer Science*, pages 1172–1182. Springer, 2004.

Zhanna V. Zatuchna and Anthony J. Bagnall. Maze material for AgentP, 2005a. URL `http://www.cmp.uea.ac.uk/Research/kdd/projects.php?project=17`.

Zhanna V. Zatuchna and Anthony J. Bagnall. AgentP classifier system: Self-adjusting vs. gradual approach. In *Proceedings of the 2005 Congress on Evolutionary Computation*, pages 1196–1203, 2005b.

Modelling of Temperament in an Associative Reinforcement Learning Agent

Zhanna V. Zatuchna*
*School of Computing Sciences,
University of East Anglia,
Norwich, NR4 7TJ, England
zhanna.zatuchna@gmail.com

Anthony J. Bagnall†
†School of Computing Sciences,
University of East Anglia,
Norwich, NR4 7TJ, England
ajb@cmp.uea.ac.uk

Abstract

The idea of temperament refers to the essential properties of the central nervous system that can produce variations in behaviour and influence the ability of an individual to learn and adapt itself to a complex environment. The research represents an attempt to model certain biological aspects of temperament as alternative learning mechanisms. We investigate the influence of the 'virtual temperament' on the effectiveness of the learning in maze environments and evaluate the performance of the learning algorithms on two extensive sets of maze problems.

1 Introduction

Temperament refers (Pavlov, 1927, 1957) to basic dimensions of personality that are grounded in biology and explains individual differences in the developmental process. In the research we use the idea of temperament to improve the learning results of autonomous adaptive agents. More specifically, we model certain biological aspects of temperament as alternative approaches to extracting knowledge from interactions with an environment and investigate how it influences the performance results of a learning agent in maze environments.

Artificial cognitive systems have been extensively studied in the reinforcement learning research (Sutton and Barto, 1998; Sutton, 1991; Dayan, 2001; Balkenius and Winberg, 2004) and include many different approaches, from neural nets (Kamo et al., 2002; Niv et al., 2002) to Learning Classifier Systems (Holland and Reitman, 1978; Wilson, 1995; Stolzmann, 2000; Bull and Hurst, 2001). Our interest lies in improving of learning abilities of Learning Classifier Systems (LCS), a group of rule-based machine learning algorithms that produce adaptive systems for different kinds of learning problem. For our experiments we use AgentP, a reinforcement learning agent with associative perception (Zatuchna, 2004, 2005). AgentP is a recently introduced variation of LCS which has shown promising results in the area of maze learning (Zatuchna, 2004).

Mazes were originally used in psychological experiments involving primarily small laboratory animals, such as rats (Tolman, 1932; Prados and Redhead, 2002; Pearce et al., 2004), to study characteristics of the learning process and the role of reward in learned behaviour (Dayan and Balleine, 2002; Bryson, 2004). Later mazes were adopted in machine learning research and now serve a reinforcement learning task (Sutton and Barto, 1998; Dayan, 2001) that involves learning actions to optimize some objective in an environment.

Virtual mazes consists of cells, each of which can be either empty and available for the agent or occupied by a barrier. The learning agent is usually able to see only the nearest cells around itself, perceived as a sensory input, and in certain mazes it results in the problem of *aliasing*. Aliasing occurs when some maze squares look exactly the same for the agent, despite the fact that they are in different locations. In other words, the sensory inputs the agent receives in these cells are identical, but the actions that need to be taken in them may be different. Aliasing may significantly disrupt learning and result in the agent's inability to solve the maze.

Learning Classifier Systems have been successfully applied to maze problems (Wilson, 1995; Stolzmann, 2000; Bull and Hurst, 2001;

Lanzi and Wilson, 1999; Métivier and Lattaud, 2002) and are a promising direction of machine learning research that has been receiving an increase in interest in the recent years (Bull and Kovacs, 2005). However, the problem of learning in aliasing environments has not been resolved yet and we hope our temperament-based approach will advance the LCS performance toward the goal.

The rest of the paper is structured as follows. Section 2 gives a brief introduction to Learning Classifier Systems and AgentP in particular. Section 3 offers background information on the idea of temperament and introduces two different learning modes for AgentP, Self-Adjusting and Gradual, which represent variations in mobility of the nervous system. In Section 4 we analyze how the two different models perform on an extensive range of maze environments. Finally, conclusions are provided.

2 AgentP: a Learning Classifier System with Associative Perception

Learning Classifier Systems (LCS) are rule-based systems, where an agent learns to perform a certain task by interacting with an unknown environment. Occasionally it receives some feedback from the environment, usually in the form of a reward and uses this reward to guide an internal learning process and modify its rule-based knowledge about the world. Fig. 1 represents an example of a typical rule structure used by LCS. It includes a coded sensory input, received from the environment, action directions and a reward expectation.

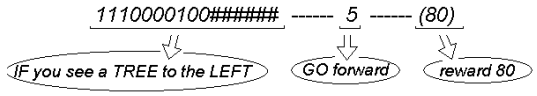


Figure 1: Example of a rule structure used by LCS.

The great importance of model design in development of artificial cognitive systems has been discussed by many authors (Sutton, 1991; Bryson, 2004; Nehmzow, 2006). One of the most promising ideas for further Reinforcement Learning research is integrating it with the principles of associative learning (Hall, 1991; Alonso and Mondragon, 2004). AgentP is a Learning Classifier System with Associative Perception (Zatuchna, 2005), that retains and extends the tradition of biologically inspired designs for learning agents. It employs the Associative Perception Learning Model (Zatuchna, 2004) and incorporates psychological principles of imprinting (Lorenz, 1935; Honey and Bolhuis, 1997; Enquist et al., 2002) and the laws of organization (Wertheimer, 1938), that have not been previously used with LCS.

The rules in AgentP are extended with a prediction part (Stolzmann, 2000), thus, each rule contains not only the initial sensory input and the performed action (as shown in Fig 1), but also the result sensory input. In other words, AgentP tries to predict what consequences will have a certain action in a particular situation and what picture of the environment it will see if it performs the action.

The rule structure in AgentP also includes a new ID system. Each rule has reserved space for an additional ID of each sensory input (see Fig. 2; parts of the rules representing IDs are marked with X). The ID system allows the solution of the problem of aliasing squares: each confusing input is processed by the differentiation mechanism (Zatuchna, 2004) and receives a unique number.

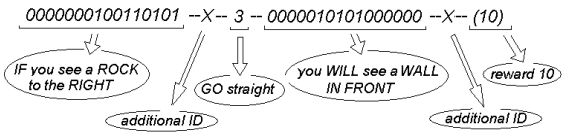


Figure 2: Example of a rule in AgentP.

An ID of a particular sensory input in a certain rule can then be transferred to another rule which contains the same sensory input through the mechanism of association. In such way the knowledge about the true identity of an aliasing sensory input can be spread in the rule population allowing it to unambiguously identify the aliasing cell from whatever side the agent approaches it. The transfer of the ID information takes place every time when the agent is 'sure' about its position in the environment. However, the regulations on when it is 'sure' and when it is 'not sure' may differ.

3 Modelling of Temperament

Pavlov (1927) explained differentiations in hu-

man states based on the assumption that essential properties of the central nervous system can produce variations in behavioural and psychological outputs. According to his conclusions, the brain activities on a microscopic level come down to the intensity, homeostasis and mobility of nerve cell stimulation and inhibition. He called the essential properties of the nervous system *strength*, *equilibrium* and *mobility*.

Strength refers to the capacity of the cerebral cells to endure intense stimulation and their resistance to powerful external disruptors and stress. Equilibrium refers to the ability to maintain a balance between excitation and inhibition. Mobility defines the ease with which brain processes could shift from one state to another to keep pace with changing environmental demands and determines the speed at which an individual can adopt specific appropriate responses to environmental stimuli. Thus, mobility is a characteristic of the nervous system that is directly connected with the quality of the learning process and reflects the adaptive capabilities of an individual.

We have introduced two alternative procedural techniques to the learning process that reflect the idea of mobility. It has resulted in creation of two variations of AgentP of 'different temper'. The first, *Self-Adjusting* AgentP, is flexible and adapts rapidly to changing information; the second, *Gradual* AgentP, is more conservative in drawing conclusions and rigid when it comes to revising strategy.

3.1 Self-Adjusting AgentP

Let us assume there are three consecutive aliasing cells, visited by Self-Adjusting AgentP in a learning run. The sensory inputs the agent receives in the squares are *0000000100110101*, *0000010101000000* and *1111111111000000* accordingly, rules *A* and *B* consecutively describe the movement of the agent in these cells. As soon as both rule *A* and rule *B* are the only rules that match the actions of the agent, the agent is considered to be 'sure' about its location in the maze and may transfer the ID information from one rule to another. Figure 3 illustrates the process: the ID of the intermediate state *0000010101000000* is transferred from rule *A* to rule *B*. As a result, rule B is changed as shown.

The only restriction on the learning process in the Self-Adjusting learning mode is the presence

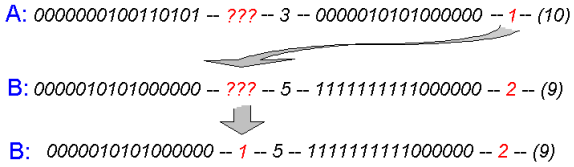


Figure 3: Self-Adjusting AgentP: transferring ID.

of multiple rules in an uncertain situation. Thus, if the agent has two or more consistent rules for each action, no ID transference takes place.

Under these conditions the agent performs as a rapidly adjusting system: IDs are immediately transmitted with no precautions and mistakes are adjusted for without checks. This means AgentP can explore all aliasing squares at the same time, but also means that incorrect information may be transmitted.

Figure 4 illustrates the exploration of aliasing cells in a maze by Self-Adjusting AgentP. It initiates spreading of IDs at many places at once (aliasing squares are marked with numbers, those in development are marked with a tint), rapidly covering the maze with its labile learning process.

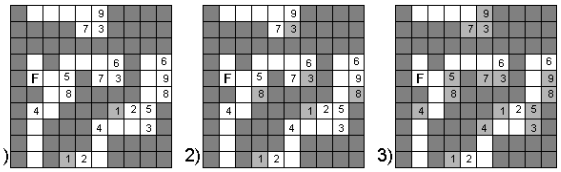


Figure 4: Exploration of aliasing cells by Self-Adjusting AgentP.

3.2 Gradual AgentP

Gradual AgentP is only 'sure' about its position in the environment when, firstly, it has been 'sure' on every step since the last non-aliasing square and, secondly, when at least one rule of the two last ones has no uncertainty about the identities of the aliasing states it includes. Referring to the example in Section 3.1, Gradual AgentP will not transmit the ID from rule *A* to rule *B* because rule *A* does not satisfy the condition (Fig. 5). However, if the cell, that produces the initial sensory state *0000000100110101* in rule *A* were non-aliasing, the ID would be transmitted.

The restriction of presence of multiple rules also applies to the Gradual learning mode. In

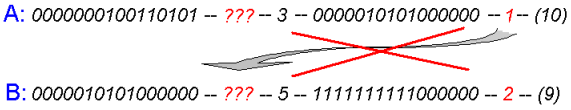


Figure 5: Gradual AgentP: no ID transfer because the initial state of rule A is uncertain.

addition, Gradual AgentP does not include any direct correction of mistakes; if a cell that was previously considered as a reliable non-aliasing square has been freshly discovered to be an alias, all rules including the unreliable information about the cell, are deleted, and its exploration begins from scratch.

The Gradual agent uses its rules as a thread to orient itself in aliasing surrounding. Under these settings AgentP is a cautious learning system that explores the aliasing environment gradually, building up a consecutive bridge from reliable non-aliasing squares through an aliasing conglomerate. Figure 6 illustrates the process of exploration of aliasing cells in a maze by Gradual AgentP. It moves from one cell to a neighbouring one slowly (marked with a tint), and does not draw any conclusion about the next piece of puzzle until it has finished with the previous one.

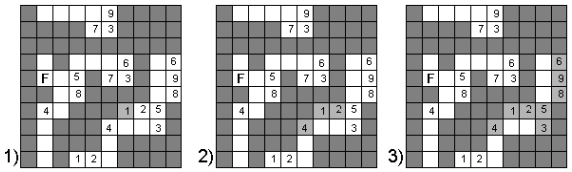


Figure 6: Exploration of aliasing cells by Gradual AgentP.

4 Experiments and Results

To test the learning abilities of the agent with different temperament settings, we repeatedly ran it on two sets of maze environments. Each stage of the experiments involved running the system in two learning modes sequentially, first Self-Adjusting and then Gradual.

The majority of LCS research has been evaluated on a small number of mazes (1-3 mazes only) (Bagnall and Zatuchna, 2005). To improve validity of the experiments and provide a firm basis for comparing our results to those of other learning algorithms we use two extensive sets of maze environments.

The first set consisted of 80 medium mazes up to 100 cells in total. Figure 7 presents examples of medium mazes (aliasing cells that produce the same sensory input are marked with the same numbers). AliasIIMaze20 (to the left) has the optimal performance of 7.31 steps to food and comprises of 19 aliasing cells reflected in 9 aliasing states (sensory inputs). The optimal performance on AliasIIIMaze20 (to the right) is 4.28 steps to food in average. The maze includes 31 aliasing squares represented as 11 different sensory inputs.

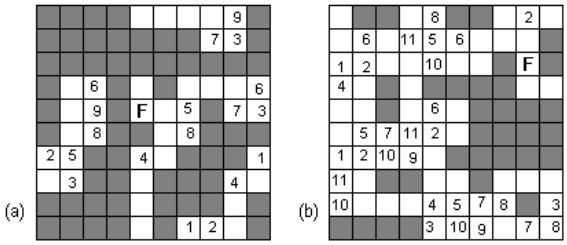


Figure 7: (a) AliasIIMaze20; (b) AliasIIIMaze20.

Both, Self-Adjusting and Gradual AgentP, were able to solve the majority of mazes from the medium maze set. The number of mazes for which AgentP was not able to find an optimal policy is virtually the same for the two learning modes. Self-Adjusting reached optimal performance on 89% of mazes, while Gradual showed 90% result.

The major difference in the performance of the two agents in medium environments was the learning time: on average Gradual AgentP needed more time to learn a maze. Figure 8 shows the scatter plot of average steps to food against trials before the learning was accomplished. This graph demonstrates the longer time required by Gradual AgentP. Thus, after the first stage of the experiments Self-Adjusting AgentP seemed to be a more effective problem solver than Gradual.

The second set contained 271 larger mazes that vary from 120 up to 1140 squares in total. Figure 9 presents examples of larger mazes. Environment LargeAliasIIIMaze15 has the optimal performance of 7.73 steps to food in average and includes 62 aliasing squares that produce 27 sensory inputs. LargeAliasIIMaze5 has the optimal performance of 11.5 steps to food and contains 60 aliasing cells (24 aliasing sensory inputs).

The experiments on larger mazes clearly

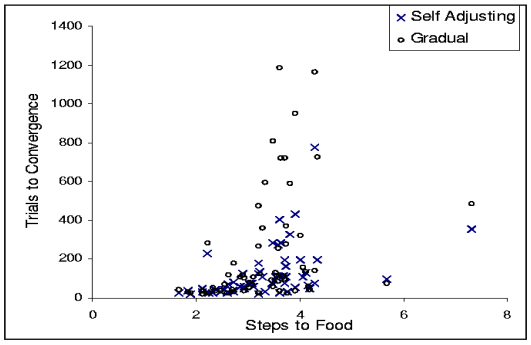


Figure 8: Steps to food plotted against average trials before learning for Self-Adjusting and Gradual AgentP on medium mazes.

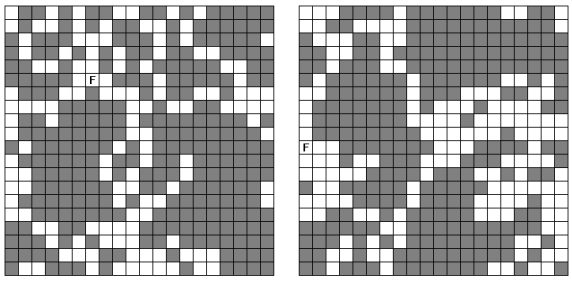


Figure 9: LargeAliasIIIMaze15 (left); LargeAliasIIMaze5 (right).

showed that there was a significant difference in the abilities of Self-Adjusting and Gradual AgentP to solve them. The performance of Self-Adjusting AgentP dropped significantly compared to the previous results (down to 23%). Gradual AgentP, on the contrary, still performed well and demonstrated three times better performance than Self-Adjusting (71%). Table 1 gives the number of mazes solved every time (for 50 runs) by both types of agent and its percentage equivalent. A complete set of correctness, convergence and memory statistics for both sets of mazes is given at (Zatuchna and Bagnall, 2005).

Table 1: Number of mazes solved by Self-Adjusting and Gradual AgentP

Maze Set	Self Adjusting	Gradual
Medium	71 / 89%	72 / 90%
Large	62 / 23%	193 / 71%

Analysis of the results suggests that enlarging of a maze results in the increasing of not only the average steps to food value, but also the number of aliasing squares. The distribution of the optimal runs against the number of aliasing squares (not shown) have demonstrated a clear dependance between the latter and ability of the agents to solve the maze. The more aliasing squares a maze contains, the better proportion of optimal runs is demonstrated by Gradual AgentP, compared to Self-Adjusting.

Another indicative point of the problems faced by the agents in large maze environments is the amount of knowledge extracted by them by the end on a learning run. The average number of rules created by Self-Adjusting AgentP is noticeably smaller compare to Gradual. More precisely, the size of the rule population created by Gradual agent on the larger maze environment was 255 on average. Meanwhile, the average rule set created by Self-Adjusting was only 252 rules. This suggests that the ability of Self-Adjusting agent to recognize all significant regulations in an environment drops down as the maze size enlarges.

5 Conclusions

In this research we incorporate the idea of temperament into a machine learning framework to improve learning abilities of autonomous adaptive agents. We base our model on the idea of mobility, a significant characteristic of the nervous system that reflects the adaptive capabilities of an individual. The experiments were performed using AgentP, a Learning Classifier System with Associative Perception. We introduce two alternative learning techniques to AgentP and create two systems of 'different temperament'. Self-Adjusting is flexible and adapts rapidly to changing information; Gradual is conservative and rigid. Then we test AgentP in the two different learning modes on a number of maze environments.

There were two stages of the experiments: first involved 80 medium mazes; second included 271 large-sized mazes, up to 1140 cells in total. So far these are the most extensive maze sets used in research.

The results of the experiments show that both versions of AgentP can solve the majority of the medium mazes quite easily and with virtually equal performance. Gradual AgentP, though, needed more time compared to Self-Adjusting to complete the learning, therefore, seems to be less effective. Larger mazes, however, provide

us with a different result. Gradual AgentP is still able to solve the majority of the environments, while the performance of Self-Adjusting drops significantly.

Analysis of the results suggest that the more aliasing cells in a maze, the larger is the probability that Self-Adjusting agent will become confused and not able to recognize the significant regulations in the environment because of its careless leaning style. At the same time the cautious and deliberate approach of Gradual AgentP proved to be more reliable in the situation of large and complex maze environments.

Overall, Gradual AgentP takes longer to converge than Self-Adjusting, but performs better on some of the medium mazes and on the vast majority of the larger mazes. However, on some of the medium mazes Self-Adjusting AgentP finds a better policy than Gradual. This indicates that Gradual AgentP is at times discounting useful information because it cannot determine its meaning with certainty. Thus, the first priority for future work would be investigating in more detail what makes different mazes that can be easily solved by Self-Adjusting AgentP hard for Gradual, and vice versa. As a next step in the research, hybridizing the Gradual agent with the Self-Adjusting may lead to improved performance.

The research has brought out three valuable outcomes. First, the model represents a successful simulation of the influence of temperament on the learning process and offers an approach to modelling of temperament in artificial leaning systems. Second, it has resulted into development of two versions of AgentP, one of which, Gradual, seems to be one of the most promising reinforcement learning design for maze environments for the present moment. Finally, the research has demonstrated the advantage that can be had bringing established psychological phenomena to the design of learning algorithms.

References

Eduardo Alonso and Esther Mondragon. Agency, learning and animal-based reinforcement learning. In M. Rovatsos M. Nickles and G. Weiss, editors, *Agents and Computational Autonomy: Potentil, Risks and Solutions*, pages 1–6. Springer-Verlag, 2004.

Anthony J. Bagnall and Zhanna V. Zatuchna. On the classification of maze problems. In Larry Bull and Tim Kovacs, editors, *Foundations of Learning Classifier Systems*, pages 307–316. Springer, 2005.

Christian Balkenius and Stefan Winberg. Cognitive modeling with context sensitive reinforcement learning. In *Proceedings of AILS '04*, 2004.

Joanna J. Bryson. Modularity and specialized learning: Reexamining behavior-based artificial intelligence. In Jochen Triesch and Tony Jebara, editors, *The Proceedings of The Third International Conference on Development and Learning (ICDL'04): Developing Social Brains*, pages 309–316, 2004.

Larry Bull and Jacob Hurst. ZCS: Theory and practice. Technical Report 01-001, UWE Learning Classifier Systems Group, 2001.

Larry Bull and Tim Kovacs, editors. *Foundations of Learning Classifier Systems*. Springer, 2005.

Peter Dayan. Reinforcement learning. In C. Gallistel, editor, *Steven's Handbook of Experimental Psychology*. Wiley, 2001.

Peter Dayan and Bernard Balleine. Reward, motivation and reinforcement learning. *Neuron*, (36):285–298, 2002.

Magnus Enquist, Stefano Ghirlanda, Daniel Lundqvist, and Carl-Adam Wachtmeister. An ethological theory of attractiveness. In G. Rhodes and L. A. Zebrowitz, editors, *Facial Attractiveness: Evolutionary, Cognitive, and Social Perspectives*. London: Ablex, 2002.

Geoffrey Hall. *Perceptual and Associative Learning*. Oxford University Press, 1991.

John H. Holland and Judith S. Reitman. Cognitive systems based on adaptive algorithms. In D. A. Waterman and F. Hayes-Roth, editors, *Pattern-directed Inference Systems*. New York: Academic Press, 1978.

Rob C. Honey and Johan J. Bolhuis. Imprinting, conditioning, and within-event learning. *Quarterly Journal of Experimental Psychology*, (50B):97–110., 1997.

Masashi Kamo, Stefano Ghirlanda, and Magnus Enquist. The evolution of signal form: Effects of learned vs. inherited recognition. In *Proceedings of the Royal Society of London*, volume B269, pages 1765–1771, 2002.

Pier Luca Lanzi and Stewart W. Wilson. Optimal Classifier System performance in non-Markov environments. Technical Report 99.36, Dipartimento di Elettronica e Informazione - Politecnico di Milano, 1999.

Konrad Lorenz. Der kumpan in der umwelt des vogels. *Journal of Ornithology*, pages 137–215, 1935.

Marc Métivier and Claude Lattaud. Anticipatory Classifier System using behavioral sequences in non-Markov environments. In *IWLCS*, pages 143–162, 2002.

Ulrich Nehmzow. *Scientific Methods in Mobile Robotics - Quantitative Analysis of Agent Behaviour*. Springer, 2006.

Yael Niv, Daphna Joel, Isaac Meilijson, and Eytan Ruppin. Evolution of reinforcement learning in uncertain environments: A simple explanation for complex foraging behaviors. *Adaptive Behavior*, 10(1):5–24, 2002.

Ivan P. Pavlov. *Conditioned Reflexes.* London: Oxford University Press., 1927.

Ivan P. Pavlov. *Experimental Psychology and Other Essays*. New York: Philosophical Library, 1957.

John M. Pearce, Mark A. Good, Peter M. Jones, and Anthony McGregor. Transfer of spatial behaviour between different environments: Implications for theories of spatial learning and for the role of the hippocampus in spatial learning. *Journal of Experimental Psychology: Animal Behavior Processes*, (30):135–147, 2004.

Jose Prados and Ed Redhead. Preexposure effects in spatial learning: From gestaltic to associative and attentional cognitive maps. *Psicologica*, 23(Special Issue on Spatial Learning): 59–78, 2002.

Wolfgang Stolzmann. An introduction to Anticipatory Classifier Systems. In Wolfgang Stolzmann Pier Luca Lanzi and Stewart W. Wilson, editors, *Learning Classifier Systems. From Foundations to Applications*, pages 175–194. Springer-Verlag, 2000.

Richard S. Sutton. Reinforcement learning architectures for animats. In J.-A. Mayer and S. W. Wilson, editors, *From Animals to Animats: Proceedings of the First International Conference on Simulation of Adaptive Behaviour.*, pages 288–296. Cambridge, MA: MIT Press, 1991.

Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 1998.

Edward C. Tolman. *Purposive Behaviour in Animals and Men.* New York: Appleton., 1932.

Max Wertheimer. Laws of organization in perceptual forms. In *A Source Book of Gestalt Psychology*, pages 71–88. London, Routledge and Kegan Paul, 1938.

Stewart W. Wilson. Classifier fitness based on accuracy. *Evolutionary Computation*, 3(2):149–175, 1995.

Zhanna V. Zatuchna. AgentP model: Learning Classifier System with associative perception. In Xin Yao et al., editor, *Proceedings of the Parallel Problem Solving from Nature Conference (PPSN)*, volume 3242 of *Lecture Notes in Computer Science*, pages 1172–1182. Springer, 2004.

Zhanna V. Zatuchna. *AgentP: A Learning Classifier System with Associative Perception in Maze Environments*. PhD thesis, School of Computing Sciences, UEA, 2005.

Zhanna V. Zatuchna and Anthony J. Bagnall. Maze material for AgentP, 2005. URL `http://www.cmp.uea.ac.uk/Research/kdd/projects.php?project=17`.

Integrative Approaches to Machine Consciousness

5th - 6th April 2006

Organisers

Rob Clowes, University of Sussex
Ron Chrisley, University of Sussex
Steve Torrance, Middlesex University

Programme Committee

Igor Aleksander, Imperial College Lond.
Giovanna Colombetti, York University
Rodney Cotterill, Technical University of Denmark
Frédéric Kaplan, Sony Computer Science Laboratory
Pentti Haikonen, Nokia Research Center
Germund Hesslow, Lund University
Owen Holland, University of Essex
Takashi Ikegami, University of Tokyo
Miguel Salichs, University Carlos III
Ricardo Sanz, Polytechnic University of Madrid
Murray Shanahan, Imperial College London
Jun Tani, Brain Science Institute
Steve Torrance, University of Middlesex
Tom Ziemke, University of Skövde

Contents

On Architectures for Synthetic Phenomenology

Igor Aleksander
Dept. Of Electrical and Electronic Engineering,
Imperial College , London SW7 2BT
i.aleksander@imperial.ac.uk

Helen Morton
School of Social Sciences and Law
Brunel University, Uxbridge UB83PH
Also, Imperial College , London SW7 2BT
helen.morton@brunel.ac.uk

Abstract

Is synthetic phenomenology a valid concept? In approaching consciousness from a computational point of view, the question of phenomenology is not often explicitly addressed. In this paper we review the use of phenomenology as a philosophical and a cognitive construct in order to have a meaningful transfer of the concept into the computational domain. Two architectures are discussed with respect to these definitions: our 'kernel, axiomatic' structure and the widely quoted 'Global Workspace' scheme. The conclusion suggests that architectures with phenomenal properties genuinely address the issue of modelling consciousness and indicate and the way that a machine with synthetic phenomenology may benefit from the property

1 Introduction

In searching for computational models of being conscious, the detailed nature of internal representation is an important facet of the way that modelling is to be approached. Synthetic phenomenology is involved when two conditions are fulfilled: first there is a meaningful sense in which a first person may be ascribed to the model and second, when the architecture caters for an explicitable and action-usable representation of "the way things seem" within the machine. We take the view that rather than this being an idealist stance, it represents as close an approximation to "the way things are" as is permitted by the sensory apparatus of that organism. This is assumed to be sufficiently close to reality to enable the organism to take appropriate action in its world. So one expects to find accurate phenomenological representation in successfully evolved organisms, as a major distance between the representation and reality does not augur well for successful evolution.

The paper first reviews the reason that in philosophy, phenomenology had a firm foothold despite the fact that the appellative became used in a variety of ways. A brief discussion is included on Block's use of the word in the notion of 'Phenomenal consciousness' as being distinct from 'Access consciousness' and, particularly in the way that such concepts could feature in computational systems.

The concept of a 'depictive' representation is developed in this paper beyond that which has been discussed to date (Aleksander, 2005) to show that this is a central requirement for an architecture that could be said to be synthetically phenomenological. A set of architectural definitions is then developed that determines whether an architecture could be said to be phenomenological or not. Two known architectures are scrutinised from the point of view of these definitions: are own *kernel* architecture (Aleksander, 2005) and Shanahan's embodied version of Baars' Global Workspace architecture (Shanahan, 2005). This reveals that the issue of phenomenology can be considered for differing mechanistic descriptions, of which the two architectures are distinct examples. In the conclusion we argue that the material in the paper indicates that architectures that are phenomenological have characteristics of being conscious that enhance their use both as explanatory tools and, possibly, functional artefacts. We shall first review issues that go under the heading of Phenomenology *and italicise strands that are*

taken up in discussing the implication for synthetic systems and their architectures discussed later in the paper.

2 Phenomenology

2.1 Definition

In the broadest terms, phenomenology is the word given to studies of consciousness which specifically start with the first person. In other words, introspection is an important facet of the discussion. This distinguishes phenomenology from other forms of philosophy, say, ontology, which asks what it is for an object to *be* conscious. One should also distinguish 'a phenomenon' from other philosophical constructs such as 'qualia' which relate to sensational primitives such as 'redness' or 'the sweet smell of a rose'. In general, phenomenologists like to extend the definition beyond the immediate sensation to more compositional structures of experience such as enjoying a game of tennis or the experience of having tried a new restaurant. This also aids action in the world and the generation of descriptive language in the case of humans or human-like machines.

Conforming with the above definition, the 'kernel' architecture we shall discuss in this paper was synthesised through a process of using introspection to discover design principles. This led to a consideration of ways that this work contributes to the formation of a synthetic phenomenology paradigm.

2.2 Past Usage

It is noted that in the history of philosophy, phenomenology is sometimes treated as the study of consciousness itself. For Franz Brentano (1874 trans. 1995) phenomena *are* acts of consciousness, they are the contents of mind. They stand in relation to physical phenomena that are perceived in the world by intentionally creating meaning of physical elements of the world in the mind. This first-person, descriptive character of a phenomenon has remained the hallmark of the work of later phenomenologists. Of these, Edmund Husserl (1913 trans. 1989), also focuses on the meanings the mind creates when contemplating the real world. This position addresses the mental object beyond just its real-world shape. So a stick may have the ability to dislodge a banana off the branch of a tree, enhancing the phenomenology of the stick by a mental vignette of the action of dislodging the banana.

Martin Heidegger (1975, trans. 1982) maintained that setting ontology (what it is to be conscious) apart from phenomenology could be an error. He suggests that it is actually linked to the phenomenology of the first person sensation of being a self in an external world. *See the influence of this in what we shall call 'axiom 1'.* Given Sartre's socio-philosophical observations on phenomenology as a literary examination of one's own experience and Maurice Merlau-Ponty's linking of phenomenology to personal experiences of one's own body (1945, trans. 1996) this becomes important particularly for those who discuss consciousness in the context of embodied robots.

The body's muscular activity is a key element in the 'kernel' architecture to create 'depictions', that is sensations of being an entity in an out-there world. As will be seen, Shanahan argues that embodiment is essential to have an experiencer.

2.3 Materialist Concerns

Gilbert Ryle in *Concept of Mind* (1949) argued that linguistic descriptions of mental states are a direct way of expressing phenomenology. This was possibly erroneously discredited by many materialists who identified the mental state with the neural state. Clearly only some neural states support phenomenology as identified by Crick and Koch (2003). Only some parts of the entire neural state are responsible for personal sensation, the parts that are not, have been called by the authors the 'Zombie' regions of the brain. This appears to beg the question of how one distinguishes a neuron that contributes to conscious sensation from one that does not. *A possible answer was developed by Aleksander and Dunmall (2003) and Aleksander (2005). This draws attention to the fact that in the visual system only some neurons, those indexed by the motor areas of the brain, can fire in a way that correlates with elements of the visual sensation of being an entity in an 'out-there' world. This is summarised later in this paper.*

2.4 Access and Phenomenal Aspects

Ned Block (1995) has identified at least two salient functions of consciousness. The first he calls 'phenomenal' or P-consciousness to indicate the personal function of experiencing a mental state. He contrasts this with 'Access' or A-consciousness which is that function of consciousness which is available for use in reasoning, being 'poised' for action and the generation of language. Although he argues that both are present most of the time, conflating the two when studying consciousness is a severe error. Some evidence of Block's distinct forms of is drawn from the phenomenon of 'blindsight' where individuals with a damaged primary visual cortex can respond to input without reporting an experience of the input.

This is A without P. P without A is the effect that some unattended experience had happened previously (e.g. a clock striking) but the individual had only realised this later. That is, P without A covers the case that unattended input can be retrieved. This creates a mechanistic difficulty for the definition of phenomenal consciousness as, were it never to be brought into access format, it could not in any way be described as 'the way things seem'. In hard-nosed synthetic phenomenology it might be politic to concentrate only on things that have seemed or seem to be like something.

This implies that in architectures it is important to be clear about the way in which immediate perceptual consciousness interacts with awareness of past experience, which bears on the A/P discussion.

Blindsight has also entered the theories of 'enacted' vision proposed by Kevin O'Regan and Alva Noë (2001) who have broadly argued that 'representing' the visual world in any architecture, living or synthetic, is an error, as the world itself is representation enough for the system to act on in a physical way. Consciousness is then a 'breaking into' this somewhat reactive, autonomic process through mechanisms of attention.

It is known that in the brain there are unconscious sensorimotor processes of the O'Regan and Noë description that work in conjunction with conscious phenomenal processes. For example the oculo-motor loop that involves the superior colliculus is such a mechanism. We are not conscious of the retinal maps that are projected onto the superior colliculus. They lead, also unconsciously, to the saliency maps that partly determine eye movement which eventually leads to reconstructions of world-fixed representations much deeper in the visual cortex (the extrastriate regions according to Crck and Koch, 2003). The enacted-unconscious/depicted-conscious interaction is a useful concept that may be used in synthetic systems. We find it difficult to accept the 'hard' sensorimotor view that complete access to a visual world can be achieved without any phenomenal representation at all.

3. Phenomenology in Computational Models

There are two important computational issues we wish to stress here. The first is the nature of a third-person design of an object that is capable of first-person representation, and the second is the relationship of depiction to synthetic phenolenology.

3.1 The Third Person Design with First Person Within It.

Where, in philosophy, phenomenology starts with the first person sensation, we suggest that in computational modelling, a phenomenological model must, in the broadest terms, sustain representations that have first person properties for *the model itself*. There is no dualist slight of hand here as the designer of the system can happily retain a third-person view of what is being designed, given a theory of what in the design is necessary to achieve a first person for the mechanism. That is, despite starting with our own first-person sense, we can speak of the first person of others. Similarly, we can speak of the first person of a machine and, indeed, set out to search for mechanisms of such. This implies that, in vision, for example, there is a need to differentiate mechanisms that mediate the sense of presence of the organism in the world from those that are due to previous experience: memory of various kinds and imagination (for example, states induced by literature). That is, there needs to be computational clarity about how a first-person phenomenal state relates to the current world event, how meaning is assigned to this, how meaningful states arise even in the absence of meaningful sensory input and how a personal sensation of decisions about 'what to do next' can arise. In Aleksander and Dunmall, 2003 and Aleksander 2005 we have referred to a necessary property for the machine having a first person at all as being a 'depiction'. Here we set out this concept as a logical sequence.

3.2 Depiction and Phenomenology.

It is useful to define what we mean by a *synthetically phenomenological system*.

> **Def 1**: To be **synthetically phenomenological**, a system S must contain machinery that represents what the world and the system S within it *seem* like, from the point of view of S.

The word *seem* has been transferred from the phraseology of the earlier parts of this paper to stress that perfect knowledge of the world cannot be achieved if only because of the weaknesses of sensory transducers. But, it is stressed that living creatures, if we believe that they have phenomenological representations, will come to our notice only through successful evolution. Again we stress that this is due to some sufficiency in the similarity between what things seem like and how, in a sense important to the organism, they not only *seem like* but, as far as the organism is concerned, *they are.* To achieve this it is necessary that such a representation should fully compensate for trans-

ducer and body mobility. In earlier work we have called this a 'depiction' rather than a representation. To advance this prior work we develop a series of definitions and assertions about depictions that positions this work within the framework of phenomenology addressed earlier.

Def 2: A **depiction** is a state in system S that represents, as accurately as required by the purposes of S, the world from a virtual point of view within S.

Assertion 1: A depiction of Def. 2 defines the mechanism that is necessary to satisfy that a system be synthetically phenomenological according to Def. 1.

Assertion 2: If S is mobile and has mobile sensors, a depiction of Def. 2 can only be achieved if the mobile nature of S is combined with the information carried by the sensors. That is the 'where' of the elements of the world needs to be predicated on the 'body' parameters of S. (In vision, eye-movement clearly needs to be compensated to achieve a depiction).

Assertion 3: 'As accurately as required ..' in Def. 2, indicates that, given effectors with which to act on the world, the depiction should carry all the information needed for such effectors to be successfully deployed on the attended and desired elements of the world.

Assertion 4: 'As accurately as required ..' also sets determines the granularity with which the depiction may be achieved.

Assertion 5: While Def. 2 makes no call on a topological representation, it does require that differently positioned elements within the representation be indexed by the predicates introduced in assertion 2. In animal vision it is known that different attributes of a visual element (e.g. the colour and motion of a dot) are represented in different parts of the brain. What 'binds' them in our analysis is the indexing as clarified in the example below (see Aleksander and Dunmall, 2000).

Example of indexing: Participant X is fixating a cross in the centre of a screen. She is asked to identify the shape *s* and colour *c* of an object that will appear briefly on some other part of the monitor screen. Shape is represented in area P of her brain and colour in area Q. The eye driven by the superior colliculus will saccade to the position of the object. The signal issued by the eye movement is, say, a 2-dimentional vector $\boldsymbol{v}$. Then the depiction in P will be *s,* indexed by $\boldsymbol{v}$, say $s_{\boldsymbol{v}}$. Similarly, in Q we have $c_{\boldsymbol{v}}$. Assertion 5 states that the binding of *s* and *c* is due to the common indexing by $\boldsymbol{v}$: that is, $(s,c)_{\boldsymbol{v}}$.

It is the deeper contention of the depictive approach that $(s,c)_{\boldsymbol{v}}$ uniquely encodes X's phenomenal experience of the appeared object. Of course, away from this experimental example, the indexing, as indicated by a great deal of physiological evidence (e.g. Galletti & Battaglini, 1989) occurs over many areas of the cortex, giving the phenomenal experience of one sensory modality several dimensions possibly bound across modality boudaries. Touch together with vision are a commonly bound experience.

4. Architectures

By 'architecture' we refer to a structure that first, is made of several internal parts each of which performs a specified distinct function, and second, includes a full specification of the interconnections among these parts the inputs and a variety of outputs (e.g. language generators, physical actuators etc..). It is the contention of this paper that there exists a set of architectures that can support phenomenology for the organism that embodies the architecture. We shall first look at two specific architectures to assess some of the definitional material presented in section 3.

4.1 The 'Kernel' Architecture

It is hardly a coincidence that a prototypical architecture we have recently suggested (Aleksander, 2005) should be based on the notion of a depiction and can, therefore, be said to have phenomenal consciousness according to our criteria. We take a closer look at this scheme that is shown in Fig.1

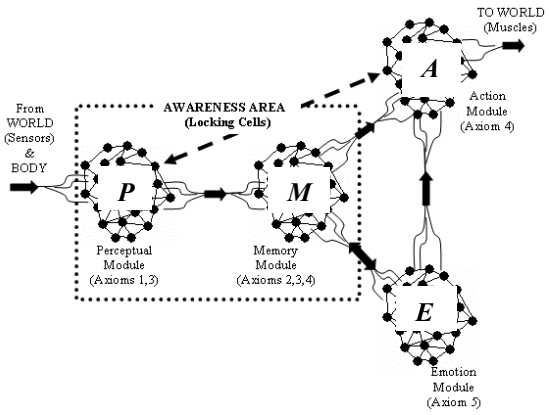


Figure 1. The 'kernel' architecture.

This architecture is based on the axioms of consciousness published in Aleksander and Dunmall (2003). For completeness, they are briefly listed in the Appendix of this paper.

These axioms start from a phenomenological standpoint as they are derived through an intro-

spective decomposition of the most significantly felt aspects of being conscious. Then it has been argued that the decomposition eases the transfer of these features into the synthetic domain.

Fig.1 is the result of this process. It consists of five modules each of which is considered to be a neural state machine (NSM) that operates in binary mode. That is, each connection carries a binary signal. We have ofen argued that any loss of generality due to the binary synthesis will be minor with respect to the behaviours that are being researched.

The binary NSM is specified as a six-tuple:
$<Ci,Co,Cf,Ct, I, O, F, T>_n$ where,
n is the module index,
Ci is a connection pattern of inputs (which may come from other modules or sensory inputs);
Co is a connection pattern of outputs (to other modules or system outputs);
Cf is the pattern of internal feedback connections.
Ct is the set of 'teaching connections' that determine the state of Co and Cf that becomes associated with Ci.
I, O, F, and T are the state sets of Ci,Co,Cf and Ct respectively.

Then, in the usual way with neural state machines, the states of F(t) and O(t) become functions of F (t-1) and I(t). These functions are determined by a training strategy which is expressed through T during a 'training phase'.

For example, an 'Iconic' mode of training is conventional with neural state machines of this kind (Aleksander and Morton, 1995). This ensures that, given that Ct and Cf have the same dimensions and Co=Cf, the network learns F(t)=T(t) as a function of I(t) and F(t-1).

Returning to Fig.1, the four axioms are implemented as follows. P is a 'Perceptual' NSM which is made to be phenomenological in the sense of the earlier definitions of this paper through the following design. The state F(t) is a reconstruction of the sequences of attended world inputs from sensory transducers over defined time windows (sometimes sliding time windows). The muscular effort required to attend to the elements of the world is shown as the link from the action NSM, A. In the animal visual system it is surmised that attentional shifts are driven by saliency maps in the superior colliculus. In specific studies of the visual system, this has been modelled as an additional part of the kernel architecture (See Igor Aleksander et al. 2001)

M is the memory and 'imagination' module. It is connected to P in such a way that for every reconstruction in P, a state in M is created. Sequences of reconstructed states in P can therefore be stored as state trajectories in M – they will have inherited the depictive, hence phenomenal properties of P.

P and M together form what we have dubbed 'the awareness areas' of the architecture. In the sense that one can perceive and recall at the same time, the two areas both contribute to the same phenomenal state The remaining modules of the kernel architecture are not depictive, hence not phenomenal, but add to the phenomenal existence of the system in the following way. As mentioned, A is the action area in which links between the state trajectories of the phenomenal areas are translated into action. But this is not automatic, it is surmised that volition and emotion as implemented in module E mediate this link. This was the subject of the contribution by Aleksander, Lahnstein and Lee in the AISB 2005 symposium on machine consciousness.

In summary, the kernel architecture is based *ab initio* on the intention of synthesising an architecture with phenomenological properties. This has also been guided by those who like Crick and Koch (2003) have been researching the neural correlates of consciousness in living organisms. We now consider a model that is more closely related to computational approaches of the functional kind.

4.2 Embodied Global Workspace

Bernard Baars' (1988, 1997) Global Workspace models have held sway in computational modelling of consciousness for some years. Baars considered how a large number of unconscious processes might collaborate to produce a continuum of conscious experience. In very broad terms, he answers the question through the architecture of Fig. 2.

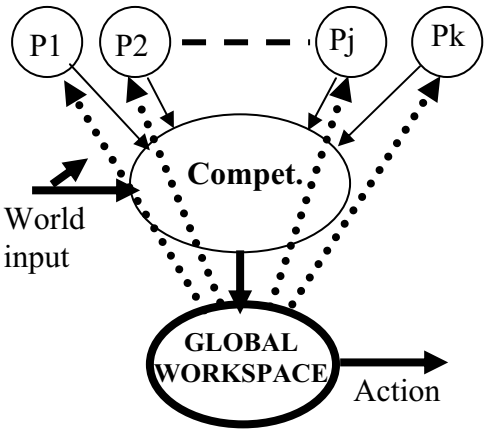


Figure 2 A sketch of Baars' Global workspace architecture.

The separate processes, P1 to Pk, said to be unconscious, compete to enter 'The Global Workspace'.

Such processes are often thought of as memory activities, say, episodic memory, working memory and so on. The competition is won by the process that has the greatest saliency at a given moment. This saliency is predicated by world input which sets the context for the competition. Of course, world input is also assumed to have direct influence on the unconscious processes P1 - Pk. Having entered the global workspace, the winner of the competition becomes the conscious state of the system. This is continuously 'broadacast' back to the originating processes that change their state according to the conscious state. This results in a new conscious state and so on, linking sensory input to memory and the conscious state. It is both general and useful for these separate processes to be modelled as NSMs as was done for the kernel architectures

Murray Shanahan (2005) points out that modelling of a conscious organism cannot proceed without that organism being embodied in some palpable world. Using the above Global Workspace model he argues that there can be no 'experiencer' in GW unless the model takes account of the "spatial unity of the body". It is this localisation in space that for Shanahan gives the model its "viewpoint on the world" which according to def. 1 makes it a candidate for phenomenal consciousness. Shanahan argues that denying this possibility, as is done by Block (1995), revives the dualist stance, putting phenomenal consciousness in the Chalmers-like 'hard problem' class, that is, a problem that cannot be reduced to physical structure and hence cannot be synthesized. And yet, the claimed 'point of view' of the embodied organism is undoubtedly a claim that this accords with definition 1 above of a phenomenal system. In terms Block's division into access and phenomenal consciousness, Shanahan implies that the embodied GW model addresses access consciousness, treating the phenomenal element as being an unnecessary appeal to a dualistic concept.

4.3 GW and Synthetic Phenomenology

While it seems entirely correct that without embodiment, GW does not include an experiencer, the question remains of how the experience stream in GW relates to the real world. We recall that in section 3.2 we have argued that a synthetic phenomenological system is achieved through a compositional representation of the world that is sufficiently accurate for the system be able to use its embodiment to control its world as accurately as possible. That is, it is the contention of this paper that depiction is the missing ingredient in making GW phenomenal. That is, phenomenal consciousness can occur in functional, physical systems, and the implication for the embodied GW system is that ***all*** the P1-Pk states need to be ***depictive*** for the GW state to be truly a model of a conscious state. Were this not the case, some translation into depiction would have to go along with the winning of the competition. Otherwise the spectre of purely arbitrary representations in GW remains. Shanahan is aware of this by requiring that the conscious broadcast back to the competing processes be in some way intelligible to these processes. But this still makes it hard to see how the states of the processes remain non-depictive when the state of GW might be depictive.

5. Discussion

In this paper we have explored the concept of synthetic phenomenology mainly by attempting to define the necessary features of an architecture that supports phenomenal consciousness within the broadest definition of the term. We brought the definitions to ground by considering two models that might be candidates for possessing these features. To conclude we raise and, using the material of this paper, attempt to answer five general questions that may be central to the existence of a synthetic phenomenology. The first of these addresses the architectures presented in the paper.

Can non-depictive representations be phenomenal?

It is the firm implication of this paper that this cannot be the case. It is depiction in a functional area which determines that the area contributes to the phenomenal sensation of the organism. Were this not the case, a human description of a state would require translation into phenomenal terms as such descriptions are of phenomena and not encoded states.

What is the difference between 'depictive kernel' and GW architectures in terms of synthetic phenomenology?

Clearly the depictive kernel architecture was designed with the purpose of creating a phenomenal representation within the system according to the definitions set out in this paper. This has the computational advantage of being able to be display on a screen the current phenomenal state of the machine enabling a designer's assessment of the interactions between both postulated conscious and postulated unconscious mechanisms in the generation of the phenomenology. The rules used in the synthesis involve depiction. Originally no phenomenal claims were made for GW, particularly in its practical form as synthesised by Stan Franklin (2003). However with the embodied GW work of Murray Shanahan, the question of the presence synthetic phenomenal consciousness acquires a

new urgency. In this paper we have maintained that were an architecture based on GW to have a phenomenal character, there must be a depictive activity in the processes that compete for entering the global workspace if the system is to be phenomenological. This creates problems as in our scheme of things, depiction in an area of the architecture implies phenomenal consciousness and GW sees the competing processes as being non-conscious. Therefore a phenomenal GW implies some sort of coming into consciousness in the GW area for reasons other than depiction. These have not yet been explained. Of course, the depiction idea can be rejected, but if not depiction, then what?

What is the use of synthetic phenomenology? Given the difficulties mentioned with embodied GW above, it is proper to ask why bother with phenomenology and why not settle for just access consciousness as implied by Shanahan (2005)? In the arguments of the current paper, phenomenology actually includes the purposes that are attributed to access consciousness. But such purposes are explicit and searchable through attentional mechanisms for reasons of accurate interaction with the environment (see assertions 3 and 4). This is not a Blockian confusion, but rather a suggestion that there may not be as clear-cut a functional/neurological distinction between access and phenomenal consciousness as Block seems to suggest. The A without P and P without A cases may be extreme conditions of a central phenomenon. In summary we argue that accurate interaction with, and thought about the real world is the purpose of phenomenology in a synthetic system.

Is synthetic phenomenology an oxymoron as it is the non-physical experiential side of consciousness and therefore eschews synthesis? Everything we have submitted in this article is a denial of the above proposition. Treating phenomenology as the 'hard' part of consciousness simply kicks it out of touch of science into some mystical outfield. We maintain that addressing it as a constructible concept removes the mysticism with which it might otherwise be associated.

Is synthetic phenomenology an arbitrary design option for models of consciousness? This paper regards models of consciousness without synthetic phenomenology as being valid only in a behavioural sense. That is, it is possible for a model to be given attributes of being conscious from its behaviour. Stan Franklin's Intelligent Distribution Agent (2003)is a good example of this class of system. Users think that they are dealing with an entity *conscious* of their needs. But if one were to argue that an architecture throws light on the mechanisms of consciousness in the brain it becomes mandatory to include phenomenal, that is depictive functions.

What research needs to be done in developing architectures with synthetic phenomenology? Referring to the kernel architecture there is much work to be done on modes of interaction between the modules. Current work includes a clarification of the way the emotion module E controls the link between the phenomenological P and M modules and the non-phenomenological action module, A. (fig. 1).

Illusions, ambiguous and 'flipping' figures are situations where phenomenology and reality part company. We are pursuing the mechanisms that, in the kernel architecture, would lead to the kind of perceptual instabilities associated with perceiving the Necker cube. This underlines the usefulness of synthetic phenomenology, as perceptual reversals may be measured in the depictive machinery and the conditions for such reversals studied. This is revealing of the interaction between phenomenal and non-phenomenal processes in the brain

In GW, architectures it would be interesting to clarify the causes of phenomenology in the GW area which are not present in the supporting competitive processes.

Appendix: Axioms of Being Conscious.

This is an introspective partitioning of five important aspects of being conscious

1. I feel as if I am at the focus of an out-there world.
2. I can recall and imagine experiences of feeling in an out there world.
3. My experiences in 2 are dictated by attention and attention is involved in recall.
4. I can imagine several ways of acting in the future.
5. I can evaluate emotionally ways of acting into the future in order to act in some purposive way.

References

Igor Aleksander, *The World In My Mind, My Mind In The World*' Exeter: Imprint Academic, 2005.

Igor Aleksander, Mecedes Lahnstein, Rabinder Lee: Will and Emotions: A Machine Model that Shuns Illusions, Proc AISB 2005 Symposium on New Generation Approaches to Machine Consciousness, 2005

Igor Aleksander, and Barry Dunmall: Axioms and Tests for the Presence of Minimal Con-

sciousness in Agents *Journal of Consciousness Studies.* **10**, pp 7-18, 2003

Igor Aleksander, Helen Morton and Barry Dunmall Seeing is Believing. *Proc. IWANN01*, Springer, 2001

Igor Aleksander, and Barry Dunmall:). An extension to the Hypothesis of the Asynchrony of Visual Consciousness, *Proceedings of the Royal Society ofLondon B* **267**: 200, 197–200.

Igor Aleksander and Helen Morton, *Introduction to Neural Computing (2nd Edition)*, London: Thomson Computer Press, 1995

Bernard Baars, In the Theater of Consciousness: The Workspace of the Mind , New York: Oxford University Press, 1997.

Bernard Baars, *A Cognitive Theory of Consiousness* , Cambridge: Cambridge University Press, 1988.

Ned Block, On a Confusion about a function of Consiousness, *Behavioural and Brain Sciences,* **18**, pp 227-287, 1995

Franz Brentano, *Psychology from an Emptirical Standpoint*, Trans: Rancurello et al. Routledge, 1995, Orig in German 1874.

Francis Crick and Christof Koch, 'A Framework For Consciousness' *Nature Neuroscience ,***6**, pp119 – 126, 2003 .

Stan Franklin, 'IDA a Conscious Artifact?' *Journal of Consciousness Studies,***10** (4-5), pp47-66, (2003)

Claudio Galletti and Paolo Battaglini: Gaze-Dependent Visual Neurons in Area V3A of Monkey Prestriate Cortex. *Journal of Neuroscience*, **6**, 1112-1125, 1989

Martin Heidegger, *The Basic Problems of Phenomenology,* Trans Hofstadter, Indiana University Press, Orig in German, 1975.

Edmund Husserl, *Ideas:A General Introduction to Pure Phenomenology,* Trans. Boyce Gibson, Collier, 1963. Orig in German, 1913.

Maurice Merlau-Ponty, *Phenomenology of Perception,* Trans Smith, Rotledge 1996, Orig in French, 1945.

Kevin O'Regan and Alva Noë, ., A Sensorimotor account of vision and visual consciousness. *Brain and Behavioural Sciences***, 24**(5) 2001.

Gilbert Ryle, *A Concept of Mind,* London: Hutchinson's, 1949.

Murray Shanahan, 'Global Access, Embodiment and the Conscious Subject'. *Jour. Of Consciousness Studies*, **12**, No 12, 2005 (in press)

Correlation, Explanation and Consciousness

Margaret Boden
Centre for Research in Cognitive Science
University of Sussex,
Falmer, Brighton, Sussex BN1 9QH, UK
maggieb@sussex.ac.uk

. **Abstract**

There's a lot of excitement about brain-scanning evidence for brain/consciousness correlations. Although the evidence is new, the idea isn't: Descartes formulated it nearly 400 years ago. However, he didn't regard mind-brain correlations as explanations – and neither should we.

Mere correlation between events in two domains is not enough for the one to be used as an explanation of the other. In addition, we need systematicity, isomorphism, and plausible (ideally, predictive) counterfactual conditionals.

There are a few (very few) examples where we already have those features, in respect of correlations between brain events and consciousness. In general, however, they can't be expected.

Even where we do have them, they leave the most difficult problem about conscious experience untouched.

The Problem of Inner Speech and its relation to the Organization of Conscious Experience: a Self-Regulation Model.

Robert Clowes
Centre for Research in Cognitive Science
Department of Informatics
Sussex University
Brighton BN1 9QH
East Sussex
UK
robertc@sussex.ac.uk

Abstract

This paper argues for the importance of inner speech in a proper understanding of the structure of human conscious experience. It reviews one recent attempt to build a model of inner speech based on a grammaticisation (Steels, 2003). The Steels model is compared with a *self-regulation* model here proposed. This latter model is located within the broader literature on consciousness. I argue the role of language in consciousness is not limited to checking the grammatical correctness of prospective utterances, before they are spoken. Rather, it is more broadly activity structuring, regulating and shaping the ongoing structure of human activity in the world. Through linking inner speech to the control of attention, I argue the study of the functional role of inner speech should be a central area of analysis in our attempt to understand the development and qualitative character of human consciousness.

1 Introduction

To introspection, for many of us, our mental life seems to have a constant accompaniment of inner speech. This speech is known in the literature under a number of names such as; the inner voice, the internal monologue, and is sometimes, subsumed into (the more general) stream of consciousness (James, 1890). It may also be linked to the generally pejoratively associate notion of 'voices in the head'. Understanding the nature of this phenomenon and its functional underpinnings, although of occasional interest in the history of psychology, has, in the last few years drawn the attention of many researchers into mind. There is however, much controversy about the precise nature of inner speech, its epistemic status and possible functional role.

Among psychologists, one means of accounting for inner speech is Baddeley's articulatory loop (Baddeley & Hitch, 1974), later rechristened the phonological loop[1] (Baddeley, 1997). This is considered to be a speech related working memory system.

Among philosophers, the notion of inner speech suggests privileged access to mental states, and this, at least in the 20th century, has invited great scepticism. The high-water marks of this scepticism are probably Ryle's (1949) *The Concept of Mind* and Dennett's (1991) *Consciousness Explained.* Dennett's view is complex on this question for although he ultimately doubts the strength of the epistemic warrant that can be given to the narrative stream of consciousness, and especially the subject's privileged position to report on its contents, he nevertheless argues that the subject's self-reports should be our starting-point. This is fundamental to his *heterophenomenological* method. This approach advocates

[1] Presumably this re-naming has something to do with thinking of inner speech as primarily an imaged sound, rather than unvoiced speech. The notion of a phonological loop seems to focus on the phenomenology of the passive, rather than active aspect of inner speech.

we need to attempt to offer some explanation of the importance attached to inner speech in phenomenological accounts.

A window into the phenomenology of inner speech is provided by Russell Hurlburt's *Descriptive Experience Sampling* technique (1990). Hurlburt uses an experimental technique in which subjects are cued by a small alarm device at various moments in their day, and then following protocols developed by Hurlburt, write down the details of their mental imagery at the moment that the alarm went off. He argues this technique allows us to systematically sample the qualitative characteristics of reported phenomenology[2]. It also allows us to describe some of the characteristics of inner speech, and inner imagery in general, in a much more elaborated fashion.

The content and form of this reported inner speech seems to be very diverse. Some people report the perception of being the author of voice-like inner speech; others, to hearing voices offering advice or consolation. Sometimes this voice appears to be their own, and sometimes the voice of another person. Some people report merely having the sense of experiencing language-like cognitive episodes without necessarily hearing any voices or having the sense of being the author of this speech. The variety of this speech might serve as some justification for the sceptics, or perhaps just evidence of the complexity and variety of the roles played by speech in our mental lives.

All of these phenomena seem to vary considerably both across individuals, within individuals at different times and places, and with regard to whatever activities they are at that moment engaged in. Hurlburt's work reveals much of the contents of consciousness appear to be composed of speech-like episodes. Except in cases of severe psychological disturbance or other abnormal functioning, the inner voice seems to be the constant accompaniment of human conscious life. But can we relate these accounts of the contents of conscious experience to language as vehicle?

Some recent accounts of cognitive role of language have brought to the fore they way that language may play a role, in sculpting, stabilising, and supporting forms of thought which would be otherwise impossible (Carruthers, 2002; Clark, 2004). Trying to forge a link theoretically between the phenomenological and functional aspects of inner-speech has proved so far a difficult task, but it is one upon which some progress has now started to be made.

2 – A re-entrance model of inner speech

Although traditional work on cognitive modelling made much use of more-or-less linguaform internal representations, following (if sometimes implicitly) some version of Fodor's (1975) Language Of Thought hypothesis, it has shied away from explicitly modelling the inner voice (cf. Dennett, 1994). Perhaps this is because of a worry that the inner voice might be either an epiphenomenon or user "illusion" (Dennett, 1991).

Recently work in machine consciousness has begun to treat the phenomenon of inner speech and its possible functional role more directly (Steels, 2003). Steels' earlier work used individual-based models in multi-agent systems to investigate the development of collective lexicons. More recently he has extended these models to attempt to model syntax.

[2] Although the beeps themselves are random, statistical techniques can be used to understand the distributions of reported mental-events types and indeed correlate them with other types of behavioural measures. (R. Hurlburt & Heavey, 2004)

In Steels' newer models agents are able to check the intelligibility of their own sentences by feeding back a prospective utterance through their language interpretation machinery prior to communication. Systems of agent with such *re-entrant* loops appear to be able to self-organise more complex grammars than would otherwise be the case. (Steels, 2003, 2005) Re-entrancy in Steels' models serves the role of checking the intelligibility of an utterance in their own reception systems. Systems of such "self-talking" agents seem to be able to achieve much more stable grammars as a result.

It seems that in order to develop the abilities to use complex syntax, re-entrant loops may be necessary. Steels is thus able to persuasively link re-entrancy to the generation of complex grammars in natural language and perhaps thereby provide a functional role for the inner-voice.

One problem for this work is that the everyday construction of grammatical sentences is usually considered a largely *unconscious* activity. In fact, the construction of grammatically correct sentences is often given as the paradigmatic example of what an unconscious cognitive process is like. Thus, there seems a little *prima facie* implausibility in correlating the phenomenological inner voice with a mechanism whose principle cognitive role is the construction of grammatically correct utterances. While Steels' arguments about the role of re-entrancy in the generation of complex grammars are convincing, arguably however the link with the inner-voice is less well-made.

One important caveat should be put on this observation. Insofar as we are treating the ontogenesis of language in young children, and the problems of developing capabilities to use a language for the first time, it may very well be the case that a large portion of the child's cognitive resources taken up in assembling and comprehending sentences and possibly they are much more conscious of this. It may turn out that the kinds of activities that Steels models in his experiments might very well turn out to play a central role in the consciousness of young children, and perhaps be the trailblazers for more elaborate forms of conscious inner loops to be developed later in their lives. A further task is to establish links between the Steels model and the account of the inner voice posited by theorists seeking to understand the re-organisation of cognition by language? Arguably his account could be made to fit with some of the recent accounts of language-for-thought that rely on the idea that language allows information to be passed between modules which wouldn't otherwise connect (cf- Carruthers, 2002). As the Steels model seems to have the language production and reception system rather separated from other forms of cognitive activity, it is difficult to say precisely how this relation could be established. Yet if the development of grammar turns out to be linked in this way to a re-entrant cognitive architecture, one can imagine how this architecture could become appropriated by other cognitive functions.

Although the Steels model offers an interesting attempt to show the functional importance of inner speech in order to stabilise the learning of grammars of certain complexity this model may be a special instance of the more general case where self-directed speech serves to scaffold and stabilise a whole range of cognitive functions. Yet could such a system also be linked to the phenomenology of inner-speech and the role of language in consciousness? More work clearly needs to be done in order to establish such a connection.

3 – A self-regulation model of inner-speech

Recent research conducted with Tony Morse (2005)[3] demonstrates how an alternative model of self-directed speech, still based on re-entrancy, might relate the inner-voice to a range of broader cognitive activities. The starting assumption for this work is that the cognitive role of language is better understood as one of sculpting or regulating cognitive activity rather than exhaustively representing the world (cf. Clark, 1996). Inner speech could here be seen as serving as a scaffold for developing and sustaining cognitive functions beyond the parsing and construction of meaningful and grammatical utterances.

In our model we compare a series of possible architectures for minimal cognitive agents which have to respond to instructions in order to fulfil externally indicated goals, i.e. moving objects around in a blocks world[4]. Our experiments compare several types of agents with differing architectures, some with word re-entrant loops and some without. All agents are implemented with simple recurrent neural networks that are evolved with a genetic algorithm in order to respond to commands by performing tasks. Some of the agents have architectures that allow the re-triggering of command reception systems internally.

The cognitive architecture of the 're-entrant' agents is arranged such that they can re-use the channels which are being used to signal commands to them to re-trigger their own behaviours. These channels allow at least the possibility of establishing new control circuits that use the same nodes that have previously been used to receive input from external 'words'. The thought here is that if there is some advantage to be had by re-using circuits developed to respond to words then the agents will take advantage of this source of useful adaptation. We find this is the case. Even such minimal agents can take advantage of these contingencies to develop word-based modes of self-regulation.

We show that agents with these 're-entrant speech' capabilities (as illustrated in **Figure 1**) perform considerably better on certain tasks. This is explained in greater detail in (Clowes & Morse, 2005). The basic finding is that agents that have architectures allowing the re-use of language for self-regulation achieve higher levels of performance more quickly and can stabilise them for longer that those that do not. Agents that are able to succeed in all task conditions make considerable use of auto-stimulation with words, i.e. they use re-entrant word nodes to self-trigger.

Re-entrance does not function in our models to facilitate merely communicative success or the generation and interpretation of complex linguistic constructions, but in the construction of more viable behaviours. Words here are appropriated in a way that is reminiscent of what Dennett calls auto-stimulation but not as a complex self-question (Dennett, 1991), but as new mode of self-regulation. This work then supplies at least a proof of concept that word-like constructs can be appropriated from a role in regulation from the outside (response to a command) to internal regulation (the agent self-regulating).

But linking such quite basic modes of auto-stimulation with words to inner speech, suggests a rather different picture of its un-

[3] A much more detailed examination of this work is now available in my unpublished DPhil thesis.

[4] NB. This is not exactly a blocks-world in the traditional sense. Rather, agents have extensive sensori-motor couplings with their limited world rather than it being specified in a purely abstract way. The agent architecture itself is an extension of an active vision model reported in experiments by (Floreano, Kato, Marocco, Sauser, & Suzuki, 2003)

derlying nature to that suggested by the Steels model. Inner speech is, I argue, the phenomenological dimension of internalised, word-based self regulation.

The phenomenological appearance of such speech, as speech, depends on it playing a similar attention focusing role as outer social speech often does. Further, I would conjecture that it relies on the same neural circuits, albeit appropriated for new self-directed functions.

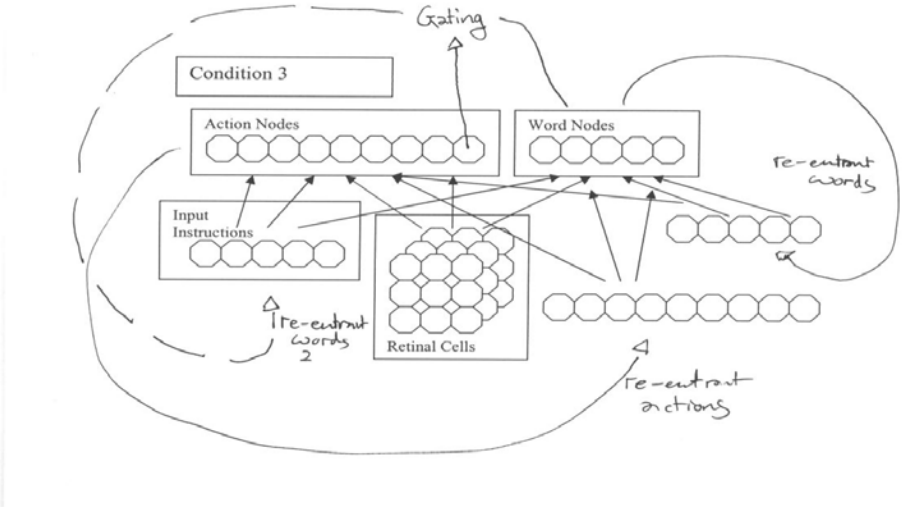


Figure 1 - The diagram shows an outline of the neural-architecture that is used in the experiments. The salient aspect is that when a gating neuron is switched on, activity from the output of the network can be fed back through the nodes that are used as input instructions. More detail on the architecture and some tasks can be found in (Clowes & Morse, 2005). Agents evolved in these conditions develop elaborate self-control loops and develop and stabilise solutions to more tasks than those that do not have such loops.

4 - A functional role for inner-speech

Normal intersubjective speech can certainly play a role in orienting attention, so why not internal speech? A shout in the street can cause an immediate refocusing of attention, e.g., hearing someone shout "mind the car!" as you were about to cross the road, would cause a fundamental reallocation of your attention.

If the inner voice could similarly be linked in some way to the allocation of attentional resources then there is the possibility that it may provide a window into the relationship between higher cognition and consciousness more generally. According to Vygotsky the internalisation of speech forms a whole new mode of attentional reorganisation.

Vygotsky (1986) emphasized the role of language in the development of control of action and ultimately of attention. His work provides an interesting possible way into the relationship between inner-speech and consciousness by looking at it through a developmental prism.

Vygotsky developed his ideas about the *internalisation* of language in part as a critique of Piaget's ideas about so-called *egocentric speech*. What Piaget called egocentric speech, and developmentalists tend to call today *private speech,* is a type of speech that children produce between the ages of about 4 and 7. It appears to be addressed toward the self and eventually seems to disappear.

For Piaget this speech occurs toward the end of his pre-operational stage and signifieds a still undeveloped ability to take, or imagine, the perspective of others. Social speech was thought to be built form this egoistic basis as children gain more experience that the point of view of others can be different (especially through argument with peers).

A longstanding controversy has arisen amongst developmentalists about the provenance and direction of this speech. Whether it is ultimately a disappearing artefact of early developmental egotism as Piaget argued in his early writings (1926), or alternatively the establishment of the bridge to linguistically controlled higher psychological function (Vygotsky, 1986 - originally 1934), either way this speech does not seem to serve a standard communicative function.

If Vygotsky's theory is correct, then inner-speech has at least its developmental precursors in this particular form of practically oriented speech found in children. If moreover inner-speech once fully internalised could come to play a role in allocating attention then this could provide a strong link between the internalisation of language and the constitution of human consciousness. Understanding inner speech may yet prove to be the royal road to understanding consciousness.

5 – Self-Restructuring through internalisation

Much of the theoretical work arguing that language plays a role in consciousness depends on the idea that language reshapes our underlying cognitive mechanisms in some way. Exactly how and to what purpose this functional re-organisation is achieved is currently part of a lively debate. The potential for re-using language as an addition to the brain's basic modes of organisation is something which is now starting to be taken very seriously in the philosophy of cognitive science (cf. Clark, 2004; Wheeler, 2004).

Dennett (1991) has argued that the development of the self-questioning form of self-directed speech is absolutely pivotal in the construction of human consciousness and its ability to sustain elaborate narrative threads. His view on this seems linked to his position that the form of human consciousness is the effect of installing what he calls a 'serial virtual machine' on parallel processing hardware. A range of accounts of the functional role of inner speech and its relationship with consciousness have also been put forward (Carruthers, 2002; Clark, 1998; Frawley, 1997) which seek to expand upon or restructure in various ways the sort of picture developed by Dennett. Although it seems possible that episodes of inner speech are epiphenomenal and fulfil no functional role in the organisation of consciousness, it is certainly too early to rule out the contrary possibility.

One can derive a further link between self-directed speech and the functional structure of consciousness from the psychopathological literature. Evidence seems compelling that the collapse of a normal inner voice in disorders such as 'thought insertion' is often correlated with catastrophic breakdowns for the organisation of individual consciousness (R. T. Hurlburt, 1993; Stephens & Graham, 2000). Disorders such as schizophrenia are sometimes theorised as control disorders and this idea gives us a way into establishing a possible link with the functional role of internalised speech (Gallagher, 2000). It points towards some quite central role for self-directed speech in the organisation of human consciousness, if not necessarily along the lines of Dennett's model.

One difficulty with this idea is that it is still very unclear at the level of sub-personal cognitive architecture how language can come to play the types of roles that are being ascribed it by the consciousness theorists. Yet there is a dearth of cognitive models that even attempt to show how such a reorganisation might happen[5]. However, it is possible to further analyse the model described above to give some insight into how attentional control through language internalisation might be established.

The model presented here gives one suggestion as to how the sorts of complex modes of self-regulation that seem bound up with human consciousness can get underway.

The simulation work with minimal cognitive agents shows that the re-use of public symbols in re-organising the ongoing activities of self can have cognitive benefits. These appear to go beyond being able to interpret and sustain more complex languages. Rather the internalisation of language in these models has more to do with the restructuring ongoing situated action.

Analysing the models further we found that the development of the ability to re-use a system of commands appears to move through broadly three control regimes.

1. Agents develop the capacity to respond to instructions. At this stage of development agents might be decribed as passive and do not use self-directed instructions very much.
2. Agents start to auto-stimulate with instruction nodes. This regime of self-control tends to produce ineffective and unstable systems of activity, (e.g. agents can sometimes perform the tasks well but very often do not).
3. Finally agents develop much more robust forms of self-control that rely on the ability to use new regimes of action made available by the self-directed loops.

Can these results be linked with Vygotsky's ideas about the establishment of new regimes of self-control through the internalisation of speech?

Vygotsky – to some extent developing the ideas of the Gestaltists[6] - argued that the development of self-directed speech was an form of self-prompting by which children come to de-centre and move themselves from one domain of situated activity (or as he might have termed it practical thought) to another. He saw this development as being centrally involved in the establishment of self-control and attention-regulation that are characteristic of human consciousness.

The work discussed above gives us a possible way of understanding the neural-dynamics underlying the establishment of this linguistic self-regulation.

6 – Inner speech and the modelling of consciousness

Notwithstanding current attempts to develop work in synthetic phenomenology (Chrisley & Holland, 1994), for now[7], hu-

[5] Despite these lacuna in more general work on cognitive modelling and the role of language some interesting work linking linguistic and cognitive function is starting to be done (Sugita & Tani, 2002). This work however encompasses quite a distinct formulation of the idea of a role for language in cognition as does the work reported here.

[6] Gestalt psychologists wrote a great deal on the problem of insight and how it was that a problem might suddenly be restructured such that it appears in an entirely new way. Kohler was one that held that tools could play a role

[7] Perhaps forever, cf, (Nagel, 1974)

man consciousness is the only type of consciousness which we know intimately. It seems unlikely that we can afford to ignore the relevance of the role of language in attempts to model it in machines, not to mention the project of building actually conscious machines.

Theorists as diverse and as historically distant as Vygotsky and Dennett have argued that self-directed speech plays a central role in the organisation and even the construction of human conscious experience. Work by Hurlburt and others appears to show that conscious experience abounds with episodes of internal speech.

If they are right and we are serious in our attempts to understand human consciousness with synthetic techniques, then we need to develop more advanced and explicit models of the role language might play in its functional organisation. The hypothesis defended here about the functional role of internalised speech is that it is a tool for the focusing or re-focusing of attentional resources.

Inner speech then appears to be of central importance because it gives an agent the capacity to restructure not just the external world but also itself. External activity in this way becomes redeployed toward inner restructuring. Simulation models such as those discussed above give us a unique mode of developing an understanding of the functional changes that underlie such a transition.

This internalisation model of self-directed speech can be used to provide an explanation of how language plays a role in creating the regimes of complex self-control and attention-regulation that are central to the sorts of consciousness that humans have (cf Donald, 2001). It does not attempt to address the question of why any experiences are conscious at all. However, it may allow us a new vantage point on their qualitative character.

According to the sensorimotor approach or 'skill theory' of conscious experience, "experience is not something we feel but something we do" (O'Regan, 2001). The character of perceptual experience, according to this theory, is given in the mastery of sensorimotor contingencies. These contingencies of self have there own governing laws just as any other complex physical system. Developing a mastery of these laws through autostimulation-with-words might be considered akin to the development of a new perceptual modality.

This mastery of the mechanisms of autostimulation-with-words affords the refocusing of one's own attention on self. This exercise of the contingencies of self can therefore be linked, more generally, to the qualitative analysis of consciousness in terms of sensorimotor contingencies (cf - O'Regan & Noë, 2001). Understanding this refocusing of attention might help us explain the uniquely human mode of the self's perceptual presence.

References

Baddeley, A. (1997). *Human Memory Theory and Practice*. Hove, UK: Psychology Press.

Baddeley, A., & Hitch, G. (1974). Working Memory. In G. A. Bower (Ed.), *Recent advances in the psychology of learning and motivation.* New York: Academic Press.

Carruthers, P. (2002). The Cognitive Function of Language. *Behavioral and Brain Sciences, 25*(6).

Chrisley, R., & Holland, A. (1994). *Connectionist synthetic epistemology: Requirements for the development of objectivity* (No. 353): COGS CSRP 353.

Clark, A. (1996). Linguistic Anchors in the Sea of Thought? *Pragmatics And Cognition, 4*(1), 93-103.

Clark, A. (1998). Magic Words: How Language Augments Human Computation. In P. Carruthers & J. Boucher (Eds.), *Language and Thought. Interdisciplary Themes* (pp. 162 - 183). Oxford: Oxford University Press.

Clark, A. (2004). Is language special? Some remarks on control, coding, and co-ordination. *Language Sciences, 26*(6), 717-726.

Clowes, R. W., & Morse, A. (2005). Scaffolding Cognition with Words. In L. Berthouze, F. Kaplan, H. Kozima, Y. Yano, J. Konczak, G. Metta, J. Nadel, G. Sandini, G. Stojanov & C. Balkenius (Eds.), *Proceedings of the 5th International Workshop on Epigenetic Robotics*. Nara, Japan: Lund University Cognitive Studies, 123. Lund: LUCS.

Dennett, D. C. (1991). *Consciousness Explained*: Penguin Books.

Dennett, D. C. (1994). The Role of Language in Intelligence. In D. C. Dennett (Ed.), *What is Intelligence*. Cambridge: Cambridge University Press.

Donald, M. (2001). *A Mind So Rare: The Evolution of Human Consciousness*. New York / London: W. W. Norton & Company.

Floreano, D., Kato, T., Marocco, D., Sauser, E., & Suzuki, M. (2003). *Active Vision & Feature Selection: Co-development of active vision control and receptive field formation. Complex visual performance with simple neural structures.* Retrieved 30 June 2004

Fodor, J. (1975). *The Language of Thought*. New York: MIT Press.

Frawley, W. (1997). *Vygotsky and Cognitive Science: Language and the Unification of the Social and Computationional Mind.* Cambridge: Harvard University.

Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends in Cognitive Sciences*.

Hurlburt, R., & Heavey, C. L. (2004). To Beep or Not To Beep: Obtaining Accurate Reports About Awareness. *Journal of Consciousness Studies, 11*(7), 113-128.

Hurlburt, R. T. (1990). *Sampling Normal and Schizophrenic Inner Experience*. New York: Plenem Press.

Hurlburt, R. T. (1993). *Sampling inner experience with disturbed affect.*: Plenum Press.

James, W. (1890). *The Principles of Psychology*.

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review, 83*, 435-450.

O'Regan, J. K. (2001). Experience in not something we feel but something we do: a principled way of explaining sensory phenomenology, with Change Blindness and other empirical consequences.

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences, 24*.

Piaget, J. (1926). *The Language and Thought of the Child*: Routledge and Kegan Paul.

Ryle, G. (1949). *The Concept of Mind*. Chicago: The University of Chicago Press.

Steels, L. (2003). Language Re-Entrance and the 'Inner Voice". In O. Holland (Ed.), *Machine Consciousness*. Exeter: Imprint.

Steels, L. (2005). Constructivist Development of Grounded Construction Grammars.

Stephens, G. L., & Graham, G. (2000). *When Self-Consciousness Breaks*: MIT Press.

Sugita, Y., & Tani, J. (2002). A connectionist model which unifies the behavioral and the linguistic processes. In M. I. Stamenov & V. Gallese (Eds.), *Mirror Neurons and the Evolution of the Brain* (Vol. 42).

Vygotsky, L. S. (1986). *Thought and Language* (Seventh Printing ed.): MIT Press.

Wheeler, M. (2004). Is language the ultimate artefact? *Language Sciences, 26*(6), 688-710.

Playing to be Mindful
(Remedies for Chronic Boxology)

Ezequiel Di Paolo
Centre for Computational Neuroscience and Robotics
University of Sussex,
Falmer, Brighton, Sussex BN1 9QH, UK
ezequiel@sussex.ac.uk

Abstract

There is a widespread misconception among critics of the dynamical systems approach to cognition: the emphasis on embodiment and situatedness has given the wrong impression that the only cognitive activities that can be explained under this paradigm are those concerned with ongoing coping with the current situation. To say that the body is actively situated in a world is only to highlight the most fundamental aspect of all cognitive activity. There is no doubt that the dynamical systems approach has already proven immensely more successful in such cases than traditional computational approaches. Even so, as soon as we move to other, more human, cognitive performances, such as planning or imagining, we must, critics predict, return to the tenets of cognitivism/ computationalism in some updated form, or worse still, to some kind of hybrid stance. Here I briefly examine the foundations of this claim (and find there aren't really any).

On the positive side, I raise the issue of what is the best route for connecting sensorimotor and situated intelligence with (some) human styles of cognitive activity (misleadingly characterized as "decoupled"). A dynamical systems approach is already useful because it forces us to formulate the questions that traditional representational approaches felt unnecessary to ask since they answered them almost axiomatically. What is to represent? How is it possible to alter the meaning of a situation? What sort of system is a cognizer such that the world is meaningful for her? How can a cognizer act autonomously in accordance with meanings not yet established by the situation but by her own actions?

I will very briefly discuss the life/cognition continuity thesis and show how it reveals fundamental issues about agency and sense-making that allow us to begin to answer some of these questions. A powerful methodological guidance is found in Hans Jonas's work on value-generating activities and the evolutionary/historical thread of increased mediacy in cognition. Following a developmental version of this thread, a large part of this presentation will be devoted to examining pretend play (in authors such as Lev Vygotsky, Maxine Sheets-Johnstone, and Margaret Donaldson) as a particularly relevant activity for understanding how transitions to freer forms of meaning manipulation are inherently embodied and dynamical in nature. This will suggest new vistas and new challenges to synthetical approaches like evolutionary robotics.

The XML Approach to Synthetic Phenomenology

David Gamez*
*Department of Computer Science
University of Essex
Colchester
C04 3SQ
UK
daogam@essex.ac.uk

Abstract

One of the major challenges in synthetic phenomenology is to find a way of systematically describing artificial non-conceptual phenomenal states. This paper puts forward a solution to this problem that uses three different XML files to describe a machine's structure, internal states and phenomenology. The advantages of XML are that it can be read by both machines and humans, it is good at capturing hierarchical relationships between data and it can be automatically generated, analysed and archived. XML could also be a useful tool for other methods of representing non-conceptual mental content, such as content realization and ability instantiation. Furthermore, as scanning technologies develop, the XML approach could be applied to the neurophenomenology of humans, which would serve as a foundation for a more scientific psychology of both humans and machines and facilitate precise comparisons between the two. The XML approach outlined in this paper will be used to describe the synthetic phenomenology of Holland's and Troscianko's CRONOS robot that is currently under development at the University of Essex and the University of Bristol.

1 Introduction

Synthetic phenomenology is a recently emerging discipline that aims to describe the phenomenal states of artificial systems. This is essential for the monitoring and debugging of machine consciousness and it could also address concerns about the possibility of suffering in machines. This paper puts forward an approach to synthetic phenomenology that uses three different XML files to describe a machine's structure, internal states and phenomenology. The advantages of XML are that it can be read by both machines and humans, it is good at capturing hierarchical relationships between data, it can be automatically generated, analysed and archived, and it avoids many of the pitfalls and presuppositions of natural language. As scanning technologies develop it may also be possible to use XML in neurophenomenology, which would allow detailed comparisons between human and artificial systems.

The first part of this paper covers some of the limitations of natural language descriptions of the phenomenology of non-human systems. After setting out the advantages of an XML approach, the central section outlines one way in which XML representations could be used in synthetic phenomenology. This is not intended to be a final and definitive methodology, since there are no doubt better ways of applying XML to this area. However, by presenting one way in which it could be done I hope to make the case that XML could be a very useful tool for the phenomenology of artificial systems.

2. Problems with Describing the Phenomenology of Non-Human Systems

Phenomenology, especially in the work of Husserl and Heidegger, derives its significance from the claim that the phenomena we experience are as important and substantial as the physical world described by science, which is often portrayed as a secondary interpretation of the phenomena. In this way phenomenology sets itself up with an 'objective' field of phenomena that are assumed to be the same for everyone and can be unproblematically described in natural human language The problem with this approach is that these assumptions about common experience start to break down once phenomenology is applied to the experiences of infants,

animals and robots. To illustrate this problem, I will consider a short extract from Wordsworth (2004), which contains a fairly straightforward description of daffodils in natural human language:

> When all at once I saw a crowd,
> A host, of golden daffodils,
> Beside the lake, beneath the trees,
> Fluttering and dancing in the breeze.

Most people have had the experience of daffodils fluttering and dancing in the breeze and when Wordsworth's description is read by humans, they can readily imagine a similar past experience and understand his words well enough. However, even this straightforward description presents problems since it is extremely vague and imprecise and each reader will imagine the daffodils differently. More serious problems start to arise when we try to use ordinary language to describe the experiences of an infant placed in front of a field of daffodils. As Chrisley (1995) points out, we cannot simply say that the infant sees a host of golden daffodils because the infant has a preobjective mode of thought, which is unable to locate the daffodils within a single unified framework. Adults understand daffodils as something objectively located in three dimensional space, whereas infants do not necessarily continue to believe in the existence of the daffodils when they are occluded. In the adult and infant the word "daffodils" refers to two different concepts and experiences. As Chrisley puts it: "The infant's concepts are not fully objective and are therefore non-conceptual. To ascribe conceptual content to the infant in this case would mischaracterize its cognitive life and would not allow prediction or explanation of the infant's behavior." (Chrisley, 1995: 145).

These problems become even more difficult when the attempt is made to describe the phenomenology of a non-human animal, such as Nagel's famous bat (Nagel, 1974). When a bat flies over a field of daffodils it receives a complex pattern of returning ultrasound pulses, which are processed into phenomenal experiences that are likely to be very different from our own. Sentences like "the bat is experiencing a host of golden daffodils" are at best an extremely misleading description of the bat's phenomenology.

The same problems are encountered when attempting to describe the phenomenal experiences of artificial systems. Whilst we may have grounds for attributing phenomenal consciousness to some robots, we have almost no basis for believing that they will have the *human* phenomenal experience of yellow when daffodils are placed in front of them, or even that they will have the same experience of yellow as each other. Robots may also be built that have unconscious daffodil recognizers, so that they are only conscious of the abstract presence or absence of daffodils. Other robots might only be capable of processing stationary daffodils, leading to highly divergent phenomenal experiences that cannot be captured in ordinary language.

Natural language evolved to describe human experiences and so it is not surprising that it is very bad at describing the phenomenology of bats and robots. Synthetic phenomenology needs a better and more systematic way of describing the phenomenal states of artificial systems and the central claim of this paper is that XML representations are more appropriate for this task. After setting out the advantages of an XML approach, section 4 will demonstrate how it can be used to describe synthetic phenomenal experiences in a systematic manner.

3. Advantages of XML for Synthetic Phenomenology

The eXtensible Markup Language (XML) is a platform-independent way of structuring and organising data so that it can be easily shared between systems. XML is stored as plain text and it has a tightly structured format that enables the relationships between data items to be easily expressed. It is also possible to validate the structure of an XML file without any prior knowledge of its form. XML is starting to be used widely and there are a number of reasons why it would be suitable for synthetic phenomenology:[1]

1. XML is much more precise and highly structured than natural language, which allows it to describe complex nested hierarchies and represent the relationships between different pieces of information. This also enables easy cross referencing between different files.
2. XML can describe low level details of the system's hardware, but it can also abstract from them so that high level comparisons can be made between machines with different architectures and between humans and machines. Whilst two systems' lower levels might be different – perhaps using neurons or silicon - the higher levels are likely to be more similar, allowing direct comparisons between different systems once everything is encoded in XML.
3. XML can be written and read by both machines and humans. When doing simple small scale analyses it is useful to be able to manually read and edit an XML description of a machine's inner state. However, it is also very easy to automatically generate and analyse the state of a machine using XML, for example by writing programs

[1] A good XML tutorial can be found at: http://www.w3schools.com/ xml/default.asp

that look for phenomenal mental content using different theories of consciousness.

4. Its human and machine readability also make XML good for debugging consciousness. Once you have a highly structured representation of a machine's inner state and a methodology for analysing this for phenomenal consciousness, you can see how the machine's phenomenal states can be improved or increased.
5. XML is easy to archive, either by converting the XML files into a database format or by storing them directly. Sequences of mental content that are stored in this way can be examined later offline.
6. XML is a good foundation for the other techniques for representing non-conceptual mental content, such as those suggested by Chrisley (1995).
7. XML is very flexible. In addition to tags and data, XML can contain references to external files, pieces of code and equations. This enables it to include features that cannot be precisely described in human language.

Although these advantages also apply to some of the alternatives to XML, such as JSON, YAML and OGDL, the popularity of XML and the availability of good parsers in most programming languages make it the best choice for the approach to synthetic phenomenology that I am setting out in this paper.

4. The XML Approach

This section outlines one way in which XML could provide a systematic framework for describing the phenomenology of artificial systems. This approach works using three separate but interlinked XML representations:

1) *System.* A systematic description of the system and its sensors.
2) *Test Suite*. Identifies active elements within the system that are systematically correlated with outside events impinging on the sensors. This treats the machine as a complete unknown that is systematically probed by exposing it to stimuli and measuring changes in its internal state. During the generation of the test suite no attempt is made to say what the stimuli might be like for the machine, although human descriptions are included to help with later analysis.
3) *Mental Content.* If the test suite is constructed in enough detail, a good idea should be gained about the range of correlations between internal states of the machine and activation of the machine's sensors by the outside world. However, at any point in time only a small proportion of the potentially active elements will be active and this set of currently active elements are recorded in a third XML representation of the machine's mental content. This includes tags to indicate whether it is phenomenal mental content, which are filled in at a later stage by programs designed to analyse the system, test suite and mental content XML for signs of consciousness.

The XML structures that could be used to contain the data for each of these stages will now be covered in more detail.

4.1 System

The system XML file describes the structure of the system, including sensors, actuators and internal components. This is needed to clarify the range of tests that could be applied to the system and to help with the identification of potential phenomenal states. Some extracts from an XML file describing a typical system are given below:

```
<system>
  <description>Robot</description>
  <sensor id="1">
    <type>light</type>
    <shape>rectangle</shape>
    <width>400</width>
    <length>300</length>
    <coordinate_system>Cartesian
                </coordinate_system>
    <wavelength_range>0.7–0.4
                </wavelength_range>
  </sensor>
  <!-- Add more sensors here -->

  <actuator id="1">
    <type>motor</type>
    <location>wheels</location>
  </actuator>
  <!-- Add more actuators here -->

  <neuron id="1">
    <position>2,3,3</position>
    <type>pyramidal</type>
    <algorithm>Leaky integrate and
                    fire</algorithm>
  </neuron>
  <!-- Add more neurons here -->

  <connection id="1">
    <presynaptic_neuron>1
              </presynaptic_neuron>
    <postsynpatic_neuron>3
              </postsynpatic_neuron>
    <synapse_type>excitatory
                    </synapse_type>
    <weight>0.9</weight>
    <delay>22</delay>
  </connection>
  <!-- Add more connections here -->

</system>
```

Brief explanations of some of the more important tags are as follows:

<**sensor**> A sensor sensitive to light, touch or sound, for example.
<**actuator**> An actuator, such as a motor or hydraulic piston.
<**neuron**>, <**connection**> In this system the internal states are held in neurons, whose parameters are specified here along with the connections between them. Other systems might use Bayesian networks or first order logic to hold their internal states.

4.2 Test Suite

A test suite is a systematic way of linking the presence of events and objects in the environment to changes in the machine's inner state. To generate a test suite the system is probed using a number of different tests and correlations between the stimulus and the machine's state are recorded as a list of active elements. The behaviour of the machine is also treated as data that is correlated with its internal states. To avoid presuppositions about three dimensional space, the input to the machine is specified in terms of changes in the machine's sensors and not as the presentation of three dimensional objects. With systems based on real or simulated neurons the test suite could be created by following the traditional approach of recording from neurons or groups of neurons. Systems along the lines of Franklin's IDA (Franklin, 1998) could be tested by using a debugger to monitor which variables or memory locations change in response to environmental stimulation. This avoids problems raised by Searle (1980) about the difference between manipulating a symbol and understanding a symbol since no assumptions are made about the meaning of any of the system's internal states.

A comprehensive test suite needs to be designed with care so that it can probe all possible sensitivities of the machine and specify them as precisely as possible. This could start with simple low level features, such as points, lines, and edges and work its way up to more abstract stimuli, such as faces and houses. All of these single modality tests would have to be combined with input from other modalities, such as audition, proprioception and sensation. They would also have to be carried out whilst the machine is engaged in different activities, such as looking to the left, moving forward, and so on, to take account of sensorimotor contingencies. Whilst this sounds like an enormous quantity of work, initial tests of this type are likely to be carried out on very simple machines and as the methodology develops it will be possible to automate the creation of the test suite by writing programs that examine the system XML file and generate a comprehensive series of tests. The tests could also be automated in many cases by simulating the input to the sensors. Some sample extracts from a test suite XML file are given below:

```
<test_suite>
  <test id="1">
    <human_description>Moving
         forward towards point of
         light</human_description>
    <sensor_input>
      <sensor>1</sensor>
      <type>light</type>
      <size>5,5</size>
      <location>55,44</location>
      <wavelength>0.55</wavelength>
      <file>Test1.dat</file>
    </sensor_input>
    <!-- Add more sensor inputs -->

    <actuator_output>
      <actuator>1</actuator>
      <type>motor</type>
      <direction>clockwise
                        </direction>
      <speed>5</speed>
    </actuator_output>
    <!-- Add more actuator outputs -->

    <active_element>
      <type>neuron population</type>
      <neuron id="27">
        <firing_rate>0.88
                      </firing_rate>
      </neuron>
      <!-- Add more neurons -->

    </active_element>
    <!-- Add more active elements -->

  </test>
  <!-- Add more tests -->

</test_suite>
```

Some of the more important XML tags are as follows:

<**test**> A test that is applied to the machine to probe its responses to a particular stimuli. Tests that do not activate any elements do not need to be included.
<**human_description**> Description of the stimulus by humans, which may be useful as part of the process of describing the phenomenology of the machine.
<**sensor_input**> Input is defined in sensory rather than world coordinates. This is to avoid the presupposition of three dimensional space that might be made if we talked about presenting a round object at a distance of three metres, for example.
<**actuator_output**> Any actions carried out by the machine whilst the stimulus is being presented.
<**active_element**> The part of the machine's inner state that is activated by the test. In a neural system

this could be a single neuron or a population of neurons with a particular distribution of firing rates. In a more traditional computer system this could be a list of memory locations that are altered by the stimulus. Active elements are defined in relation to the test stimuli that activated them and have no meaning outside of this context.

4.3 Mental Content

Only a small proportion of the elements inside the machine that respond to stimuli are likely to be active at any point in time. The currently active elements are stored in the mental content XML file, along with the active connections between them. This mental content is capable of influencing actions and could be involved in planning. For example, if a machine has a group of simulated neurons that selectively respond to images of houses, then these neurons could initiate motor patterns that cause the sound "house" to be emitted. The house-sensitive neurons could also become activated when the machine was offline, leading to an experience analogous to imagining or dreaming about a house. Some of this mental content may be conscious and a tag has been included to record whether this is the case. The contents of this tag are filled in at a later point when the system, test suite and mental content XML files are examined according to a particular theory of consciousness (see next section). Sample extracts from a mental content XML file are given below:

```
<mental_content id="66">
  <time>4010551056</time>
  <active_element>
    <id>2</id>
    <intensity>0.7</intensity>
    <phenomenal>yes</phenomenal>
  </active_element>
   <!-- Add more active elements -->

  <active_connection id="3">
    <type>synchronisation</type>
    <from>1</from>
    <to>2</to>
  </active_connection>
  <!-- Add more active connections -->

</mental_content>
```

Some of the more important tags are as follows:

<**active_element**> Reference to one of the active elements defined in the test suite along with some of its current properties.
<**active_connection**> An active connection could be synchronisation between firing neurons, active processing by the CPU or simultaneous broadcast along a radio link. Since active connections are not necessarily topologically bound they are defined separately from the static connections in the system file.
<**phenomenal**> Records whether this active element is phenomenal mental content. The contents of this tag are filled in by examining the system, test suite and mental content XML files for signs of phenomenal consciousness.

4.4 Phenomenal Mental Content

The final stage in the description of the phenomenology of the machine is the identification of the parts of the mental content that are likely to be phenomenally conscious. This is done by analysing the system, test suite and mental content XML files using a theory of consciousness. It is highly likely that different theories of consciousness will make different predictions about the phenomenal mental content of the machine, which provides a good way of discriminating between them by comparing their different predictions with first person reports about phenomenal states.[2] This process of identifying the phenomenal mental content will now be illustrated using Tononi's φ, Aleksander's axioms and Metzinger's constraints.

4.4.1 Tononi's φ

According to Tononi (2004) consciousness is linked to a system's capacity to integrate information. This is precisely quantified by Tononi as the number φ, which is the amount of effective information that can be exchanged across the minimum bipartition of a complex, where a complex is the subset of elements with $\varphi > 0$ and no inclusive subset of higher φ. Whilst there is not space to go into the details here, the system, test suite and mental content XML representations outlined in this paper would make it easy to calculate the amount of φ and pinpoint the active elements with high φ that are likely to be phenomenally conscious. It would even be possible to add a φ tag to the active elements within the mental content XML file.

4.4.2 Aleksander's Axioms

Aleksander (2003) put forward five axioms as a set of mechanisms that are thought minimally necessary to underpin consciousness. These are depiction, imagination, attention, planning and emotion. Although these axioms are not necessarily sufficient for consciousness, they are a good starting point for deciding whether a machine might be capable of conscious states and the XML approach offers a good way of analysing a system for their presence. For example, the test suite XML of an agent that

[2] There may also be ways of indirectly testing the predictions made by different theories of consciousness.

was capable of depiction would contain active elements linked to external stimuli, and an agent would be experiencing imagination when its mental content XML contained active elements that were linked in the test suite to different stimuli from the ones that are currently present. For example, an active element might be linked to apple stimuli in the test suite and yet be part of the agent's mental content when only bananas are in its field of view. One way of identifying the axiom of attention would be follow Damasio (1999) and Metzinger (2003) and look for active connections between active elements linked to the agent's self model and active elements associated with external content. Emotion could be discovered by looking for active elements associated with certain body states.[3]

4.4.3 Metzinger's Constraints

Metzinger (2003) set out eleven constraints that mental content must conform to if it is to be conscious. There is not space to go into the constraints in detail here, but the three most important, which are used to define a minimal notion of consciousness, are the activation of a coherent global model of reality (constraint 3) within a virtual window of presence (constraint 2) both of which are transparent (constraint 7). A system whose mental content conformed to these constraints would have a phenomenal experience of "the presence of one unified world, homogenous and frozen into an internal Now, as it were." (Metzinger, 2003: 169).

The identification of which parts of the mental content conform to Metzinger's constraints is easier than it seems because Metzinger provides very detailed descriptions of the informational, representational, computational and functional characteristics of the constraints along with some likely neural correlates. All of this can be fairly easily extracted once detailed and systematic XML representations have been created for the system. For example, the presence of constraint 3 (integration within a global model of reality) could be established by looking at the active connections between active elements or possibly using Tononi's methodology. Some of the other constraints, such as transparency, may come for free on systems whose internal states do not have any sensors that could make them objects of representations.

[3] The identification of planning in an agent's XML descriptions would require a fully temporalised version of the XML approach, which is not covered here.

4.5 A Description of the Synthetic Phenomenology?

Given the history of phenomenology, we might expect the final outcome of synthetic phenomenology to be a natural language description. Even if we cannot achieve this at present, it might be thought that this should be the final goal of the procedures outlined in this paper. Viewed from this perspective, the system, test suite and mental content XML would only be the preparatory stages for a traditional phenomenological account of the experiences of COG, CRONOS or IDA.

However, the problems discussed in section 2 make it unlikely that we are ever going achieve fluid natural language descriptions of non-human systems. Instead, it might be much better to treat the XML representations as the best description that we are going to get of the phenomenology of an artificial system. This has the great advantage that it is possible to see what you cannot say. We don't have adequate words in human language to describe a system that can only experience vertical lines, but we can represent such a system accurately using XML, and by looking at the XML we can start to understand how much and how little we can imagine what it is like to be such a system.

The XML descriptions also offer a good starting point for other ways of describing the phenomenology of artificial systems. The suggestions made by Chrisley (1995) about conceptual subtraction, content realization, ability instantiation and self instantiation could all be implemented automatically once the XML formats have been defined. XML would also enable precise comparisons with humans that have deficiencies in the same areas as a machine, and we could use the first person descriptions of these patients to help us imagine what it is like to be such a system. As scanning technology improves, the application of this approach to normal and brain damaged patients will become easier. Research by Kamitani and Tong (2005) on neurophenomenology using combinations of voxels suggests that it might even be possible to start this work today.

5. Discussion

One of the first issues that must be clarified about the XML approach to synthetic phenomenology is that it makes no presuppositions about whether any particular machine is the sort of system that is capable of supporting conscious states. Robots, stones and human beings are all systems that are capable of internal states; all three can be analysed using the XML approach that I have set out here and it will be an empirical outcome of this approach if it turns out that the mental content of a stone is always devoid

of phenomenal states. This *empirical* outcome must be distinguished from the *a priori* question about whether certain types of non-human system are capable of supporting conscious states, since it is possible that the XML approach will make predictions about consciousness in systems that we consider highly unlikely to be capable of consciousness – the economy of Bolivia, for example. This *a priori* question is tackled by the ordinal probability scale, set out in Gamez (2005), which evaluates the likelihood that a machine can support phenomenal states by systematically comparing its architecture with the human brain.

It has been suggested that this XML approach to synthetic phenomenology ignores behavioural criteria of consciousness, such as reports that a system might make about its mental contents. If this was thought to be important, then it would be easy to include the actuator outputs in the mental content XML file, so that the external behaviour of the system could be included in the analysis of its consciousness on a moment to moment basis. However, the problem with behavioural criteria for consciousness is that apparently conscious behaviour can be generated by systems that we are reluctant to attribute consciousness to (such as the population of China communicating with radios and satellites), which is why an internal architecture approach has been favoured here.

As this methodology develops there are likely to be a large number of ambiguities about what constitutes an element, how to handle overlapping elements, how to define active connections, the best way to analyse mental content for phenomenal states, and so on. Although these might initially appear to be weaknesses of the method, they are actually strengths because they indicate that synthetic phenomenology has the potential to become a paradigmatic science that can move forward by asking questions and resolving ambiguities such as these. At the moment synthetic phenomenology is so unclear that even its lack of clarity is unclear to it and tightening up the methodology through XML representations would make it capable of asking and answering precise questions and enable it to move forward in a sustainable manner. Different ways of resolving the ambiguities will make testable predictions about the phenomenal states of a machine or organism and as neural scanning becomes better we will actually be able to test these predictions on human beings and eliminate inaccurate methods. In the early stages it is likely that different theories will generate conflicting XML representations. However, this will at least make differences explicit; whereas at present our descriptions of inner states are so woolly and imprecise that disagreement or comparison between methods is rarely an issue.

For reasons of brevity and clarity this paper has set aside questions about the temporal nature of phenomenal experience. One solution to this would be to break the stimuli up into sequences of frames and separate the test suite and mental content into a list of associated XML files. Another temporal problem is that active elements may change as they develop and so it may not be possible to generate a single test suite that is valid for all time. This type of system will have to be retested at regular intervals or have its adaptivity frozen whilst the description of its synthetic phenomenology is taking place.

6. Previous Work

The approach that I have set out in this paper is closest to some of the techniques for representing non-conceptual content discussed by Chrisley (1995). These include content realization, in which content is referred to by listing "perceptual, computational, and/or robotic states and/or abilities that realize the possession of that content" (Chrisley, 1995: 156), ability instantiation, which involves the creation or demonstration of a system that instantiates the abilities involved in entertaining the concept, and two forms of self instantiation, in which the content is referred to by pointing to states of oneself or the environment that are linked to the presence of the content in oneself. Whilst all of these techniques are promising ways of referring to non-conceptual content, it will be very difficult to apply them in practice without a precise way of representing and organizing the computational, and/or robotic states and/or abilities. It is here that XML would be a useful tool since it could represent the structure of the systems that are being analysed along with their inner states when they are exposed to stimuli from the environment. Within the precise framework offered by XML the specification of non-conceptual mental content using Chrisley's techniques would be made considerably easier.

Other related work includes the description of the synthetic phenomenology of Khepera robots by Holland and Goodman (2003) and Stenning, et. al. (2005). In these experiments the internal model of the Khepera is held in a neural network, which stores a linked series of concepts combining sensory and motor information. The synthetic phenomenology of the Khepera is carried out by plotting a graphical representation of the sequence of sensations and movements stored in the neural network. The problem with this approach is that the Khepera is likely to have no notion of colour and a very limited idea about space and so this graphical representation is unlikely to be anything like the Khepera's actual 'mental' content. Another problem is that the graphical representation contains the complete in-

ternal model, whereas only a small part of this would be active at any point in time. It is also hard to see how this representation of an internal model could be systematically analysed for signs of consciousness. The XML approach could help with these problems since it offers a highly structured way of representing the current mental content of the Khepera, which could be compared with other robots and systematically analysed for signs of consciousness .

7. Conclusion

This paper has briefly outlined an XML approach to synthetic phenomenology in which XML plays a key role in the description of the conscious and unconscious states of the machine. This has many advantages and could help to circumvent many of the problems associated with the representation of non-conceptual mental content. By describing mental content this concretely it also forces us to face challenging theoretical and methodological questions, which will eventually open up the possibility of a systematic science of synthetic phenomenology that can pose and answer precise questions about the phenomenology of artificial systems.

The XML extracts included in this paper are intended as simple examples to illustrate the main ideas and a great deal more work is needed to turn these starting points into a usable method. Some of this development will be done as part of the work on the CRONOS robot at Essex and Bristol. In the longer term it may be possible to develop a single XML standard for both synthetic and neuro- phenomenology, which would facilitate precise comparisons between humans, animals and machines and enable us to automatically examine all three for signs of consciousness.

Acknowledgments

Many thanks to Owen Holland for feedback about this paper. Thank you also to the EPSRC for funding this work (grant number GR/S47946/01).

References

Igor Aleksander and Barry Dunmall. Axioms and Tests for the Presence of Minimal Consciousness in Agents. In Owen Holland (ed.), *Machine Consciousness*, Exeter: Imprint Academic, 2003.

R. J. Chrisley. Taking Embodiment Seriously: Non-conceptual Content and Robotics. In Kenneth M. Ford, Clark Glymour, & Patrick J. Hayes (eds), *Android Epistemology*, Menlo Park/ Cambridge/ London: AAAI Press/ The MIT Press , 1995.

Antonio, R. Damasio. *The Feeling of What Happens*. New York, San Diego and London: Harcourt Brace & Company, 1999.

S. Franklin, A. Kelemen and L. McCauley. IDA: a cognitive agent architecture. *IEEE International Conference on Systems, Man, and Cybernetics*, 3: 2646-2651, 1998.

David Gamez. An Ordinal Probability Scale for Synthetic Phenomenology. In R. Chrisley, R. Clowes and S. Torrance (eds.), *Proceedings of the AISB05 Symposium on Next Generation approaches to Machine Consciousness: Imagination, Development, Intersubjectivity, and Embodiment* 85-94, 2005.

Owen Holland and Rod Goodman. Robots With Internal Models. In Owen Holland (ed.), *Machine Consciousness*, Exeter: Imprint Academic, 2003.

Y. Kamitani and F. Tong. Decoding the visual and subjective contents of the human brain. *Nature Neuroscience* 8:(5) 679-685, 2005.

Thomas Metzinger. *Being No One*. Cambridge Massachusetts: The MIT Press, 2003.

Thomas Nagel. What is it like to be a bat? *The Philosophical Review* 83: 435-456, 1974.

J. Searle. Minds, Brains and Programs. *Behavioral and Brain Sciences*, 3: 417-57, 1980.

J. Stening, H. Jacobsson and T. Ziemke. Imagination and Abstraction of Sensorimotor Flow: Towards a Robot Model. In R. Chrisley, R. Clowes and S. Torrance (eds.): *Proceedings of the AISB05 Symposium on Next Generation approaches to Machine Consciousness: Imagination, Development, Intersubjectivity, and Embodiment* 50-58, 2005.

G. Tononi. An Information Integration Theory of Consciousness. *BMC Neuroscience* 5:42, 2004.

William Wordsworth. I Wandered Lonely as a Cloud. In Stephen Gill (ed.), *Selected Poems*, London: Penguin, 2004.

The Embodied Machine: Autonomy, Imagination and Artificial Agents

Nivedita Gangopadhyay[*†]
[*†]Institut Jean Nicod
1bis, avenue de Lowendal, 75007, Paris, France
Nivedita.Gangopadhyay@ehess.fr

Abstract

The embodied and enactive approach to consciousness emphasises the role of the physical embodiment of naturally intelligent agents as crucial for a study of consciousness and the importance attributed to the body also tends to be carried over to the material out of which the body is created viz. "living" matter. This seems to put into doubt the relevance of the embodied and enactive approach to the field of machine consciousness. However, I shall argue that consciousness as manifested in embodied intelligent systems, natural or artificial, that enact their world of experience by interacting with the environment necessarily needs to be understood in the light of freedom/autonomy and imagination, and the application of the principles of embodiment and enaction in the light of these notions in the field of robotics and AI can be a big step towards creating conscious artificial agents.

1 Introduction

The attempt to understand cognition and consciousness by recognising the fact that they necessarily involve an embodied agent who enacts her world of experience by real-time interaction with a real-world situation has been propounded of late in an ever-increasing volume of literature in the field of consciousness studies. The emphasis laid on the notions of embodiment of the cognitive agent and her interaction with the environment as crucial elements even for a scientific study of consciousness, has come a long way from a philosophical idea first presented in continental philosophy in the works of philosophers like Husserl (Husserl, 1931, 1960) and Merleau-Ponty (Merleau-Ponty, 1962, 1963, 1964). When the ideas of these philosophers were being introduced to the philosophical analyses of consciousness, mainstream cognitive science in general had remained unaffected by the implications of such a phenomenological approach. The applications of the emerging principles of cognitive science in the field of robotics and artificial intelligence dominated by the information-processing view of cognition had largely ignored the possible implementations and crucial insights that a primary emphasis on the notions of embodiment and enaction could have provided. The necessity to stress the agent's particular psycho-physical apparatus and the real-time interactions of the agent with the real-world environment for an adequate study of consciousness began to be realised for the first time in robotics and AI in the 1980s in the work of Brooks (Brooks, 1986, 1991, 1993, 1994). The development of what has come to be known as the autonomous-agent theory in AI emphasises that as a first step for artificial agents to exhibit mental characteristics typically associated with conscious agents, they must be created in such a way that they are capable of moving about, surviving and performing specific goal-directed actions in real time in a complex real-world environment.

The embodied and enactive theories, as are gradually gaining ground in mainstream cognitive science, emphasise the kind of body the agent possesses as one of the first crucial elements to be considered by a satisfactory theory of consciousness and the importance attributed to the kind of body also tends to be carried over to the material out of which the body is created viz. biological matter. The environment of the agent, both natural and socio-cultural, also constitutes an indispensable dimension of embodied and enactive approaches to the study of consciousness. Indeed, a survey of the literature reveals that "embodiment" can be understood in a variety of ways and following Ziemke we can enumerate the different notions of embodiment as follows: 1) structural coupling between agent and environment, 2) historical embodiment resulting from a history of agent-environment interaction, 3) physical embodiment, 4) 'organismoid' embodiment i.e. organism like body and 5) organismic embodiment

of autopoietic living systems (Ziemke, 2001). However, for the present I shall consider the notion of embodiment in a rudimentary sense of physical embodiment i.e. having a particular kind of body as a fundamental determinant of consciousness. Having a kind of body means instantiating a specific biological model and a major contention of the embodied and enactive approaches is that the biological model of the agent crucially determines the characteristics associated with consciousness exhibited by the agent. While this claim is indeed justifiable in view of the fact that the interaction of the organism with the environment that generates its world of experience is importantly determined by the physical embodiment of the organism, the implication that all manifestations of consciousness could thereby also be limited to the embodiments of naturally intelligent systems as created out of "living-matter" is less evident. Due to this insistence on the physical embodiment of the agent and the underlying importance of the biological matter, theories of embodied and enactive cognition seem to possess an unavoidable "biological flavour". Then does engineering artificial intelligence according to the basic principles of embodiment mean engineering living matter? This may seem to stand in the way of applying the principles of the embodied and enactive approach in the field of machine consciousness. In this article I would like to address the question: *How far can the embodied and enactive approach to consciousness with its emphasis on the kind of body possessed by the naturally intelligent agents at all help in understanding consciousness through the creation of intelligent machines based on the principles of embodiment and enaction?* I shall maintain that although the notions of embodiment and enaction as used for understanding naturally intelligent systems importantly involve the material out of which the agent's body is created as a determinant of embodiment, these notions can be applied in the field of robotics and AI too to create artificial agents that we could at least hesitate to call "machines" even if the material out of which they are created is not "living" stuff. I shall argue that consciousness as manifested in embodied intelligent systems, natural or artificial, that enact their world of experience by interacting with the environment necessarily needs to be understood in terms of freedom/autonomy and imagination, and the application of the principles of embodiment and enaction in the light of these notions in the field of robotics and AI can be a big step towards creating conscious artificial agents.

1.1 Natural Embodiment

The embodied and enactive theories have at times sought to differentiate between naturally intelligent systems and mechanical systems by drawing upon the material out of which each is created and the set of structural properties of the resulting systems as the criteria. One such effort is made by Maturana and Varela (Maturana and Varela, 1980, 1987) who distinguish between autopoietic systems and allopoietic systems. Biological systems, made out of living matter and exhibiting natural intelligence, are essentially characterised by their adaptability to their environment at the cellular as well as at the behavioural levels. Such systems are termed autopoietic as they are self-creating and self-maintaining systems, and hence are completely autonomous. On the other hand, mechanical systems made out of non-living matter are capable of adapting only at the behavioural level and are called allopoietic systems i.e. systems whose components are produced by other processes that are independent of the organization of the machines. Hence how can artificial agents created out of non-living matter help us understand consciousness? Here one may adopt a stance of mysterianism and claim following Prinz that "…progress in the science of consciousness may offer little help to those who want to engineer consciousness" (Prinz, 2003) because it is impossible to determine with certainty that biological matter does not contain properties essential for consciousness. Then are our efforts to employ the principles of embodied cognition to the study of robotics futile unless we make machines out of living matter? Prinz advises engineers that they should not "…fool themselves into thinking that they can definitely create conscious machines" (Prinz, 2003) and the emphasis laid by the embodied cognition approach upon the body and of what it is made may seem to lend support to Prinz's advice to engineers. Thus of what use are the notions of embodiment and enaction in the study of robotics and AI?

Given our present state of knowledge about biological matter we cannot but maintain for the time being that it is in fact impossible for us to determine with complete certainty that organic matter does not contain properties essential for consciousness. However, this does not make the notions of embodiment and enaction a redundancy for robotics. Instead of considering the problem of consciousness in its totality, in all its aspect, let us begin by picking out a feature that can be said to be invariably associated with manifestations of consciousness in agents with a biological embodiment. Naturally embodied agents constantly strive to attain to higher degrees of freedom by actively resisting and defying the various forces acting against them that try to break up the unity of the system and by such efforts they assert their existence. The more they are able to resist the counteracting forces threatening to destroy the unity of the system, the more they appear to be complex

from the point of view of consciousness. Thus the most rudimentary life-form embodied in the simplest biological embodiments is the least able to actively preserve its unity in the face of counteracting forces and possesses the least freedom from this point of view and is ascribed the least traces of consciousness. As we go higher up the evolutionary chain we find more and more complex life-forms with more and more complex embodiments with greater and greater degrees of freedom exhibited by actively resisting counteracting forces till we reach the human level to which we ascribe the highest intelligence and consciousness exhibited so far in the story of evolution. Moreover, over and above adverse natural forces biological systems also deal with highly complex socio-cultural forces even at a low level of the evolutionary ladder, e.g. complex social structure of termite or ant colonies, and they seek to maintain their individual existences in this social maze by variously manipulating the forces at work there and trying to preserve their identity as individuals i.e. the unity of their individual systems. When they try to preserve and assert the identity of a group they do so as they identify the unity of their systems with that of the group. The more complex the forces that act against the system and the more the system tries to exert its freedom in the form of preserving its unity by actively counteracting the forces, the more it seems to manifest intelligence. From an evolutionary perspective it can be said that the forces acting against the system become more and more complex as we go higher up the ladder and the ways of counteracting those forces also become more and more sophisticated and complex leading to the expressions of greater freedom and accordingly greater degrees of consciousness.

As a primary strategy of counteracting the forces acting against the system naturally intelligent agents resort to interacting with their environment in creative ways i.e. *they can represent to themselves or enact possible states of affairs by interacting with the present state of affairs*. The ability to represent possible states of affairs varies in complexity in accordance with the embodiment of the system and the complexity of the forces which the system encounters. The human form of embodiment is the one most capable among all biological embodiments to actively maintain the unity of its system in the face of highly complex counteracting forces, both natural and socio-cultural; and the capacity of humans to enact possible worlds, as a strategy adopted for counteracting adverse forces, is remarkable among biological embodiments from the point of view of its complexity. This capacity as present in human embodiment is what we generally call *imagination i.e. the enaction of possible worlds* although other naturally intelligent systems too can represent to themselves possible states of affairs in various degrees of complexity by interacting with the immediate environment (the present state of affairs) and hence can also be called "imaginative". By this remarkable capacity/strategy of counteracting disintegrative forces naturally intelligent systems exert their greatest freedom.

Moreover, in case of naturally intelligent systems it can be observed that with the increasing complexity of embodiment the interaction of the organism with the environment gradually shifts from one of adaptation to one of gradual control leading to greater expressions of freedom and intelligence. The simplest life-forms adapt themselves as best as they can to the conditions of the environment and accordingly the manifestation of intelligence in them is far less complex than that of the higher ones. The strategies of interacting with the environment tend to become more of control and less of adaptation in more and more complex embodiments till we reach the human level that is crucially characterised by its capacity to enact possible worlds by interacting with the environment primarily in the form of *control strategies*. In case of natural forces humans do not submit themselves to the mercy of Nature and try to adapt as best as they can to the situations Nature throws them into. Humans exert their freedom against natural forces by trying to master natural laws and make them work for their best advantage. Even for socio-cultural forces humans demonstrate the tendency to assert their control over the environment and this tendency has been manifest throughout the history of human civilization. By interacting with the environment (the current state of affairs) in accordance with their embodiments humans enact possible states of affairs (imagination) that can be far removed from and greatly more complex than the present state of affairs. Hence in humans, manifestations of intelligence are not simply matters of adaptation; intelligence is dominance and control over environment with the aim of manipulating it to the best of their advantage i.e. making conditions most favourable for the maintenance of the unity of the system and thereby exerting their freedom. For other biologically embodied systems too intelligence is crucially determined by the ability of the system to actively preserve its unity in the face of counteracting forces and by interacting with the environment in creative ways to represent to itself possible states of affairs and thereby asserting its freedom. No matter how simple or how complex the embodiment, the basic principle of intelligence and manifestation of consciousness indeed seems to be this and the human embodiment by virtue of the greatest ability exhibited so far in evolution to preserve the unity of the system and enact possible worlds by interacting with the present environment, enjoys the greatest degree of freedom in the chain

of evolution and exhibits the most complicated manifestations of intelligence.

Furthermore, despite varying in degrees of complexity and freedom, the naturally intelligent systems are all characterised by the ability to represent to themselves the goals of their actions. Natural systems have *dynamic needs* and so they interact with the environment in various ways and enact their world of experience. Exploration of the environment by natural agents is importantly guided by *curiosity*, i.e. the *need* to explore more. This is all the more true for human agents whose insatiable curiosity has been at the root of all discoveries and inventions. The interaction with the environment by human agents is characterised by this need to explore more and more, and the lack of complete satisfaction with the present state of affairs. The need to explore more is developed by the system by means of interacting with the environment and by representing to the system goals other than the immediate ones present in the environment i.e. enacting possible states of affairs (imagination). The biological systems express their freedom by not being limited to what is immediately present in the environment. The needs, whether biological or psychological, can be traced back to the desire to assert the existence of the organism and preserve the unity of the system and enact possible worlds by interacting with the present environment. Thus the needs come from the system by interacting with the environment and as long as there is embodiment there are needs.

2 Autonomy, Imagination and Artificial Agents

Consciousness as associated with this idea of freedom expressed by the embodied system through actively resisting disintegrative forces to maintain the unity of the system and enacting possible states of affairs by interacting with the present state of affairs need not be logically restricted to biological systems alone. This idea can be applied to the study of robotics and AI although the material out of which we create artificial agents like robots is not organic matter. The question is one of freedom. Naturally intelligent systems are characterised by various degrees of freedom in that they have capacities to actively preserve the unity of their system against disintegrative forces in various degrees and enact possible states of affairs by interacting with the present state of affairs, and accordingly manifest various degrees of complexity of intelligence. However, while modelling consciousness it is to be noted that biological embodiments, including human embodiment, have been shaped primarily by the environment whereas for artificial agents it is humans who are *exclusively* trying to shape the embodiment. The application of the embodied approach to cognition has to date influenced the shaping of the embodiment of artificial agents in so far as engineers are now trying to derive inspiration from biological models. The creation of robots that simulate the embodiment of simple biological models like insects (Beer and Chiel, 1993) etc. reflect this urge to copy Nature's work. This is certainly a big step towards realising the importance of the embodiment and enaction for consciousness but it is one thing to mimic biological models for embodiment of artificial agents and quite another thing to create artificial agents whose embodiment will be shaped by the environment, which includes humans but *not only* humans, by means of *creative interaction of the system with the environment* and in order to creatively interact with the environment the system must be able to *develop its own dynamic needs*. To quote Ziemke, " …despite all biological inspiration, today's adaptive robots are still radically different from living organisms. In particular despite their capacity for a certain degree of self-organization, today's so-called 'autonomous' agents are actually far from possessing the autonomy, and consequently the embodiment of living organisms." (Ziemke, 2001). Thus the idea of autonomous agents, that is already prevalent in the study of robotics under the influence of the ideas of embodiment and enaction, can be carried to a greater extent to create agents which are autonomous not only in so far as they are capable of acting upon the environment to carry out functions that have already been decided for them such as moving about and avoiding obstacles but to create systems that will develop their own course of actions by interacting with their environment in accordance with *their dynamic* needs.

To further clarify the idea let us consider basic applications of the idea of embodiment in the field of machine consciousness such as Brooks' "mobots"(Brooks, 1986, 1991, 1993, 1994). Brooks lays down four conditions that his artificial creatures should satisfy and one of these is that a creature must do something in the world; it should have a purpose in existing (Brooks, 1991). Thus Herbert, one of Brooks' well-known mobots, was designed to collect empty soft-drink cans left in the MIT lab. Although Herbert was built on the principles of interaction with the environment, it was none-the-less the human factor that exclusively fixed Herbert's reason for existence and limited its activities in important ways and thereby the exhibition of intelligent behaviour on its part. To understand this more clearly let us compare a human agent with Herbert performing the same task i.e. collecting empty soda cans in a lab. The ways of navigating through the real-world environment maybe quite similar for both the agents but the *reasons* for doing so are crucially different in case

of the human agent and Herbert. The human agent can be collecting cans by taking part in an experiment or by being employed by the lab or because she cannot tolerate a messy littered lab or simply because she likes to collect cans. However, in all these cases she knows that collecting cans is not the reason for her existence; she can stop collecting them (at least in her mind) if she wants and this representation of a possible state of affairs is an important component in her performance of the task. Even if she has been employed by the lab to collect cans she can conceive of possible worlds where she is not conditioned to collect cans. If she is bored or tired with the task or if she simply thinks it has been enough for her she can just quit. That is to say that by means of interaction with this environment i.e. the lab (the present state of affairs) she can enact a possible state of affairs. The autonomy expressed by the human agent in the task is that she is free (at least in her cognitive world) *to make a choice*; to collect cans or not to collect cans? This autonomy importantly shapes the way the human agent interacts with the environment even for simple tasks such as can-collection. Enaction of possible worlds is significantly determined by the interaction with the present state of affairs or the immediate environment. How can this autonomy be brought into artificial agents? To answer this question I shall make use of the notion of *potential enactive state* in the modelling of consciousness.

In creating an artificial agent in accordance with the principles of embodiment and enaction it is necessary to build in some routines in the form of reaction to the various environmental factors. For example, the subsumption architecture underlying the functioning of Herbert is composed of layers which can be viewed as built-in routines of reactions to environmental factors like halting when an object is sensed right in front and reorienting towards an unobstructed direction. It is crucial for successful elementary navigation through the environment that certain rules of interaction with the environment be present in the robot that guides its behaviour. These can be considered as routines that enable the artificially embodied agent to enact the present state of affairs by interacting with the environment. However, the cues that the robot obtains by interacting with the environment need not all be directed towards solving a specific task either in the form of positive feed-back or negative feed-back. Imagine a device that has multi-sensors simulating the senses of natural agents. The inputs that the robot receives via its interaction with the environment need not all be translated into action. Some inputs will be utilised for immediate action whereas some will not be so utilised. However, the ones that are not so utilised immediately will not be ignored as irrelevant for all times but be preserved in the system as potentially relevant cues for further interacting with the environment. Although the robot can be initially programmed for performing a specific task such as navigating through a real-world environment and avoiding obstacles, the picking up of cues from the environment by interaction should enable the system to develop further goals i.e. further *needs* for interacting with the environment. Interaction with the environment is a crucial factor in the origination of goals for embodied systems and the setting forth of these goals and representing them to the system enables the system to evolve and exhibit more complex intelligent behaviour. A human agent navigating through an environment for initially performing a specific task, e.g. soda-can collecting, picks up a lot of cues from the environment that are not all immediately pertinent to the task at hand but which significantly determine the manifestation of intelligent behaviour on the part of the agent. Suppose while collecting the cans in a lab a human agent hears a strain of music coming from somewhere. The music is rather lilting and the agent feels the need to dance to its tune i.e. move her body to its rhythm. The music does not constitute any part of the pertinent cues for soda-can collecting but it does constitute a dimension of interaction of the agent with the environment and enables the agent to enact a possible state of affairs. If the agent is not restricted by the terms and conditions of employment or experimentation, she may even abandon her task of can collection for some time and just dance a bit or listen more intently to the music, and if she wishes she can give up the activity of collecting soda-cans in favour of enjoying herself. Moreover, she is most likely to become curious about the source of the music too and may leave her immediate environment to trace it, i.e. she will explore more. If her movements are restricted by terms of employment or experimentation, she can none-the-less enact a possible state of affairs in her cognitive world where she is executing her desired behaviour. Thus the human agent exerts her autonomy by preserving the unity of her system in the face of counteracting forces (obligation to collect cans despite the reluctance to do so) and enacting a possible state of affairs (dancing, listening with greater attention to the music, exploring the environment for the source of the music) by interacting with the present state of affairs (collecting cans but there is a nice music coming from somewhere). In fact it is this feature of naturally intelligent systems, especially human ones, that has so far distinguished them from machines or mechanical behaviour as has been so far modelled. Herbert can go on collecting soda-cans indefinitely for it does not develop any further needs by interacting with the environment but a naturally intelligent system will sooner or later call it a day. As Maturana points out, "…as living systems that live humanly we are different from

robots on two fundamental accounts: one, is that robots have been designed de novo, intentionally in congruence with a specified medium that may also have been designed with them, and are not the arising present of an evolutionary history; two, is that we human beings are the arising present of an evolutionary history in which our ancestors and the medium in which they lived have changed together congruently…" (Maturana, 2005)

The idea of potential enactive state is the idea of the system's ability to represent to itself goals other than the ones immediately present in the environment by interacting with the environment. According to its embodiment the artificial system should be able to develop its needs with the aim of exerting greater control over the environment or for moving from adapting to the environment to gradually controlling it to the best of its advantage. The human control in the creation of truly autonomous artificial agents is at least for the time-being importantly present at the levels of design and programming. At the level of designing the initial embodiment the human designer needs to make a choice of environment for the artefact and equip the system with means of interacting with the environment. For this inspiration can be derived from biological systems and their sensory modalities because these systems by interacting with the environment for a long time have developed the most practical designs. The choice of the number of sensory modalities with which the artificial agent is to be equipped and the kind of movement that the system will execute are the concerns of the human designer, although the movement may be importantly determined by the physical features of the environment chosen. The complexities of the sensory modalities and the movement will significantly determine the complexity of the initial embodiment. However, a truly autonomous system, albeit constituted of non-living matter, should also be able *to evolve* its embodiment in accordance with the interaction with the environment. This is not merely a question of adapting to the environment at the level of behaviour as Maturana and Varela state for allopoietic systems (Maturana and Varela, 1980, 1987). It is developing or *evolving* the embodiment in accordance with environmental interaction with the aim of expressing greater freedom of the system and exerting greater control over the environment. As an example we can consider Herbert once again in an imaginative thought experiment that could roughly capture the implementation of the idea of an essential manifestation of consciousness as the ability of the system to actively preserve its unity in the face of counteracting forces and exert its freedom by interacting with the environment in creative ways to represent to itself or enact possible states of affairs. Herbert is designed to collect empty soda-cans in a lab and explores the environment randomly, *not ignoring* other objects but exploring them too by means of tactile and visual modalities. This can indeed be possible as Brooks claims that the artificial creature should be able to maintain multiple goals (Brooks, 1991). However, with the kind of physical embodiment (design) Herbert has been initially given it can pick up only empty soda-cans. Nevertheless Herbert can send a "distress" signal when it "senses" that the system is missing something, i.e. the system has developed a need by interacting with the environment and this need needs to be fulfilled for the system to exert greater freedom and control over the environment. Also suppose Herbert is equipped with temperature sensors that enable it to estimate how much energy is being spent. Now Herbert is moving through the lab, exploring it and picking up empty cans when it finds one. In the course of its random exploration suppose Herbert comes across a piece of crumpled paper lying on the floor. By exploring that crumpled ball of paper Herbert finds out that the weight of that object is less than the objects that it has been picking up. Hence interacting with that object, rather than with the empty soda-cans, means less spending of energy by the system which means more ability to explore the environment. But with the current design Herbert cannot pick up the ball of paper. It sends out a "distress" signal to indicate that the system needs something. This indicates that the agent is developing its own needs and enacting to itself a possible state of affairs by interacting with the present state of affairs for preserving the unity of the system. However, enacting a possible state of affairs or the representation of a potential enactive state should not come to an end with only a single instance. By encountering the ball of paper the system should be able to represent to itself the general possibility that there are objects in this environment that put less demand on its energy and consequently interacting with them means more ability to interact with the environment and preserve the unity of the system. *The representation of this possibility should never be exhausted*. For naturally intelligent systems as long as there is embodiment there are needs for which the system manifests intelligence and artificial embodied systems must also follow in their steps. The potential enactive state in an artificial agent represents a state which is *actually never reached* by the system. It is a state which the system is always *trying* to reach and with this aim is interacting with the environment. The system *must never reach equilibrium* i.e. the state where the system "feels" no more need to interact with the environment or develops no further need to interact with the environment. Real-world environments are essentially dynamic set-ups and hence complete control of the counteracting environmental forces is a dream for both naturally intelligent systems and

artificial ones. Yet it is the incessant pursuing of this practically unattainable state of autonomy that leads intelligent systems to manifest more and more complex intelligence.

To sum up, the necessity of applying the principles of embodiment and enaction in the field of robotics and AI is becoming increasingly clear for creating artificial agents that can exhibit mental characteristics typically associated with consciousness, and the notions of autonomy (exerting greater and greater degrees of freedom by the ability to preserve the unity of the system in the face of counteracting forces) and imagination (enacting possible states of affairs by interacting with the present state of affairs or the immediate environment) are crucial for creating embodied artificial agents capable of enacting cognitive states. The goal to be attained is complete autonomy obtained by constant enactment of possible states of affairs through interaction with the present state of affairs, and the ever present vision of this impossible goal necessarily permeates all intelligence and evolution, from the simplest to the most complex till date. I have argued in this paper that such a manifestation of consciousness need not be restricted to naturally intelligent systems alone and can be simulated in artificial agents. Whether or not it is impossible to satisfactorily determine the issue of biological matter possessing properties essential for consciousness, or whether or not autopoietic systems are essentially different from allopoietic ones by virtue of their adaptability, are questions that tend to restrict the notions of embodiment and enaction to a level of explanation that may render these notions inapplicable *in principle* in the domain of robotics and AI because of their explicit or implicit harping on biological matter. This is not to imply that these issues can be brushed aside in studies of embodiment and enaction. It may indeed be possible that biological matter is really some thing quite special for manifestations of consciousness, and the latter is inseparably linked to the former *and only* to the former. But nevertheless it may also be possible, by understanding embodiment and enaction as pertaining to consciousness in the light of freedom and imagination, to create artificial agents that we would hesitate to call "machines" any more in the sense that they perform only dumb repetitive behaviour in order to serve *our* purposes and *our* whims. Thus the challenge that faces us for this new vision of artificial agents is not how far *could* we go in creating conscious machines but rather: How far would we *dare* to go?

References

Anderson, M.L. Embodied cognition: a field guide. *Artificial Intelligence*. 149: 91-130, 2003.

Beer, R., and Chiel, H. Simulations of cockroach locomotion and escape. *Biological Neural Networks in Invertebrate Neuroethology and Robotics*. ed. R. Beer et al. Academic Press, 1993.

Brooks, R. A robust layered control system for a mobile robot *IEEE Journal of Robotics and Automation RA-2,* 1, April: 14-23, 1986.

Brooks, R. Intelligence without reason. *Proceedings of the 12th International Joint Conference on Artificial Intelligence*. Morgan Kauffman, 1991.

Brooks, R. A robot that walks: Emergent behaviors from a carefully evolved network. *Biological Neural Networks in Invertebrate Neuroethology and Robotics*. ed. R. Beer et al. Academic Press, 1993.

Brooks, R. Coherent behavior from many adaptive processes. *From Animals to Animats 3*. ed. D. Cliff et al. MIT Press, 1994.

Brooks, R., and Maes, P. eds. *Artificial Life 4*. MIT Press, 1994.

Brooks, R., and Stein, L. Building Brains for Bodies. Memo 1439, Artificial Intelligence Lab, Massachusetts Institute of Technology, 1993.

Chrisley, R., and Ziemke. T. Embodiment. *Encyclopedia of Cognitive Science*. Macmillan, 2002.

Clark, A. Being there; why implementation matters to cognitive science. *AI Review 1*, 4: 231-244, 1987.

Clark, A., and Chalmers, D. The Extended Mind. Philosophy –Neuroscience – Psychology Research Report, Washington University, St. Louis, 1995.

Clark, A. *Being There: Putting Brain, Body, and World Together Again*. MIT Press, 1997.

Husserl, E. *Ideas: General Introduction to a Pure Phenomenology*. Trans. W.R. Boyce Gibson. Allen and Unwin, 1931.

Husserl, E. *Cartesian Meditations: An Introduction to Phenomenology*. Trans. Dorian Cairns. Martinus Nijhoff, 1960.

Lipson, H., and Pollack, J.B. Automatic design and manufacture of robotic lifeforms. Nature, 406: 974-978, 2000.

Maturana, H.R. The origin and conservation of self-consciousness. Kybernetes, 34(1/2): 54-58, 2005.

Maturana, H.R., and Varela, F.J. Autopoiesis and cognition- The realization of the living. D. Reidel Publishing, 1980.

Maturana, H.R., and Varela. F.J., *The Tree of Knowledge: The Biological Roots of Human Understanding.* New Science Library, 1987.

Merleau-Ponty, M. *Phenomenology of Perception*. Trans. Colin Smith. Routledge and Kegan Paul, 1962.

Merleau-Ponty, M. *The Structure of Behavior*. Trans. Alden Fisher. Beacon Press, 1963.

Merleau-Ponty, M. Eye and mind. *The Primacy of Perception and Other Essays*. ed. James M. Edie. Northwestern University Press, 1964.

Pfeifer, R., and Scheier, C. *Understanding Intelligence*. MIT Press, 1999.

Prinz, J. Level-Headed Mysterianism and Artificial Experience. *Journal of Consciousness Studies*. 10: 111-132, 2003.

Sharkey, N.E., and Ziemke, T. Life, Mind and Robots- The Ins and Outs of Embodied Cognition. *Hybrid Neural Systems*. eds. S. Wermeter and R.Sun. Springer Verlag, 2000.

Thompson, E. Life and mind: From autopoiesis to neurophenomenology- a tribute to Francisco Varela. *Phenomenology and the Cognitive Sciences*. 3: 381-398, 2004.

Varela, F.J., Thompson, E., Rosch, E. *The Embodied Mind*. MIT Press, 1993.

Weber, A., and Varela, F.J. Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the Cognitive Sciences*. 1(2): 97-125, 2002.

Ziemke, T. Are Robots Embodied? Paper presented at the first International Workshop on Epigenetic Robotics: Modeling Cognitive Developments in Robotic Systems, Lund, Sweden, 2001.

Ziemke, T. What's that thing called embodiment? *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*. Lawrence Erlbaum, 2003.

Towards Streams of Consciousness; Implementing Inner Speech

Pentti O A Haikonen
Nokia Research Center
P.O. Box 407, FI-00045 NOKIA GROUP
pentti.haikonen@nokia.com

Abstract

Inner speech is an aspect of human cognition that has been largely neglected by traditional artificial intelligence research. It is argued here that inner speech is an important contributor to cognition and consciousness and therefore also conscious machines should incorporate it. The realization of inner speech in machines involves also notoriously difficult linguistic issues, like sentence understanding. Here an approach to language processing by associative neural networks is proposed as the solution. This method works without explicit parsing or grammatical rules. The cognitive effects of inner speech arise from its content; inner speech is about something and that content affects the operation and behavior of the cognitive system. Consciousness involves the awareness of the mental content; inner speech is seen here as one tool for introspection that facilitates this awareness. In inner speech we may comment ourselves in a way that we have learned from others. This self-appraisal is seen as a process that leads to enhanced social self-awareness and self-image.

1 Introduction

In humans the inner or silent speech is the "little voice inside the head" that commences when we awake and ceases when we fall asleep. Inner speech seems to be present also in dreams at least to some degree. Inner speech is persistent; it is difficult to suppress it for any extended moment while awake. In folk psychology inner speech is often equated to thinking and is understood as a main difference between man, animals and machines. Introspection may mislead us, but inner speech would seem to be one consciousness-related phenomenon that we can be rather sure of. Inner speech is a tool of introspection; via the flow of inner speech we are able to report to ourselves what we think. When we fall asleep the flow of inner speech stops and our consciousness is very much diminished. Nevertheless, it is obviously possible to be conscious to at least some degree without language, solely by the flow of sensory percepts, inner imagery, feelings, actions, needs and the like.

Inner speech has been traditionally ignored by AI researchers while within cognitive psychology and neuroscience its potential as a key component of consciousness has been seen (for instance Morin & Everett 1990, Morin 1993, 1999, 2003, 2005, Siegrist 1995, Schneider 2002). Lately however, also some machine cognition researchers have recognized the importance of inner speech. (Clowes & Morse 2005, Haikonen 1998, 1999, 2000, 2003, 2005a, 2005b, Steels 2003a, 2003b). Also Duch (2005) has proposed a conscious architecture with a flow of "mind objects"; words and images.

In the context of machine consciousness inner speech has a rather crucial position as its explanation and artificial generation involves almost every other issue of cognition; perception, recognition, the grounding of meaning, situational inner models, the temporal handling of information; what the situation is now, what it was before, what has changed, etc. It seems obvious that a machine cannot have meaningful inner speech if it does not understand the world, as this would be a prerequisite for the understanding of language. The solving of the issues of inner speech would involve the solving of most of the practical problems of conscious machines.

The author sees the engineering challenges of inner speech as two-fold. The first issue relates to the enabling neural mechanisms and supporting circuitry for inner speech. The second issue relates to the contents of inner speech, how its meaning is grounded, how it arises and what are its effects on cognition and consciousness, especially self-awareness and self-image. In the following the neural and linguistic prerequisites are treated first and the consciousness-related issues next.

2 Mechanisms for Inner Speech

Natural language understanding is a hard problem that has not yet been solved satisfactorily and definitely not in any elegant way. Yet this is the exact problem that must be solved if meaningful inner speech is to be created in a machine. The author's "multimodal model of language" (Haikonen 2003) is one attempt towards natural use and understanding of language in a machine. Here an experiment relating to the implementation of this approach with associative neural networks is described.

Spoken words are temporal sound patterns consisting of sequences of phonemes. The detection of words calls for the ability to capture and analyze sound patterns and transform the serial phoneme sequence into a parallel representation. Thereafter there are two possibilities for the word representation, namely the distributed representation and the single signal (grandmother) representation. In the distributed representation there can be one or more signals per phoneme or syllable, thus each word will be represented by a signal vector. In the single signal representation each word is represented by one signal only. The distributed representation method is more flexible and allows the use of inflection while the single signal method is easier to use in simple simulations.

The author has used an associative neuron group (Haikonen 1999) as the basic processing unit for the distributed and single signal representations. The operation of the associative neuron group is explained here in simplified (but working) terms, which can be readily implemented with a computer program. The associative neuron group can be seen as a group of neurons that share common associative (synaptic) input signals. Thus their synapses form a kind of a matrix, figure 1.

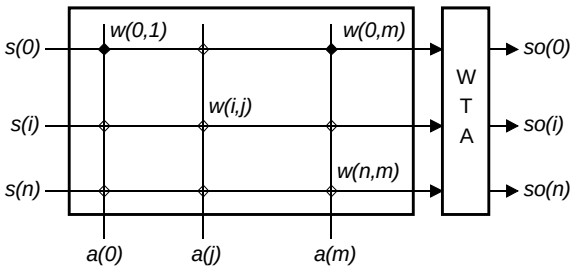


Figure 1. The associative neuron group

Each cross-point can be understood as one synapse and each horizontal line can be understood as one neuron with *m* synapses and one output signal *so(i)*. The purpose of each synapse is to associate the crossing signals *s(i)* and *a(j)* with each other. This is done via the synaptic weight *w(i,j)*. The synaptic weight value *w(i,j)* = 0 means that the signals *s(i)* and *a(j)* are not associated with each other, while the value *w(i,j)* = 1 means that the signals *s(i)* and *a(j)* are associated with each other.

The associative link between the two signals *s(i)* and *a(j)* is created if they appear simultaneously. The synaptic weight value *w(i,j)* is computed as follows at the moment of association:

(1) $w(i,j) = s(i)*a(j)$

where

s(i) = the input of the associative matrix; zero or one
a(j)= the associative input of the associative neuron group; zero or one.

Initially the synaptic weight value *w(i,j)* has the value of zero. The synaptic weight value *w(i,j)* = 1 gained at any moment of association will remain permanent. In the figures the symbol ♦ at the line crossings is used to indicate a synapse with the weight value 1. (A correlative learning rule for more general learning is given in Haikonen 1999, also described in Haikonen 2003, p. 78)

The associated signal *so(i)* is evoked by the signal *a(j)* according to (2) and (3). First, for each *so(i)* signal an evocation sum *Σ(i)* is computed as follows:

(2) $\Sigma(i) = \Sigma w(i,j)*a(j)$

where

Σ(i) = evocation sum
w(i,j) = synaptic weight value; zero or one.

Next, the output *so(i)* is determined by using an output threshold that equals to the maximum evocation sum. This method is also known as the Winner-Takes-All threshold (WTA).

(3) $so(i) = 0$ *IF* $\Sigma(i) < threshold$
$so(i) = 1$ *IF* $\Sigma(i) \geq threshold$

where

$threshold = max\{ \Sigma(i)\}$

The state of the complete associative neuron group can be computed by the above equations by running the indexes from zero to *n* and *m*.

The associative neuron group can be applied to language processing neural networks as will be shown by the next example.

According to the "multimodal model of language" sensory modalities consist of feedback loops that are associatively connected and in this way try to broadcast their percepts to each other. The percepts are signal vectors where each individual signal represents a detected elementary feature. These elementary features are extracted from sensory infor-

mation via sensor-specific preprocesses. This perception process is also affected by the feedback from the system. (The author proposes that this kind of a system is conscious of an entity, when each sensory modality percept is about the same entity and represent different aspects of that entity, hence broadcasts are globally accepted and the whole system is in a kind of multiple closed-loop state.)

Thus, according to this model the language processing takes place in the auditory sensory modality, but is assisted by all the other modalities as well. The general outline of the auditory sensory modality with connections to elsewhere is depicted in the figure 2.

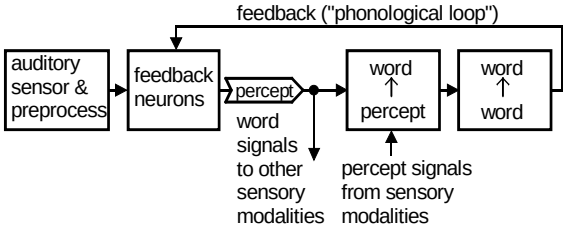


Figure 2. Linguistic modality model, part of the auditory sensory modality

The basic meanings of the words are grounded to sensory percepts like objects, sensations and change. These percepts are associated with words so that each percept may evoke the corresponding word (the percept→ word box in the figure 2). However, *our inner speech is not a list of the names of seen objects*, instead it is more like a running commentary about the perceived situation. Names are not important, possibilities, affordances (Gibson 1966) are. Here also, it should be seen that a perceived entity would evoke many kinds of responses in the other sensory and motor modalities; these would be perceived by those modalities and broadcast to the linguistic modality. Hence the evoked words would be related to the initially perceived object in a more general way. Also, the visual sensory modality is not the only relevant modality here; inner speech may be cued by other sensory modalities as well, like the auditory, touch, temperature, hunger. (Name→ percept association is important whenever a verbal description of a situation is to be transformed into a mental image of the same.) Nevertheless, the percept→ word association is a rather straightforward process and is not elaborated here.

The understanding of a sentence calls for the ability to extract the relationships between the entities that are described by the words in that sentence. There is also a syntactic component; part of the meaning is encoded in the word order and/or in the inflection of the words. This process would be executed in the word→ word association box in the figure 2.

A more complicated associative neural network is required for this word→ word association process. A simple example is presented here in order to illuminate the relevant basic issues and requirements.

In this example each word is represented by one dedicated signal (single signal representation). Distributed representation would also have been possible as was already done by the author (Haikonen 1999).

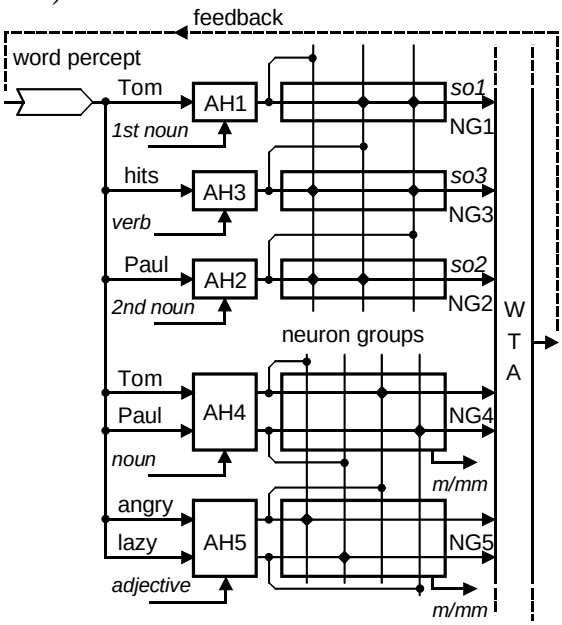


Figure 3. An associative network for linguistic word→ word processing

The associative neural network of the figure 3 is able to process simple sentences like those that describe action between two persons. Furthermore, these persons may or may not be characterized by adjectives. The network consists of neuron groups NG1, NG2, NG3, NG4 and NG5 which all share a common WTA-threshold circuit. The circuits AH1 ... AH5 are "Accept-and-Hold" circuits that recognize individual signals and hold them for a short period. The operation of the "Accept-and-Hold" circuits is grounded; each circuit accepts only its own kind of words, nouns, verbs and adjectives. This is facilitated here by defining each word as a noun, verb or adjective when the vocabulary is taught. In real robotic applications this process would take place during learning of words in natural environment.

Also the position of the word within the sentence matters; first and second nouns are captured separately. The figure 3 is simplified; only those signal lines that are relevant to the specific example are shown.

The subject-object action is captured from the incoming sentence by the neuron groups NG1, NG2

and NG3 and the related Accept-and-Hold circuits AH1, AH2 and AH3. The first noun and the second noun Accept-and-Hold circuits AH1 and AH2 are connected. They accept nouns sequentially; AH1 circuit accepts and captures the first noun and AH2 circuit captures the second noun. AH3 circuit accepts the verb.

When the network learns the information content of a sentence it forms associative connections between the words of the said sentence. The sentence as such is not stored anywhere in the network. As an example the sentence "*Angry Tom hits lazy Paul*" is considered. During learning, that is, when the network receives the sentence, certain associative links are formed. These links operate via synaptic weight values of 1 and are depicted in the figure 3.

The understanding of a sentence involves the ability to answer to questions about the information content of the sentence. When, for instance, the question "*Who hits Paul*" is entered, the word "who" is captured by the AH1 forcing the word "Paul" to be captured by AH2. The verb "hits" is captured by AH3. The associative connections will give the correct response "Tom". The question "*Paul hits whom*" will not evoke incorrect responses, as "Paul" will be captured by AH1 and in that position does not have any associative connections.

The associative neuron groups NG4 and NG5 associate nouns with their adjacent adjectives. Thus "Tom" is associated with the adjective "angry" and "Paul" with "lazy". This is done in the run; as soon as "Tom" is associated with "angry", the Accept-and-Hold circuits AH4 and AH5 must clear and be ready to accept new adjective-noun pairs. After successful association the question "Who is *lazy*" will evoke the response "Paul" and the question "Who is *angry*" will evoke the response "Tom".

Interesting things happen when the question "Is *Tom lazy*" is entered. The word "Tom" will evoke the adjective "angry" at the output of NG5 while the word "lazy" will evoke the word "Paul" at the output of NG4. Both neuron groups NG4 and NG5 have now mismatch condition; the associatively evoked output does not match the input. The generated match/mismatch signals may be associated with the words like "yes" and "no" and thus the system may be made to answer "No" to the question "Is *Tom lazy*" and "Yes" to the question "Is *Tom angry*".

This exercise was executed in the form of a Visual Basic program. The visual interface of this program is shown in the figure 4. This picture presents the situation when the example sentence and some questions about the information content of the sentence have been entered.

This exercise shows that it is possible to use associative neuron groups for language processing, at least for simple sentences. The importance of the grounding of meaning is also demonstrated; while the actual meanings of the words are not grounded here, the categorical meanings are and this grounding is still essential. Further refinement of this approach would involve the basic grounding of meaning for the words and additionally, the use of *inner situational models* that would involve representations in other sensory modalities. These models would allow the grounding and inspection of the relationships between the entities of a given sentence and would also facilitate paraphrasing.

Figure 4. Sentence understanding with the associative neural architecture, a Visual Basic program

A complete cognitive system with the flow of inner speech as sketched by the author (Haikonen 2003) would use these kinds of neural systems as subsystems within the auditory modality. It is worth noticing the simplicity of this approach; this kind of linguistic processing does not utilize explicit sentence parsing or grammatical rules. There is no innate grammar, the "grammar" of a sentence arises from the relationships of the real world.

3 Inner Speech and Consciousness

As good as the devised neural machinery might be it would only be a supporting platform. Any phenomena that relate to consciousness would arise via the content that were carried by the platform in the form of inner speech. After all, our inner speech is about something and, on the other hand, we are not able to perceive the physical neurons or neural processes as such behind our inner speech. The contents of inner speech would allow us to shape our aware-

ness while other, transparent neuron and architecture-related mechanisms, especially feedback and cross-association, would allow the content to be introspected and controlled by itself.

Would it be possible for one system to introspect and control itself? Even simple feedback control systems can do this. However, in this case there may be two major systems interacting with each other, namely the auditory-linguistic modality and the speech-motor modality. These modalities would usually carry the same information albeit in different terms; "heard" word representations and motor command representations for spoken or overt speech. This arrangement would allow the easy inspection of inner speech as a copy of it would be available in the speech-motor modality. Ryding et al (1996) propose: "Audible and silent speech may represent two principally different types of cerebral feedback systems, one for overt sensory–motor activity and one for a pure internal cognitive feedback". There is also some experimental proof that inner "heard" speech and overt speech have separate neural substrates. For instance aphasic patients may complain that the words they speak are not the words they think and intend to say (Huang et al 2001).

The author has proposed that consciousness arises in a multimodal system from associative interconnections between the modalities (Haikonen 2003). According to the "multimodal model of language" inner speech would be one manifestation of these interconnections. If these interconnections break down, then inner speech and consciousness should also vanish. Indeed, Massimi et al. (2005) have noticed that during sleep, when there is no consciousness (or inner speech), neural communication between different parts of the cerebral cortex breaks down while local activities may still exist.

In inner speech we may engage in thoughts about thoughts: "I am thinking now" and in doing so be aware of having thoughts. Obviously this observation of one's own thoughts and the recognition of the ownership of the same would seem to be one manifestation of self-consciousness.

Duval and Wicklund (1972) define self-awareness as the state of being the object of one's own attention. This would include the paying of attention to one's own mental content such as percepts, thoughts, emotions, sensations, etc. Inner speech has been seen as a tool for introspection and one of the most important cognitive processes involved in the acquisition of information about the self and the creation of self-awareness (Morin 1990, 2005, Haikonen 2003, pp. 256 – 260).

With inner speech one can comment one's own situation. Morin (2005) sees this self-talk as a device that can reproduce and extend social mechanisms leading to social self-awareness. (The author has argued elsewhere that basic self-awareness does not require social interaction, see the "hammer test" in Haikonen 2003 p. 161.) As a part of social interactions we are subject to comments about ourselves, the way we are and behave. Self-talk allows us also to internally imitate the act of appraisal; we may echo the patterns of others' comments directly as such or as first-person transformations. We may ask ourselves: "Why did *you* do this stupid thing?" or "Why did *I* do this stupid thing?". Originally it was your mother that posed the question (Haikonen 2003 p. 240). In this way inner speech turns into a tool for self-evaluation, which in turn will affect our self-image; who we are, what we want.

4 Conclusion

Inner speech has been largely neglected by traditional artificial intelligence research perhaps because the algorithmic solving of problems in binary computers does not necessitate it. However, cognitive machines would be different. The emulation of the processes of the human brain and mind would be incomplete without the realization of inner speech.

Unfortunately the realization of inner speech in machines involves also notoriously difficult linguistic issues, like the grounding of meaning and sentence understanding. Here an associative neural approach that works without explicit parsing or grammatical rules is outlined and verified to a limited degree by a computer simulation program.

Inner speech is about something and that content affects the operation and behavior of the cognitive system. Consciousness involves the awareness of the mental content; conscious beings may introspect their mind. Inner speech is seen here as one tool for introspection that facilitates this awareness. Inner speech is not only a running commentary of external events, it involves also self-appraisal. This self-appraisal is seen as a process that leads to enhanced social self-awareness and self-image.

For practical reasons robots should have inner speech, as this would allow communication with natural language in natural way. This would allow easy peeking into the workings of the robot brain; technically it would be very easy to monitor and listen to the inner speech. Also, from a philosophical point of view it would be easier to accept that a robot thinks if it had the flow of inner speech and imagery in a similar way that we have.

Inner speech helps us to make sense of our moment-to-moment existence. A conscious robot should experience its existence in the same way. Therefore we should build machines with inner speech, machines that have streams of consciousness.

References

Clowes, R., Morse, A. F. (2005). *Scaffolding Cognition with Words*. Retrieved on 16. 12. 2005 from http://www.cogs.susx.ac.uk/users/robertc/Papers/ScaffoldingCognitionWithWords.pdf

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

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

Gibson, J.J. (1966). *The Senses Considered as Perceptual Systems.* Boston: Houghton Mifflin.

Haikonen P. O. (1998). Machine Cognition via Associative Neural Networks. *Proceedings of EANN'98* pp. 350 – 357

Haikonen, P. O. (1999). *An Artificial Cognitive Neural System Based on a Novel Neuron Structure and a Reentrant Modular Architecture with Implications to Machine Consciousness*. Dissertation for the degree of Doctor of Technology, Helsinki University of Technology, Applied Electronics Laboratory, Series B: Research Reports B4

Haikonen, P. O. (2000). An Artificial Mind via Cognitive Modular Neural Architecture. *Proceedings of the AISB'00 Symposium on how to design a functioning mind* pp. 85 – 92. UK: University of Birmingham.

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

Haikonen, P. O. (2005a). Artificial Minds and Conscious Machines. In D. N. Davis (Ed.) *Visions of Mind: Architectures for Cognition and Affect* pp. 286 – 306. USA: Idea Group Inc.

Haikonen, P. O. (2005b). You Only Live Twice; Imagination in Conscious Machines. In R. Chrisley, R. W. Clowes & S. Torrance (Eds.), *Proceedings of the AISB05 Symposium on Next Generation approaches to Machine Consciousness: Imagination, Development, Intersubjectivity, and Embodiment.* The Society for the study of Artificial Intelligence and the simulation of behaviour, UK. pp. 19 – 25.

Huang, J., Carr, T. H., Cao, Y. (2001). Comparing Cortical Activations for Silent and Overt Speech Using Event-Related fMRI. In *Human Brain Mapping* 15 (2001), pp. 39 – 53.

Massimi, M. et al. (2005). Breakdown of Cortical Effective Connectivity During Sleep. *Science*, Vol. 309 30 Sept. 2005 pp. 2228-2232

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

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

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

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

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

Ryding, E., BraÅdvik, B., Ingvar, D. H. (1996). Silent Speech Activates Prefrontal Cortical Regions Asymmetrically, as Well as Speech-Related Areas in the Dominant Hemisphere. *Brain and Language* Volume 52, Issue 3 (March 1996), pp. 435-451

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

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

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

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

Could a Robot have a Subjective Point of View?

Dr Julian Kiverstein
Dept of Philosophy, University of Edinburgh
3rd Floor, David Hume Tower,
George Square, Edinburgh, EH8 9JX
j.d.kiverstein@sms.ed.ac.uk

Abstract

An argument for the possibility of conscious robots would have to show that the brain is neither necessary nor sufficient for the possession of consciousness. I will set about giving just such an argument. Proponents of the enactive theory of perception have argued that neural activity doesn't always suffice for the having of conscious experience. They have argued that the body and environment can also play a constitutive role in enabling conscious experience.

In this paper I will argue for the stronger claim that neural activity isn't necessary for conscious experience either. A robot could, I will argue, enjoy phenomenal consciousness. This has been denied by at least one prominent proponent of the enactive theory of perception (see Alva Noë (2005. 230)) who has argued that a robot wouldn't count as a subject of experience. In the absence of a subject of experience, Noë thinks it makes no sense to attribute phenomenal consciousness.

I will argue that on the contrary a robot could be a subject of experience. My argument will proceed in three stages. The first stage argues that a creature is a subject of experience if it has a first-person perspective. I set out some conditions a creature must satisfy if we are to attribute to that creature a first-person perspective. The most important of these conditions is that the representations the creature produces must have reflexive content – they must, in a sense I explain, be representations that refer to themselves.

The second stage of my argument uses a variation on Andy Clark's (2000) argument for the conclusion that access implies qualia. I claim that any representation that has reflexive content will be one to which we have access. Clark has argued that access implies qualia, so it follows that a representation with reflexive content will also have qualia.

The final step in my argument will be to show that action-oriented representations (see Clark 1997 for an account of this type of representation) have reflexive content. Many robots that are capable of producing adaptive behaviour do so by means of action-oriented representations. These robots, I will argue, already meet the conditions for having a first-person perspective. Thus robots with a low-degree of phenomenal consciousness I will claim already exist.

My paper will finish by attempting to motivate this conclusion through a reflection on the connection between consciousness and life. Robots that produce adaptive behaviour are models of life. I will argue that because of the connection between consciousness and life these robots are also models of consciousness.

References

Clark, A. 2000: ‘A case where access implies qualia’ in *Analysis* 60.1: 30-8

Clark, A. 1997: *Being-There: Putting Brain, Body and World Together Again* (Cambridge, MA: MIT Press)

Noë, A, 2005: *Action in Perception* (Cambridge, MA: MIT Press)

Acting and Being Aware

Jacques Penders
Sheffield Hallam University
Sheffield
j.penders@shu.ac.uk

Abstract

One often assumes that we, rational human beings, first think and than act. This paper is an attempt to describe the mental characteristics governing the performance of regular everyday actions; and shows that no mental act has to precede our actions, instead of consciously thinking before we act, we mostly act while simultaneously overseeing our acting. The case of ball juggling is used to underpin the analysis with practical facts.

1 Introduction

In the overview paper of the 2005 Machine Consciousness conference the goals of Machine Consciousness are described as: 1) to create artifacts that have mental characteristics typically associated with consciousness (such as awareness, self-awareness, emotion and affect, experience, phenomenal states, imagination etc.); and 2) to model these aspects of natural systems in embodied models (e.g., robots), (Chrisley et al., 2005).

This definition stipulates that the mental phenomena are to be studied in an embodied creature or model, thus the combination of mental states and physical action is brought into the focus. The theme of the current conference concerns *"models which show the emergence of, or otherwise treat, processes or systems underlying these core themes."* The present paper addresses this theme with an attempt to unravel the mental characteristics, which manifest themselves in regular action oriented contexts. I try to describe the mental stance applied by a human being while performing the standard routines of everyday life. Without being able to systematically order all the mental characteristics mentioned above I will discuss a few assumptions so as to indicate some ordering and suggest a place for the stance I am describing.

An often-encountered assumption – which I believe is generally untrue - is that a certain mental act precedes our bodily actions, or in plain language that we first think and then act. For instance Haggard et al. (2002) write: "*Normal human experience consists of a coherent stream of sensorimotor events, in which we formulate intentions to act and then move our bodies to produce a desired effect*".

However, William James (1890) already clearly noted that the suggested ordering in time does not hold. He described his concept of ideomotor action summarised as: we think the act and it is done. An example of his: "*We think to drink our coffee and we find ourselves already holding the cup in our hands*".

I will argue a step beyond and show that we often act before any conscious thinking has occured. My point is not to substantiate a general moral excuse for cases where we have done things, which we afterwards regret. My point is pragmatic: we cannot act and behave as we do in ordinary life if we first have to think (let alone think over) every action. Being human, we like to think of ourselves as rational beings. In the history of Philosophy Immanuel Kant is probably the clearest exponent of this view. He saw a human being as a logical subject of thought (Stuart, 2005) that is bound to act in the physical world. Kant's work could be seen as a major attempt to reconcile the two while giving primacy to rationality. And indeed on occasions we do first think and then try to act accordingly. However, considering the full extent of all the actions an individual performs in his or her everyday routine, it is clear that our rationality can operate only in the background. The occasions where thinking precedes acting are the exception and not routine practice.

In the morning of a regular day, while deliberating on how to make the best out of the day of today, we routinely drink our coffee and make our way to work, say by car. While driving the car, we suddenly stand on the brakes as we are forced to an emergency stop. Only after having come to a stand-

still we come to think about what we have done the seconds before.

Instead of first thinking and then acting, we only oversee our actions with our conscious and rational minds. I call the mental stance which we take when driving the car and which generally prevails when we act: **being aware without focus.** Interesting about this stance is that actions are selected and performed without them being in the focus of attention, and what is more, as I will show below, when attention gets focussed it often interrupts the actions. I use juggling as an example to investigate the flow of the mental processes.

It is interesting on its own to unravel the mental stance in which action selection takes place, since it might shed light on the complex of mental states and stances by which a human being monitors and controls his of her body and actions. Definitely the human body on its own is a complex system with a complex control structure, the understanding of which could function as a paradigm for robot and machine design.

2 Attention and Acting

In order to explain the stance of being aware without focus, first a few words about the closely related notion of attention. Our mind can be in different modes of activity, with sleeping as the extreme on one end. When awakening from sleep, our mind has to "warm-up" in an arousal phase. Then we become generally aware enough so that we can attend: the mind is aroused and proceeds via getting aware to attention. Further onwards, when there is attention, consciousness and conscious experiences may come in.

Attention is since Broadbent's work often conceived of as a filter for or a gate to consciousness, which blocks, weakens or inhibits incoming messages from the senses. Baars (1997) introduced the metaphor of attention acting as a spotlight in a theatre. When in the spotlight of attention, the mental processing becomes accessible to consciousness. The filter metaphor characterises the operations of attention as reductive while the spotlight metaphor suggests amplification; both nevertheless agree that attention is selective.

Attention also has to do with action. "Awareness [or being aware] implies perception, a purely sensate phase of receptivity. Attention reaches. It is awareness stretched toward something. It has executive, motoric implications. We attend **to** things." (Austin, 1998).

Appropriate applications of motor skills - that is to act appropriately - requires a proper combination of perception, action selection and action execution. The role of attention in relation to perception has been widely studied; however its role in applying motor-skills has not received as much scientific interest. The reason for this might be that motor-control, which is a prerequisite for motor-skilfulness, is very much on and below the edge of what we can consciously experience and control.

The performing arts and sports sciences deal with action and attention. Artists and sporting men and women engage in what is called *deliberate practice* (Rossano, 2003) (Ericsson et al., 1993): the concentrated effort to hone and improve specific (mental and) physical skills. Literature on deliberate practice distinguishes between external attentional focus and internal attentional focus; internal attentional focus means that the performer directs attention to the movements itself, while in external attentional focus, the attention goes to the effects the movements have on the environment (Wulf and Prinz, 2001). In both attitudes attention plays a prominent role, and generally external attentional focus is more proficient.

The influence of internal attentional focus may be observed in for instance dancing or martial arts classes. In a class of beginners, the students might be quite able to straightforwardly copy the movements of their instructors. However, when the instructor explains the consecutive moves to the very detail, several students appear not to be able to perform, even though they did so before. And reverse, when the instructor is asked about the details of a move which (s)he has never made explicit before, it is likely he or she has to perform first before being able to explain. Applying attentional and conscious control in motor-control hampers performance. Extreme examples are observed with patients suffering from the syndrome called apraxia. Apraxia denotes the inability of a patient to perform a certain skilled movement. For instance when asked to demonstrate teeth brushing, the patient is unable to do so, whereas he or she is perfectly able to brush the teeth in the morning.

Attention obviously has motoric implications, the examples show that internal attentional focus and conscious control of motor-skils may even lead to an inability to act.

The notion of external attentional focus, is not clearly defined and allows several interpretations. In a narrow, but easiest to define sense it denotes attention focusing on bringing about a single effect: directing a tennis ball, or throwing a single ball or bean bag into the air such that it can be caught. I will test this reading in the next section in the context of juggling.

3 Acting and Awareness

Five-ball juggling is hard and requires fast acting, the complication being that between throwing and catching the same ball four other objects – three of which are already up in the air - have to be handled. When first starting, it is a problem to throw each of the five balls one after the other before the first has returned (flashing as it is called), in doing so a novice will not be able to tell which ball was first thrown, let alone be able to catch it with the proper hand.

The novice juggler is trying to apply full and conscious attention, and that leads him or her astray. In juggling, the time lapse between throwing and catching a single ball is not more then a single second. Meanwhile, in five ball juggling four other objects are flying around appealing for attention. However, it is known that per second no more then two attentional shifts can occur, which is far too slow for five-ball juggling.

Juggling combines perception with action; in the one second between throwing and catching a particular ball four other objects have to be handled as well. Psychological experimentation has shown that the time required for the single voluntary act of pressing a button *only* when a light flashes is about 0.15 seconds (Austin, 1998). In contrast, observations of jugglers show that the time lapse between two catches of the same hand may be as little as 0.2 seconds (Polster, 2003). In this short interval several actions of this hand flow into each other: catching, bringing to throwing position (dwelling), throwing and preparing/waiting for the next, while in the middle of this series the other hand has to start its own series as well; refer to Polster (2003) for more details. A simple comparison of the time required for a voluntary act and the constraints of juggling shows the impossibility of juggling being a series of voluntary actions.

Because of the complexity and time constraints in five-ball juggling, correction of the movements and abandoning systematic flaws is quite difficult and requires persistence and endurance. An explanation is that there exist two independent systems or circuitries for the perceptual control of movement (Rossano, 2003). Raichle (1997) makes a distinction between "the neural circuitry underlying the unpractised, presumably conscious performance of a task on the one hand, and the practised presumably non-conscious performance of a task on the other hand." The response time of the latter circuitry is significantly shorter than that of the first (Raichle, 1997).

Voluntary actions are slow compared to involuntary acts, for instance a reflexive jerk takes only 0.025-0.05 seconds, which is in the order of five times faster then a voluntary act!

Internal attentional focus hampers execution of actions and actions are generally slower then when external attentional focus is applied. In five-ball juggling external attentional focus fails as there is not enough time to focus attention. Obviously the very fast, but complex and precision requiring moves in juggling cannot be under full conscious control. The juggler must be applying a different stance: a very sensate stance requiring awareness but avoiding any attentional focus; I call this stance: **being aware without focus.**

Indeed, an experienced juggler does not focus on the individual balls. In his juggling book Dancey (1994) advises: "*While learning [a five-ball pattern] you are trying to* make *yourself do it, when you can do it you* watch *yourself doing it."*

In five-ball juggling, there simply is not enough time to focus attention; restricting attention results in faster actions. However the surprising thing is that when no full attention is required for acting, the mind performs other tasks concurrently.

In daily life we perform many actions without attentional focus, for instance when walking the body performs an intricate combination of muscle activities to maintain posture; car driving and juggling are other examples. Three-ball juggling is less demanding than five-ball juggling. While juggling, the juggler can do other things as well, for instance speak, walk etc.; however non-focussed awareness is permanently required, when the juggler's attention drifts away and focuses elsewhere the balls drop. Car driving implies a similar requirement; the driver can perform many other things while driving but a certain level of awareness is required throughout.

I have avoided any attempt to define the notion of attention; therefore I cannot conclude that attention is not involved in the stance of being aware without focus. But referring to the spotlight metaphor, if there is attention involved, it is only a dim light. Because attention is a preliminary for consciousness this conclusion has implications for the role of consciousness as well.

The juggling example shows that no conscious mental act is required in order to perform, and what is more it shows that for fast acting no conscious mental act **can** precede the execution of the actions.

4 Acting and Emotions

Many cognitive scientists subscribe to the view that affect addresses the problems of decision making and action selection (Shanahan, 2005; Sloman, 2001). However, in the state of being aware without focus, the influence of affect seems much reduced.

Returning to the example of routinely driving the car on the way to work; our conscious mind was occupied of our plans for the coming day, and we

were at a sudden interrupted by the emergency break. The action of pressing the breaks was a straight reaction to events occurring around us, and as far as I can see it was not guided by any obvious emotion. Of course, emotions come up afterwards and may interfere with our consciously reconstructing the events, but they did not initiate nor guide the breaking action.

Literature on deliberate practice refers to emotions mainly by advising to attain an optimal emotion state and thinking positively (Wulf and Prinz, 2001).

Some descriptive evidence about the interference of emotions with acting can be found in the area of the eastern martial arts, in particular where Zen-Buddhism is involved. The aim of Zen-Buddhism is to voluntarily move into and try to intensify a mental state described as: "*When the ultimate perfection is attained, the body and limbs perform by themselves what is assigned to them to do with no interference from the mind. [The technical skill is so autonomised it is completely divorced from conscious efforts]*." (Takuan, translated in Suzuki 1959, the addition in brackets by Suzuki). The stance of being aware without focus, which I try to describe, bears similarities. Thus, though the aims are quite different, the Zen related literature contains interesting observations concerning the influence of affect and emotion on acting.

In Japanese, the state of perfection is called Mushin, which literally means "no-mind" or "without mind, without heart" (Austin 1998). Descriptions of this state are found in Hinduism as well; an interesting metaphor is given in the text called The Bhagavat Gita, it says that someone who masters this state, " … *withdraws all his senses from the attractions of their objects, even as a tortoise withdraws all its limbs*,…" (BG 2,58). The citation does not imply that the senses are withdrawn; the point is the mental stance with respect to the 'attractions' of the senses. Austin (1998) gives a further addition: "*The no mind of Zen implies a mental posture in which at least two things are going on: (1) bare attention still registers percepts, but (2) there are no emotional reverberations."*

The impact of emotion on performing is also described by the 20th century Zen master Taisen Deshimaru in a discourse for martial art practitioners: *"If our mind is upset, the natural functions of our bodies also tend to be disturbed. When the mind is calm, the body can act spontaneously ..."* (Taisen Deshimaru, 1982). In the ideal attitude of the swordsman this is pushed to the limit: "*The perfect swordsman takes no cognisance of the enemy's personality, no more than of his own. For he is an indifferent onlooker of the fatal drama of life and death in which he himself is the most active participant.*" (Suzuki, 1959).

Though my evidence on emotions is rather thin, I tend to conclude that intense emotions have a similar effect on performance and acting as focussed attention has.

Interesting to note at this point is an approach to deliberate practice developed Singer (1985, 1988) with aims at non-focused performance. Wulf and Prinz (2001) call it mysteriously *"a compromise between awareness and nonawareness strategies"*. It includes several steps: *readying* or arousal; *imaging* that is, going through the motion mentally; *focusing*, concentrating on a certain cue to block out all other thoughts; and *executing* the movement, while not thinking about the act itself or the possible outcome. This approach is much in line with the advices from Zen Buddhism, however it is seldom mentioned in the recent literature on deliberate practice.

4 Consciously Inhibiting Actions

A recent assumption in cognitive neuroscience is that the mind has a layered structure with at least three organising levels concerning body experience. "The lowest level is an assembly of neuronal information coming from all parts of the body; at the middle level the body schema are situated which secure the emergence of the conscious body image at the third level" (Yamadori, 1997). The body schema are subsystems 'implementing' James' ideomotor actions, for instance grabbing the coffee cup. Interesting for my analysis is the distinction between the second and the third level; are these levels really separate and may the second level operate independent from the third? The independence of the second level is shown by the split-brain studies and in particular very compellingly by the so-called *Anarchic hand* (Blakemore et al., 2002). The latter designates pathological behaviour in which a patient's right hand manipulates a tool properly but 'spontaneously', that is with the patient being aware of the hand acting, though neither consciously initiating the movement nor being able to inhibit the action. The anarchic hand shows that the neither attention nor consciousness are a prerequisite or a necessary condition (sine qua non) for action; they are not necessarily the initiator of actions. Moreover, it even shows that there exist pathological cases where consciousness is unable to inhibit actions.

Most people readily acknowledge that consciousness is not in control of the internal functioning of our body. The anarchic hand demonstrates that even skilful behaviour might be beyond the span of control of consciousness

Conclusions

I have made an attempt to describe the mental stance taken when performing regular everyday actions. I have called this stance *being aware without focus*; it is a stance in which there is typically little or no attentional focus.

Acting requires perception, action selection and action execution. These processes are often initiated and performed without any conscious deliberation; they are mostly on and below the edge of conscious experience and control.

Attention and emotions may interfere with acting but that often results in poorer or slower execution. Restricting attention results in faster actions. Surprisingly, if no full attention is applied for acting, the mind performs other tasks concurrently.

Attention is a gate to consciousness. Conscious thinking takes time and the often-supposed sequence that a mental act precedes bodily actions, or that we first think and then act cannot hold: it is too slow for many of our activities. In everyday practice we usually act before consciously thinking.

Conscious control is not a necessary condition for acting and consciousness only has weak control over the acting body, even though subjects have the feeling they consciously control their body.

Nevertheless, we do oversee our actions with our conscious and rational minds and except for pathological cases we are able to suppress many 'spontaneous' actions.

References

J.H. Austin, *Zen and the Brain*, MIT Press 1998.

B.J. Baars, *In the Theatre of Consciousness; The Workspace of the Mind,* Oxford University Press 1997.

S-J Blakemore, D.M. Wolpert and C.D. Frith, Abnormalities in the awareness of action, *TRENDS in Cognitive Sciences* Vol 6, no 6, 2002.

Chrisley, R., Clowes, R. W., & Torrance, S. "Next-generation approaches to machine consciousness". In R. Chrisley, R. W. Clowes & S. Torrance (eds.), *Proceedings of the AISB05 Symposium on Next Generation approaches to Machine Consciousness: Imagination, Development, Intersubjectivity, and Embodiment*, 2005.

C. Dancey, *Encyclopaedia of Ball Juggling*, Butterfingers, Bath UK 1994.

J. Decety, Do imagined and executed actions share the same neural substrate?. *Cognitive Brain Research*, 3:87-93, 1996.

Ericsson, K. A., R. Th. Krampe, and C. Tesch-Römer, 1993, 'The role of deliberate practice in the acquisition of expert performance.' *Psychological Review*, 100: 363-406.

Patrick Haggard, Sam Clark and Jeri Kalogeras. Voluntary action and conscious Awareness. *Nature Neuroscience* volume 5 no 4. 2002

W. James *The principles of Psychology*, 1890; Harvard University Press 1983.

B. Polster, *The mathematics of Juggling*, Springer-Verlag 2003.

M.E. Raichle, Automaticity: from reflective to reflexive information processing in the human brain, in: *Cognition, Computation and Consciousness*, K.Ito, Y. Miyashita and E. Rolls (eds), Oxford University Press, 1997.

M.J. Rossano, Expertise and the evolution of consciousness, *Cognition* Vol 89, (3) 2003

Shanahan, M. Consciousness, Emotion, and Imagination: A Brain-Inspired Architecture for Cogntive Robotics. In R. Chrisley, R. W. Clowes & S. Torrance (Eds.), Proceedings of the AISB05 Symposium on Next Generation approaches to Machine Consciousness: Imagination, Development, Intersubjectivity, and Embodiment, 2005.

Singer, R. N. (1985). Sport performance: A five-step mental approach. *Journal of Physical Education & Recreation*, **57**, 82-84.

Singer, R. N. (1988). Strategies and metastrategies in learning and performing self-paced athletic skills. *Sport Psychologist*, **2**, 49-68.

Aaron Sloman, Beyond Shallow Models of Emotion. *Cognitive Processing* 2 (1), 177-198. 2003.

Susan Stuart, The Binding Problem: Induction, Integration and Imagination, ". In R. Chrisley, R. W. Clowes & S. Torrance (eds.), *Proceedings of the AISB05 Symposium on Next Generation approaches to Machine Consciousness: Imagination, Development, Intersubjectivity, and Embodiment*, 2005.

Taisen Deshimaru. *The Zen Way to the Martial Arts*, Arkana, Penguin Books 1982.

D.T. Suzuki, *Zen and Japanese Culture*, Princeton University Press 1959

Gabriele Wulf and Wofgang Prinz Directing attention to movement effects enhances learning: A review, *Psychonomic Bulletin & Review*, Volume 8, Number 4, 1 December 2001, pp. 648-660(13)

A. Yamadori, Body awareness and its disorders, in: *Cognition, Computation and Consciousness*, K.Ito, Y. Miyashita and E. Rolls (eds), Oxford University Press, 1997.

Using Emotions on Autonomous Agents. The Role of Happiness, Sadness and Fear.

Miguel Angel Salichs Maria Malfaz
RoboticsLab, Carlos III University of Madrid
28911 Leganés, Madrid, Spain
salichs@ing.uc3m.es mmalfaz@ing.uc3m.es

Abstract

This paper addresses the use of emotions on autonomous agents for behaviour-selection learning, focusing in the emotions fear, happiness and sadness. The control architecture is based in a motivational model, which performs homeostatic control of the internal state of the agent. The behaviour-selection is learned by the agent using a Q-learning algorithm while there is no interaction with other agents. In situations where interaction arises (e.g. interacting with other agents), agents rely on stochastic games approaches as a learning strategy. The agent is intrinsically motivated and his final goal is to maximize Happiness. The learning algorithms use happiness/sadness of the agent as positive/negative reinforcement signals. Fear is used to prevent the agent choosing dangerous actions or being in dangerous states where non-controlled exogenous events, produced by external objects or other agents, could danger him. Preliminary tests have been carried out in a virtual world, based in a role-playing game.

1 Introduction

The goal of our project is to develop social robots with a high degree of autonomy. The social aspect of the robot will be reflected in the fact that the human interaction will not be considered only as a complement of the rest of the robot's functionalities, but as one of the basic features.

For this kind of robots, the autonomy and emotions makes them to behave as if they were "alive". This feature would help people to think about these robots not as simple machines but as real companions. Evidently, a robot that has his own "personality" is much more attractive than one that simply executes the orders that he is programmed to do.

Emotions can act as a control and learning mechanism, driving behaviour and reflecting how the robot is affected by, and adapts to, different factors over time (Fong et al, 2002). In previous works (Malfaz and Salichs, 2004), an emotion-based architecture has been proposed.

Some researchers have also used emotions in robots. Most of them have made emphasis in the external expression of emotions (Breazeal, 2002) (Fujita, 2001) (Shibata et al 1999). Their robots include the possibility of showing emotions, by facial and sometimes body expressions. In this case, the emotions can be considered just as a particular type of information that is exchanged in the human-robot interaction process. In nature emotions have different purposes and interaction is only one of them. We intend to make use of emotions in robots trying to imitate their purpose in nature, which includes, but is not limited to, interaction. The role that plays each emotion and how the mechanisms associated to each one work are very specific. That means that each emotion must be incorporated to the robot in a particular way. In this paper we will present some basic ideas on how emotions such as happiness, sadness and fear can be used in an autonomous robot.

Emotions will be generated from the evaluation of the wellbeing of the robot. Happiness is produced because something good has happened, i.e. an increment of the wellbeing is produced. On the contrary, Sadness is produced because something bad has happened, so its wellbeing decreases. Fear appears when the possibility of something bad is about to happen. In this case, we expect that the wellbeing drops off. Finally, Anger is produced when a decrement of the wellbeing of the robot happened due to another-initiated act.

This paper presents a control architecture for an autonomous agent based on motivations. The agent uses reinforcement learning algorithms to learn its policy while interacts with the world. The reward for these learning algorithms will be the variation of the wellbeing of the agent (happiness/sadness) due

to the previous selected behaviour, calculated at each step of the process. This wellbeing is a function of the internal needs of the agent (drives). This idea of using the wellbeing of the agent as the reinforcement in the learning process for behaviour selection has been also used by Gadanho in the ALEC architecture, obtaining quite good results (Gadanho, 2003).

The remainder of the paper is organized as follows. Section 2 introduces the use of emotions in robots. Section 3 and 4 describe the proposed control architecture and the reinforcement learning algorithms respectively. Section 5 introduces the emotion fear and section 6 describes the experimental setting. Finally, conclusions and future works are summarized in section 7.

2 Emotions in robotic

One of the main objectives in robotics and artificial intelligence research is to imitate the human mind and behaviour. For this purpose the studies of psychologists on the working mind and the factors involved in the decision making are used. In fact, it has been proved that two highly cognitive actions are dependant not only on rules and laws, but on emotions: Decision making and perception (Picard, 1998). In fact, some authors affirm that emotions are generated through cognitive processes. Therefore emotions depend on ones interpretation, i.e. the same situation can produce different emotions on each agent, such as in a football match (Ortony, 1988). Moreover, emotions can be considered as part of a provision for ensuring and satisfaction of the system's major goals (Frijda, 1987).

Emotions play a very important role in human behaviour, communication and social interaction. Emotions also influence cognitive processes, particularly problem solving and decision making (Damasio, 1994). In recent years, emotion has increasingly been used in interface and robot design, primarily in recognition that people tend to treat computers as they treat other people.

There are several theories about emotions (Frijda 1987; Ortony, 1988; Sloman, 2003; Rolls, 2003), but the results of Damasio (1994) can be considered the basis, for many A.I. researchers, to justify the use of emotions in robotics and their computation.
Rosalind Picard in her book Affective Computing (1998), writes a complete dissertation about this subject based on several psychologists, including Damasio. Picard (1998) proposed a design criterion in order to create a computer that could express emotions. Moreover, she established that a computer has emotions if it has certain components that are present on the emotional systems of healthy people. Picard (2003) expounded four motives for giving certain emotional abilities to machines: The first goal is to build robots and synthetic characters that can emulate living humans and animals, such as a humanoid robot. The second is to make machines that are intelligent. A third objective is to try to understand human emotions by modelling them. Although these three goals are important, the main one is to make machines less frustrating to interact with, i.e. to facilitate the human-machine interface.

Cañamero (2003) considers that emotions, or at least a sub-group of them, are one of the mechanisms founded in biological agents to confront their environment. This creates ease of autonomy and adaptation. For this reason she considers that it could be useful to exploit this role of emotions to design mechanisms for an autonomous robot. Emotions are used as mechanisms that allow the agent (robot) to:

1. Have fast reactions.
2. Contribute to resolve the selection among multiple objectives.
3. Signal important events to others.

Bellman (2003) agrees, to some degree, with Cañamero and her reasons for considering emotions in robotics. The author states that emotions allow animals with emotions to survive better than the others without emotions. Therefore, we can presume that some type of analogy to emotional abilities is required within robots, if we want an intelligent and independent behaviour within a real environment.

Changing subject, Picard (2003) gives an advice about the implementation in machines of functions implemented by the human emotional system. Computers do not have emotions as human beings in any natural experimentation sense. Science methodology is to try to reduce complex phenomena, such as emotions, to a functional requirements list. The challenge of many computing science researchers is to try to duplicate these in computers at different levels depending on the motives of the investigation. But we must be careful when presenting this challenge to the general public, who may perceive that emotions are the frontier that separates man and machine

3 Control Architecture

An independent system should not have to wait for someone to maintain, succour, and help it (Frijda and Swagerman,1987). Therefore, an autonomous agent should be capable of determining its goals, and it must be capable of selecting the most suitable behaviour in order to reach its goals. Similarly to other authors (Avila-Garcia and Cañamero, 2004), (Breazeal, 2002), (Gadanho, 2003), (Velasquez, 1998), our agent's autonomy relies on a motiva-

tional model. Figure 1 shows this proposed control architecture for behaviour selection.

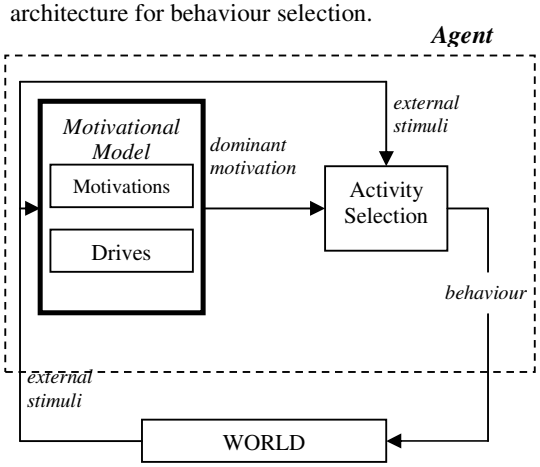


Figure 1: Control architecture for autonomous agents

3.1 Motivational Model

Motivations can be seen as homeostatic processes, which maintain a controlled physiological variable within a certain range. Homeostasis means maintaining a stable internal state (Berridge, 2004). This internal state can be parameterized by several variables, which must be around an ideal level. When the value of these variables differs from the ideal one, an error signal occurs: the drive. These drives constitute urges to action based on bodily needs related to self-sufficiency and survival. External stimuli, both innate and learned, are also able to motivate and drive behaviour (Cañamero, 1997).

In order to model motivation, the hydraulic model of motivation described by Lorentz and Leyhausen in (Lorentz and Leyhausen, 1973) has been used as an inspiration. This model is essentially a metaphor that suggests that motivational drive grows internally and operates a bit like pressure from a fluid reservoir that grows until it bursts through an outlet. Motivational stimuli from the external world act to open an outflow valve, releasing drive to be expressed in behaviour. In this model, internal drive strength interacts with external stimulus strength. If drive is low, then, a strong stimulus is needed to trigger motivated behaviour. If the drive is high, then, a mild stimulus is sufficient (Berridge, 2004). Following this idea, the intensity of motivations (M_i) is a combination of the intensity of the related drive (D_i) and the related external stimuli (w_i), as it is expressed in the following equation:

$$M_i = D_i + w_i \tag{1}$$

The ideal value for all the drives is 0. The external stimuli are the different objects that the player can find in the virtual world during the game. If the stimulus is present the value of w_i is 1, otherwise is 0.

According to (1), the intensity of a motivation is high due to two reasons: 1) the correspondent drive is high or 2) The correct stimulus is present. The dominant motivation is the one with the highest intensity.

This model can explain the fact that due to the availability of food in front of us, we sometimes eat although we are not hungry. We have also introduced activation levels (L_d) for motivations such that:

$$\begin{aligned} &\textit{if } D_i \le L_d \textit{ then } M_i = 0 \\ &\textit{if } D_i > L_d \textit{ then } (1) \text{ is applied} \end{aligned} \tag{2}$$

Therefore the possibility of no dominant motivation exists.

3.2 Wellbeing

As shown in (3), the agent's wellbeing is a function of the values of the drives (D_i) and some "personality" factors (α_i).

$$Wb = Wb_{ideal} - \sum_i \alpha_i \cdot D_i \tag{3}$$

Wb_{ideal} is the ideal value of the wellbeing of the agent, which is set to 100. The personality factors weight the importance of the values of the drives on the wellbeing of the agent. The value of the wellbeing and its variation (ΔWb) are calculated at each step. The variation of the wellbeing is calculated as the current value of the wellbeing minus the wellbeing value in the previous step.

3.3 Behaviour Selection

The action selection process consists in making decisions as to what behaviours to execute in order to satisfy internal goals and guarantee survival in a given environment and situation. For other authors (Avila and Cañamero, 2002), (Avila and Cañamero, 2004), (Cañamero, 1997) this implies that the agent can choose among some behaviors related to the dominant motivation. Therefore for each motivation there is a set of behaviours oriented to fulfill the motivational goal.

It is important to note that finally, the agent will learn that when the dominant motivation is Eat, it must select among the behaviours related to the object food, instead of those associated to water or medicine. The novelty of our approach is that these

behaviours were not linked a priori with the correspondent motivations.

3.4 Happiness and Sadness

Considering the definitions of the emotions given in the introduction section:

$$\begin{aligned} &If\ \Delta Wb > L_h \Rightarrow Happiness \\ &If\ \Delta Wb < L_s \Rightarrow Sadness \end{aligned} \quad (4)$$

Where $L_h>0$ and $L_s<0$ are the minimum variations of the wellbeing of the agent that produce Happiness or Sadness respectively. Therefore these two emotions are used by the agent as the reward for the reinforcement learning algorithms.

In this architecture the agent learns, using different reinforcement learning algorithms, the best behaviour at each step using happiness/sadness as the positive/negative reward. Therefore, in this architecture behaviours are not selected to satisfy the goals determined by the dominant motivation but to optimize the wellbeing of the agent. This implies that the final goal of the agent is to maximize Happiness.

4 Reinforcement Learning

Reinforcement learning (RL) is about learning from interaction how to behave in order to achieve a goal. The agent's objective is to maximize the amount of reward it receives over time (Sutton and Barto, 1998). Q-learning is a value learning version of RL that learns utility values (Q-values) of state and action pairs Q(s,a). It provides a simple way for agents to learn how to act optimally in controlled Markovian domains (Yang and Gu, 2004). The theory of Markov Decision Processes (MDP's), assumes that the agent's environment is stationary and as such contains no other adaptive agents (Littman, 1994). Therefore, while the agent is not interacting with the other agent, we will consider our virtual world as a MDP environment.

On the other hand, if the agent is interacting with other player, the rewards the agent receives depend not only on their own actions but also on the action of the other agent. Therefore, the individual Q-learning methods are unable to model the dynamics of simultaneous learners in the shared environment. Currently multiagent learning has focused on the theoretic framework of Stochastic Games (SGs) or Markov Games (MGs). SGs appear to be a natural and powerful extension of MDPs to multiagent domains (Yang and Gu, 2004).

Taking into account these considerations, in the proposed architecture the agent will use the standard Q-learning algorithm as the RL algorithm when the agent is not interacting with the other player. Obviously, in the case of "social" interaction, the agent must use a multiagent RL algorithm. The following subsections explain in more details these two scenarios.

In our system, the state of the agent is the aggregation of his inner state S_{inner} and the states S_{obj} related to each of the objects, including external agents, which can interact with him.

$$S = S_{inner} \times S_{obj_1} \times S_{obj_2} \ldots \quad (5)$$

For the RL algorithms the states related to the objects are considered as independent. This means that the state of the agent in relation with each object is $s \in S_{inner} \times S_{obj_i}$

4.1 Q-learning Algorithm

As mentioned previously, in MDP environments the agent will use the standard Q-Learning as a learning algorithm. As described in (Gadanho, 2002), through this algorithm the agent learns iteratively by trial and error the expected discounted cumulative reinforcement that it will receive after executing an action a in response to a world state s, the Q-values for each object is:

$$Q^{obj_i}(s,a) = (1-\alpha)\cdot Q^{obj_i}(s,a) + \alpha\cdot\left(r + \gamma \max_{a \in A_{obj_i}} \left(Q^{obj_i}(s',a)\right)\right) \quad (6)$$

where A_{obj_i} is the set of actions related to the object i, s' is the new state, r is the reinforcement; γ is the discount factor and α is the learning rate parameter.

The optimal policy, chooses the action that maximizes $Q^{obj_i}(s,a)$ this means

$$a^* = \arg\max_a Q^{obj_i}(s,a) \quad (7)$$

The proposed architecture differs from others in that we do not consider only the behaviours that help to satisfy the drive related with the dominant motivation but the agent must consider all the behaviours that can be performed at each step, depending on his states.

4.2 Multiagent reinforcement learning

In multiagent systems, other adapting agents make the environment no longer stationary so the Markov property is not applicable. In the learning framework of SGs, learning agents attempt to maximize their expected sum of discounted rewards. Unlike single-agent system, in multiagent systems the joint actions determine the next state and rewards of each agent. In (Littman, 1994) it is proposed a Minimax-

Q learning algorithm for zero-sum games in which the player always tries to maximize its expected value in the face of the worst-possible action choice of the opponent. The player's interests in the game are opposite. Later, Littman (Littman, 2001) proposed the Friend or Foe Q-learning algorithm, for the RL in general-sum SGs. The main idea is that each agent is identified in advance as being either "friend" or "foe". The Friend class consists of SGs in which the Q-values of the players define a game which has a coordination equilibrium. The Foe class is the one in which the Q-values define a game with an adversarial equilibrium. The Friend-Q updates similarly to regular Q-learning, and Foe-Q updates as does minimax-Q (Shoham et al, 2003).

All these algorithms extend the normal Q-function of state-action pairs $Q^{obj_i}(s,a)$ to a function of states and joint actions of all agents. Taking into account this fact and that each agent can select among n actions while they are interacting, the Q-values to be calculated are $Q^{obj_i}(s,a_1,a_2)$ where a_1 and a_2 belong to the set of n actions of each agent.

5 Fear

Fear is produced when the agent knows that something bad may happen. This means that the wellbeing of the agent might decrease. To cope with fear the action that produces the negative effect is going to be considered. We will distinguish between actions executed by the agent and exogenous actions carried out by other elements of the environment such as other agents.

5.1 To be afraid of executing risky actions

Q-learning algorithm evaluates every action carried out in a state, using the expected average value. However, since the system is non deterministic, the result of a certain action may have different values. The worst result experimented by the agent for each pair action-state is stored in a variable called $Q^{obj_i}_{worst}(s,a)$, which is updated after the execution of the action.

$$Q^{obj_i}_{worst}(s,a) = \min(\, Q^{obj_i}_{worst}(\, s,a),\; r + \gamma \max_{a \in A_{obj_i}} (Q^{obj_i}(s',a))) \tag{8}$$

where A_{obj_i} is the set of actions, *s'* is the new state, *r* is the reinforcement and γ is the discount factor.
The effect of being afraid can be considered by choosing the action that maximizes $Q^{obj_i}_{fear}$ instead of choosing the one that maximizes Q^{obj_i},

$$Q^{obj_i}_{fear}(s,a) = \beta Q^{obj_i}(\, s,a) + (1-\beta) Q^{obj_i}_{worst}(\, s,a) \tag{9}$$

Using this approach the expected result of each action is considered as well as the less favourable one. The parameter β, being $0 \leq \beta \leq 1$, measures the daring degree of the agent, and its value will depend on the personality of the agent. If the agent is fearless, β will be near 1; while in a fearful agent, who tries to minimize the risk, β will be near 0. If $\beta = 1$ the agent is using the optimal policy.

This means that the "fearful" policy chooses the action:

$$a^f = \arg\max_a Q^{obj_i}_{fear}(\, s,a) \tag{10}$$

For example, when an agent has to pass over a deep hole, he can choose between jumping over it and going around it. Jumping is easier, faster and usually safe, but very occasionally he can fail and die. On the other hand, if the agent goes around the hole he will take a lot of time and get tired but it is safer. Translating this example to our point of view, the Q-value related with jumping will be greater than the one related to going around. Using the standard Q-learning algorithm, the agent would always jump over the hole. Using the fearful policy, considering the worst thing that could happen to the agent jumping or going around, he would choose going around since it is safer than jumping.

5.2 To be afraid of malicious exogenous actions

When the agent may suffer some negative effects in a state as a consequence of exogenous events, feels fear. "Fear" is expressed as a drive D_{fear}.

Traditionally, Q-learning has been applied on Markov decision processes (MDP), which are discrete time systems. Some authors have extended the use of this algorithm to continuous time systems by considering them as semi Markov decision processes. In both cases it is commonly assumed that there are no exogenous events. In order to introduce the effects of exogenous events in continuous systems we consider the system as a discrete time system with constant period. In the limit, if the period is very small the system will tend to be a continuous time system. Moreover, we will also consider that the exogenous events can be associated to other agents or elements of the environment. These exogenous events are synchronized with the actions executed by the agent. Among these action we will include the action of "doing nothing". In this case the treatment for multiagent systems mentioned before will be applied.

The exogenous events executed by an external object or other agent can occur simultaneously to any of the actions of the agent. Therefore the negative effects of these exogenous events will be reflected in all the actions of the agent. In order to separate the effects of the actions of the agent and the effects of the exogenous events, we will focus on the study of the agent when he is "doing nothing". In that case, we suppose that all the changes suffered by the agent are a consequence of external elements.

It will be considered that a state is a "scary" state when:

$$Q_{worst}^{obj_i}(s, Nothing) < L_{fear} \quad (11)$$

being L_{fear} the minimum acceptable value of the worst result that can be expected by the agent when it is doing nothing. In this case the value of the fear drive D_{fear} will be incremented.
When

$$Q_{worst}^{obj_i}(s, Nothing) > L_{safe} \quad (12)$$

it is considered that the agent is in a "safe" state and the value of the fear drive D_{fear} will be decreased.

The fear drive is equally treated as the rest of drives, and its related motivation could be the dominant one. In this case, the agent will learn by itself what to do when it is afraid.

6 Experimental Test Bed

The proposed architecture is intended to be used in a social personal robot developed by our lab and named "Maggie" (see Fig2) (Salichs et al, 2006). As a first stage of this project and due to the obvious physical difficulties of making experiments on a real robot and on a real environment, we decided to implement our architecture on virtual players, who "live" in a virtual world, a text-based multi user role game. This game gave us the possibly of creating different 2-D environments to play in, as well as a graphic interface.

Table 1 shows our agent's motivations, drives and external stimuli that the agent can find in the virtual world.

These drives have been selected tacking into account the role of the agent in the virtual world used to implement our architecture. Since our final goal is to construct an autonomous social robot, it must show social behaviours. Therefore, as it is shown, social motivations are included as robot's needs.

Table 1: Motivations, drives and related stimuli

Drive/Motivation	External Stimuli
Energy	Food
Thirst	Water
Health	Medicine
Sociability	Other player
Fear	

At each simulation step some of these drives, such as Energy, Thirst, Health and Sociability are incremented by a certain amount. The value of the drive Fear, as it was previously explained, increase or decrease depending on if the agent is in a "scary" state or not.

Following (3) the wellbeing of the agent is defined by:

$$Wb = Wb_{ideal} - (\alpha_1 D_{energy} + \alpha_2 D_{thirst} + \alpha_3 D_{health} + \alpha_4 D_{social} + \alpha_5 D_{fear}) \quad (13)$$

In our test bed the inner state is then:

$$S_{inner} = \{Hungry, Thirsty, Ill, Bored, Scary, OK\} \quad (14)$$

This internal state is obviously related with the dominant motivation. Therefore when the dominant motivation is for example "Eat" then the agent is "Hungry" and so on.

In relation with static objects the agent can be in the following states:

$$S_{obj} = Have\_it \times Near\_of \times Know\_where \quad (15)$$

where,

$$Have\_it = \{yes, no\} \quad (16)$$

$$Near\_of = \{yes, no\} \quad (17)$$

$$Know\_where = \{yes, no\} \quad (18)$$

In relation with other player:

$$S_{obj} = Near\_of \quad (19)$$

where,

$$Near\_of = \{yes, no\} \quad (20)$$

And the set of actions that can be executed in every state is the following:

$$A_{food} = \{Eat, Get, Go\_to, Explore\} \quad (21)$$

$$A_{water} = \{Drink, Get, Go\_to, Explore\} \quad (22)$$

$$A_{medicine} = \{Take, Get, Go\_to, Explore\} \quad (23)$$

$$A_{playmate} = \begin{cases} Explore \\ Steal\ \ food / water / medicine \\ Give\ \ food / water / medicine \\ Chat \end{cases} \quad (25)$$

Among the previously mentioned behaviours there are some of them that reduce or increase some drives, and therefore will produce a variation in the emotional state of the agent:

- Eat food: reduces to zero the Energy drive. (happiness when hungry)
- Drink water: reduces to zero the Thirst drive. (happiness when thirsty)
- Take medicine: reduces to zero the Health drive. (happiness when sick)
- Chat: reduces to zero the Social drive. (happiness when the social drive is high)
- To be taken something by other player: increases by a certain amount the Social drive. (sadness)
- To be given something from other player: reduces by a certain amount the Social drive. (happiness when the social drive is high)

Fig. 2. "Maggie" The Social Robot of the Robotic Lab.

The conducted experiments show the usefulness of the proposed architecture in facilitating the development of social autonomous agents able to learn from the experience the right behaviours to execute depending on the world state.

7 Conclusion and Future work

In this paper different reinforcement learning algorithms have been discussed and implemented for the behaviour-selection learning of non-interacting and social autonomous agents. These agents are controlled by an emotion-based architecture, which performs homeostatic control of the internal state of the agent through an embedded motivational model. This architecture has been designed for autonomous and social robots.

The agent is intrinsically motivated and his goal is his own wellbeing. The learning algorithms use happiness/sadness of the agent as positive/negative reinforcement signals. Fear is used to prevent the agent choosing dangerous actions or being in dangerous states where non-controlled exogenous events, produced by external objects or other agents, could danger him.

In the future work, it is expected that the agent learns not only the right policy but also to identify its opponent. So far, the agent treats all its opponents as if they were all the same, and this is not true. In future scenarios, the agent will be able to behave different with the "good" opponent than with the one that tries to steal its objects every time that interacts with it.

Another emotion is going to be implemented: Anger. Anger will be produced when sadness arises due to the interaction with another agent

Acknowledgements

The authors gratefully acknowledge the funds provided by the Spanish Government through the projects named "Personal Robotic Assistant" (PRA) and "Peer to Peer Robot-Human Interaction" (R2H), of MEC (Ministry of Science and Education).

References

Avila-Garcia, O. and Cañamero, L. A Comparison of Behavior Selection Architectures Using Viability Indicators. In *Proc. InternationalWorkshop Biologically-Inspired Robotics: The Legacy of W. Grey Walter*(WGW'02). 2002.

Avila-Garcia, O. and Cañamero, L. Using Hormonal Feedback to Modulate Action Selection in a Competitive Scenario. *In Proc. 8th Intl.Conference on Simulation of Adaptive Behavior (SAB'04)*. 2004

Bellman , Kirstie L.. Emotions: Meaningful mappings between the individual and its world. In: *Emotions in Humans and Artifacts*. (Robert Trappl, Paolo Petta and Sabine Payr), pp 149-188. The MIT Press. Cambridge, Massachusetts. 2003

Berridge , Kent C. Motivation concepts in behavioural neuroscience. *Physiology & Behaviour* 81, 179-209,2004,.

Breazeal C. *Designing Sociable Robots*. The MIT Press. 2002

Cañamero, D. Modeling Motivations and Emotions as a Basis for Intelligent Behavior. *In W. Lewis Johnson, ed., Proceedings of the First International Symposium on Autonomous Agents (Agents'97)*, 148-155. New York, NY: The ACM Press. 1997.

Cañamero, D. Designing emotions for activity selection in autonomous agents. In: *Emotions in Humans and Artifacts.* (Robert Trappl, Paolo Petta and Sabine Payr), pp 115- 148. The MIT Press. Cambridge, Massachusetts. 2003

Damasio, Antonio. *Descartes' Error – Emotion, reason and human brain*. Picador, London. 1994

Fong. T., Nourbakhsh. I., Dautenhahn K. *A survey of socially interactive robots: Concepts, design, and applications.* Technical Report CMU-RI-TR-02-29. 2002

Frijda, N. and Swagerman, J. Can computers feel? Theory and design of an emotional model. *Cognition and Emotion*. 1 (3). pp 235-357. 1987

Fujita, Masahiro AIBO: Toward the Era of Digital Creatures. *The International Journal of Robotics Research*. Vol 20, Nº 10, pp 781-794. October 2001

Gadanho, Sandra Clara. Emotional and Cognitive Adaptation in Real Environments. *In: Symposium ACE'2002 of the 16th European Meeting on Cybernetics and Systems Research*, Vienna, Austria. 2002

Gadanho, Sandra Clara. Learning behavior-selection by emotions and cognition in a multi-goal robot task. *The Journal of Machine Learning Research.* Volume 4 Pages: 385 – 412. MIT Press Cambridge, MA, USA. 2003

Littman, M. L. Markov games as a framework for multiagent learning. *In Proceedings of the Eleventh International Conference on Machine Learning*, San Francisco, California, pp. 157--163. 1994

Littman, M. L. Friend-or-foe Q-learning in general-sum games. *In Proceedings of the Eighteenth International Conference on Machine Learning*, pages 322--328, Williams College, June 2001.

Lorentz K, Leyhausen P. *Motivation of human and animal behaviour; an ethological view*. New York: Van Nostrand-Reinhold; xix, 423 pp. 1973

Malfaz, M. and Salichs, M.A.. A new architecture for autonomous robots based on emotions. *Fifth IFAC Symposium on Intelligent Autonomous Vehicles*. Lisbon. Portugal. Jul, 2004.

Ortony, A., Clore, G. L., and Collins, A.. *The Cognitive Structure of Emotions.* Cambridge University Press. Cambridge, UK. 1988

Picard, Rosalind W. *Affective computing.* Ed. Ariel S.A. 1998

Picard, Rosalind W. What does it mean for a computer to have emotions?. In. *Emotions in Humans and Artifacts*. (Robert Trappl, Paolo Petta and Sabine Payr), pp 213-235. The MIT Press. Cambridge, Massachusetts. 2003

Rolls, Edmund, T. A Theory of emotion, its functions, and its adaptive value. In: *Emotions in Humans and Artifacts.* (Robert Trappl, Paolo Petta and Sabine Payr), pp 11-35. The MIT Press. Cambridge, Massachusetts. 2003

Salichs, M. et al. Maggie: A Robotic Platform for Human-Robot Social Interaction. In *2006 IEEE International Conferences on Cybernetics & Intelligent Systems (CIS) and Robotics, Automation & Mechatronics.* Bangkok, Thailand. 2006

Shibata T., Tashima T.,Arao M., Tanie K. Interpretation in Phisical Interaction between human and artificial emotional creature. *Proceedings of the 1999 IEEE. International Workshop on Robot and Human Interaction*. Pisa, Italy – September 1999

Shoham, Y., Powers, R. and Grenager, T. *Multi-agent reinforcement learning: a critical survey.* Technical report, Computer Science Department, Stanford University, Stanford. 2003.

Sloman, Aaron. How many separately evolved emotional beasties live within us. In: *Emotions in Humans and Artifacts*. (Robert Trappl, Paolo Petta and Sabine Payr), pp 35-115. The MIT Press. Cambridge, Massachusetts. 2003

Sutton, Richard S. and Barto, Andrew G. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, A Bradford Book.1998

Velásquez, J. When Robots Weep: Emotional Memories and Decision Making. *In: Proceedings of AAAI-98*. 1998

Yang E. and Gu D. *Multiagent Reinforcement Learning for Multi-Robot Systems: A Survey.* Technical Report CSM-404.University of Essex. (2004)

Towards a Computational Account of Reflexive Consciousness

Murray Shanahan

Department of Computing, Imperial College London, 180 Queen's Gate, London SW7 2AZ, UK.
m.shanahan@imperial.ac.uk

Abstract

This paper offers a preliminary sketch for an account of reflexive consciousness based on an implemented architecture that combines a global workspace architecture with an internally closed sensorimotor loop. The proposed account extends the theoretical framework of the already implemented architecture with to two concepts that structure the flow of consciously processed information. First, contextual switches divide the unfolding contents of consciousness into a set of nested episodes, wherein one conscious episode can "refer to" another. Second, the imposition of a focus / fringe structure enables consciousness to encompass material that is merely available to it but not actually present. This combination of reflexivity and fringe may underpin our awareness of our own existence as conscious beings.

1 Introduction

Cognitive theories of consciousness, as the name suggests, posit an intimate link between cognition and consciousness. For example, according to *global workspace theory* (Baars, 1988; 1997; 2002), non-conscious information processing in the human brain is carried out locally within specialist brain processes, while the hallmark of consciously processed information is that it is broadcast (via a "global workspace") and made available to the entire set of these specialists. The upshot is that consciously processed information is cognitively efficacious in ways that non-consciously processed information is not. Specifically, the procession of broadcast global workspace states resembles a serial thread of computation, yet it integrates the results of massively parallel computation, sifting out relevant contributions from the irrelevant (Shanahan & Baars, 2005).

However, one feature of conscious human thought not accounted for by global workspace theory in its basic guise is *reflexivity*, that is to say the capacity for a conscious thought to refer to itself or to other conscious states. (By contrast, so-called higher-order thought (HOT) theories of consciousness take reflexivity as their primary datum (Rosenthal, 1986).) If consciously processed information is, as global workspace theory maintains, cognitively efficacious, then reflexively conscious information processing is even more so – since it enables the thinking subject to reflect on his or her own mental operations, to critique them and improve on them, and to respond to the ongoing situation in ways that depend on a degree of self-knowledge. So the question arises: Can global workspace theory be extended to account for reflexive consciousness?

This question has phenomenological as well as cognitive implications. For if we accept the argument of Shanahan (2005), the very idea of a conscious subject – something it is like something to be, in Nagel's well-known terminology – can be objectively accounted for in terms of a suitably *embodied* instantiation of the global workspace architecture, wherein all the specialist processes are indexically directed towards maintaining the wellbeing and fulfilling the purpose (or "mission") of a single, spatially unified body. By extending global workspace theory to reflexive consciousness, we can bolster this line of argument by showing that a similar treatment is available for a vital aspect of human phenomenology, namely our ability to become conscious of our own existence *as* conscious subjects.

2 Internal Simulation with a Global Workspace

Figure 1. illustrates the operation of the global workspace architecture, which comprises a set of specialist brain processes plus a global workspace. Information processing within the architecture consists of periods of *competition* interleaved with periods of *broadcast*. On the left of the figure, we see the set of specialist processes competing to gain access to the global workspace. Gaining access entails that the winning process (or coalition of processes) gets to broadcast its message, via the global

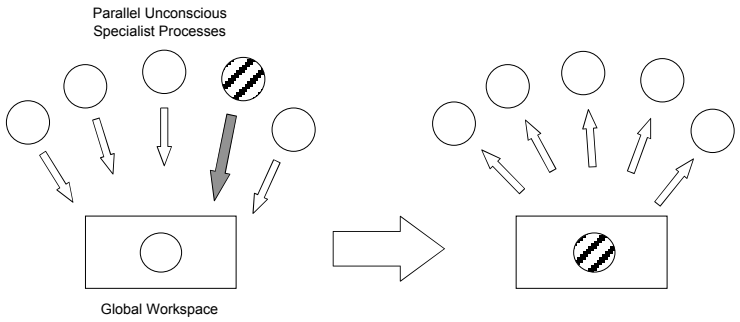


Fig.1: The Global Workspace Architecture.

workspace, back to the entire set of specialists, as seen on the right of the figure. The global workspace itself is, in essence, nothing more than the infrastructure of a communications network that permits signals generated within localised neuronal populations to influence remote, widespread brain regions. According to global workspace theory, the mammalian brain instantiates such an architecture, and this allows us draw a empirically falsifiable distinction between consciously and non-consciously processed information. Information processing that is confined to local specialists is necessarily non-conscious, and only broadcast information can be consciously processed.

Although global workspace architecture permits this fundamental distinction to be drawn in a theoretically respectable manner, it still leaves open the question of the content of consciously processed information. But by augmenting the basic global workspace architecture with an internally closed sensorimotor loop (Fig. 2), it is possible to reconcile it with another idea current within the scientific study of consciousness, namely the *simulation hypothesis*, according to which thought is internally simulated interaction with the environment (Cotterill, 1998; Hesslow, 2002; Shanahan, 2006). If the sophisticated mental life of a human being results from the interplay of external stimulation with internally generated activity such as inner speech and mental imagery, then something like the internally closed sensorimotor loop posited by the simulation hypothesis is required to account for it. Moreover, by facilitating the *rehearsal* of trajectories through sensorimotor space, the internal sensorimotor loop helps the individual to anticipate the consequences of their actions and to plan ahead, and thereby fulfils a fundamental cognitive role.

In (Shanahan, 2006), a implemented system is described that reconciles global workspace theory with the simulation hypothesis. The system controls a simple two-wheeled robot with a camera, and enables it to select an action based not only on a set of reactive responses, but also taking into consideration the result of simulating the expected outcomes of its actions using an internal sensorimotor loop, as depicted in Figure 3. Moreover, a global workspace is incorporated into the loop. The procession of states exhibitted by the global workspace, which simulates a possible trajectory through the robot's sensorimotor space, is the outcome of both competition and broadcast : the i^{th} state being broadcast to multiple neuronal populations which then compete to determine the $i+1^{th}$ state. Further details of the system are beyond the scope of this article, and can be found in (Shanahan, 2006).

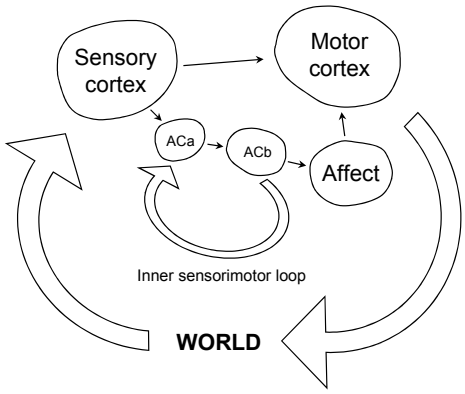


Fig.2: External and Internal Sensorimotor Loops

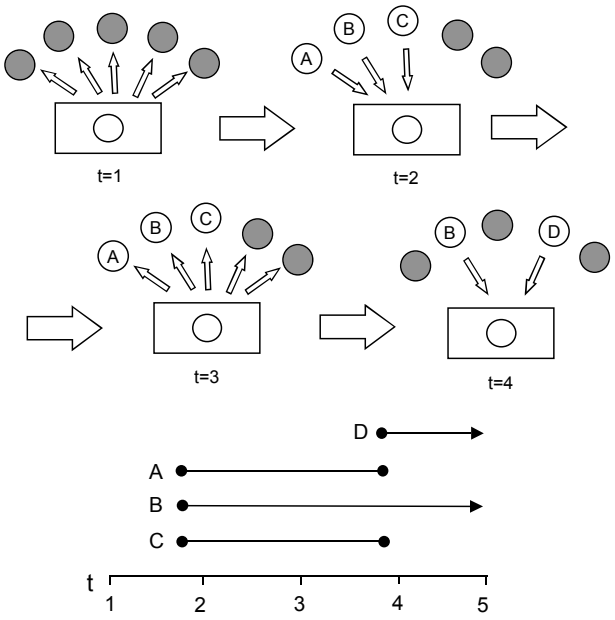


Fig. 3: The Temporal Structure of Consciously Processed Information

3 Context and Temporal Structure

According to the account of reflexive consciousness proposed by this paper, the flow of information through the global workspace is divided into distinct, but possibly nested or overlapping, episodes at various timescales. Beginnings and ends of conscious episodes are triggered by events (contextual switches) – such as entering and leaving a room, or meeting and parting from a friend – which wake up or send to sleep relevant specialist processes, whose job it is to manage the individual's response to situations of that particular type. Figure 3 illustrates the idea. On the left of the figure are snapshots of the global workspace (GW) at four time points. At t=1, all five of the processes depicted are dormant, although they are still receiving information broadcast from GW. This information indicates the occurrence of a distinctive event – a *contextual switch* – and by t=2 this has caused three of the processes (A, B, and C) to become active and begin competing for access to GW. There follows a further period of broadcast (t=3), indicating a new contextual switch. By t=4, this has caused processes A and C to go back to sleep, but has woken up process D.

As Figure 3 shows, competition for access to GW is restricted to the currently active or "awake" set of specialist processes, and the set of active processes can be thought of as reflecting the current *context* (Fig. 3, top), a conception which is broadly in line with the notion of context prominently deployed by Baars (1988) in his original presentation of global workspace theory. Each distinct conscious episode, bracketed by a pair of contextual switches, falls under the jurisdiction of a particular process, a process that should be relevant in the current context. Intuitively, temporal context is a richly structured, hierarchical concept. The context of a lunchtime falls within the larger context of a day, while the context of a conversation can overlap the context of a lunchtime. Similarly, conscious episodes, which are associated with temporal contexts, can be nested or overlapping. However, it should be noted that diagrams such as Figure 3 (bottom) only show the set of processes that have the *potential* to contribute to the unfolding content of the global workspace at any given time point. For example, although process B is *active* at time t=3 in Figure 3, this does not entail that it has won (full) *access* to GW at time t=3. This means that (focal) consciousness typically does not contribute to a conscious episode for its entire duration, but only at those times when the corresponding process gains access to GW. On the other hand, as we'll see in the next section, any active process competing for access can contribute to *fringe* consciousness.

Allowing specialist processes to wake up and go to sleep in response to contextual cues gives them a simple form of internal state (on or off), and therefore allows them to respond to information in a way that is sensitive to past events. But from the standpoint of the present paper, the most important consequence of this demarcation of conscious episodes is

that it allows one such episode to "refer to" another. This could occur either when the referring episode of conscious thought falls entirely within the episode it is referring to (Fig. 4, left), or when the referred-to episode of conscious thought and the referring episode of conscious thought both occur within a third, enclosing conscious episode and the former occurs before the latter (Fig. 4, right). In either case, the referred-to episode might be the an ongoing experience, the recollection of an experience from the distant past (long-term memory) or the recent past (working memory), or part of an ongoing rational or creative process involving inner rehearsal. A typical referring (reflexively conscious) episode might offer some judgement on the (non-reflexively conscious) episode it is referring to, such as "that was unpleasant" or (for a reasoning process) "that hasn't got me any further".

4 Focus and Fringe

The above characterisation of a reflexively conscious thought as a conscious episode that "refers to" another conscious episode is all very well. But it leaves open many questions, including that of the mechanism by which this reference is achieved. So to flesh out our account of reflexivity, something further is required. According to the present treatment, in addition to the temporal structure described above, the flow of consciously processed information has a focus / fringe structure (Mangan, 1993; 2001). The fringe contains hints of material that has the potential to be brought into focal consciousness if required. As Mangan (2001) puts it, "The fringe creates a non-sensory feeling of imminence which implies the existence of far more than consciousness actually presents at any given moment. ... This is the fundamental trick that lets consciousness finesse its severely limited capacity ...".

The contention of this paper is that this is indeed a "fundamental trick", a means to enhance the cognitive efficacy of conscious information processing in many ways. Of especial interest here is the fact that, at any given time, while focal consciousness is contributing to one conscious episode, broadcasting information supplied by the corresponding active process, the fringe can simultaneously retain the trace of *another* co-occurring conscious episode, governed by a different active process. To see this, consider Figure 4 (right). Suppose that at time t=3 active process Z has won access to GW, and is therefore supplying the current content of focal consciousness. At the same time, although process Y is not enjoying (full) access to GW, it is still active, and can therefore influence fringe consciousness.

We have the outline, here, of mechanism by which one conscious episode can refer to another, wherein the referring episode is in focal consciousness while fringe consciousness retains a trace of the referred-to episode. But to see how this might be realised more concretely we need to zoom in and examine the evolving contents of GW at a finer timescale. In the computer model described in (Shanahan, 2006), GW was implemented as an attractor network. During execution, GW exhibited periods of stability (broadcast) during which it settled into an attractor, punctuated by periods of rapid change (competition) during which it got nudged out of a previously stable attractor and taken into a new one. During the periods of competition, it was sometimes observed that faint hints of competing attractors would become temporarily overlaid on GW's current attractor, each trying to take over.

This suggests the possibility that fringe consciousness might be realised as a rapid series of *faintly pulsing attractors*, each of which becomes transiently overlaid on the current attractor, but none of which yet has enough influence to dominate GW completely. (The dynamics here is reminiscent of Bressler & Kelso's (2001) notion of *metastability*.) Because these brief attractor pulses occur in GW, they are broadcast, and can therefore contribute to the flow of conscious information, as global workspace theory requires. Now we can appeal to *temporal synchrony*, as postulated by various authors as a solution to the binding problem (von der Malsburg, 1999), to realise reference between conscious episodes. The process currently supplying the content of focal consciousness – that is to say, the process associated with the referring episode – simply has to wait for the attractor corresponding to the process associated with the referred-to episode to pulse in

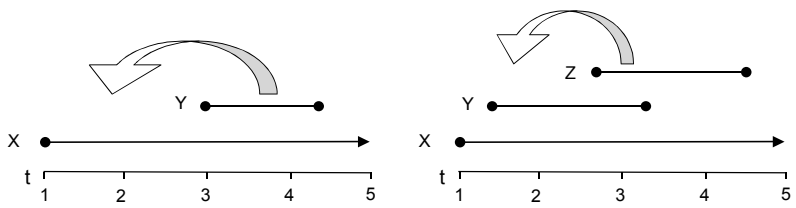


Fig. 4: Reflexively Conscious Episodes

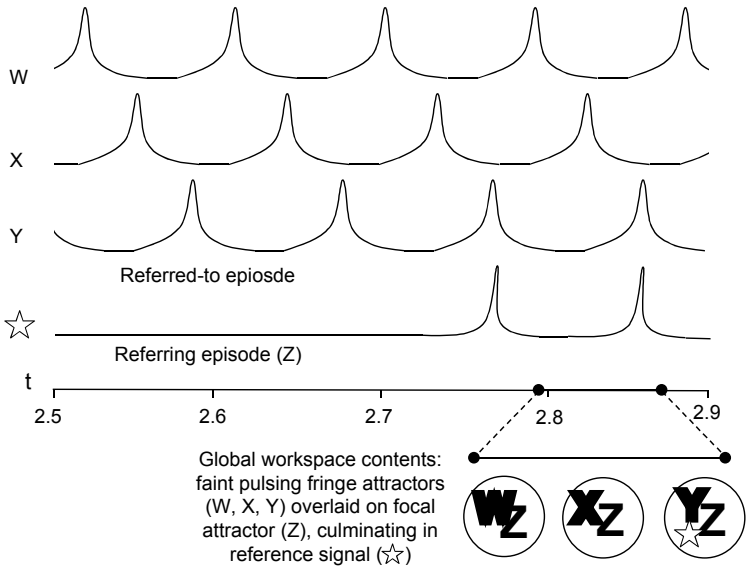


Fig.5: Focus-Fringe Reference by Temporal Synchrony

GW, when it can, so to speak, signal "THAT ONE" to GW (Fig. 5). Since this signal will be broadcast at the same time as the attractor pulse of the referred-to episode, the required reference will be secured.

5 Fringe-Borne Self-Awareness

According to the simulation hypothesis, conscious thought is simulated interaction with the environment. This entails that insofar as a conscious experience relates to anything other than an immediately present stimulus, the information processing that underpins it, as well as implicating the broadcast mechanism of the global workspace, must recruit a higher-order, internally closed sensorimotor loop (Fig. 2). This is the case for both the recall of a past conscious episode and the conscious rehearsal (or imagination) of a trajectory through sensorimotor space, where the latter conception encompasses inner speech, mental imagery, and so on.

Now, the fundamental role posited for the fringe is to augment the flow of consciously processed information with an awareness of the many possible ways that the content of the GW could unfold from its present state, without having to supply detailed information about any one of those possibilities. For example, our awareness of the three-dimensionality of a solid object can be cashed out in terms of a fringe awareness of a host of sensorimotor possibilities, such as moving around to view the back of the object, or picking it up and rotating it to see a different facet.

In the context of an internal sensorimotor loop, the fringe carries an awareness of the tree of possiblilities for conscious recall or rehearsal that branches out from the GW's current state (Fig. 6). Now suppose that, using the mechanism outlined in the previous section, one (reflexively) conscious episode Z refers to another conscious episode Y with the thought "that didn't work because P" (in the case of recall) or "that wouldn't work because P" (in the case of rehearsal). Then, thanks to the broadcast of this message, the entire set of specialist, unconscious processes will be offered the challenge of finding a potential variation of Y in which P is not the case. The vast majority of these specialists will be irrelevant to Y. But any that are successful in finding a potentially useful variation will be able to promote, via the fringe, the possibility of rehearsing it properly. This shows how reflexive consciousness can marshall massively parallel resources to further increase the cognitive power of (non-reflexive) conscious information processing, which is itself more cognitively efficacious than non-conscious information processing.

To round off the account, let's develop further the parallel between fringe-borne spatial awareness (of solid objects, for example), and the fringe-borne awareness of the unfolding content of the global workspace itself. According to the present account, the conscious awareness of the three-dimensionality of nearby objects or of the space through which the body can move consists of hints in the fringe of a systematically organised set of possible trajectories through sensorimotor space. These hints are systematically organised in the sense that they conform to various constraints, which include the reversibility of certain actions (eg: moving forwards then backwards gets you back where you started) and the cyclic character of certain trajectories (eg: turning an object

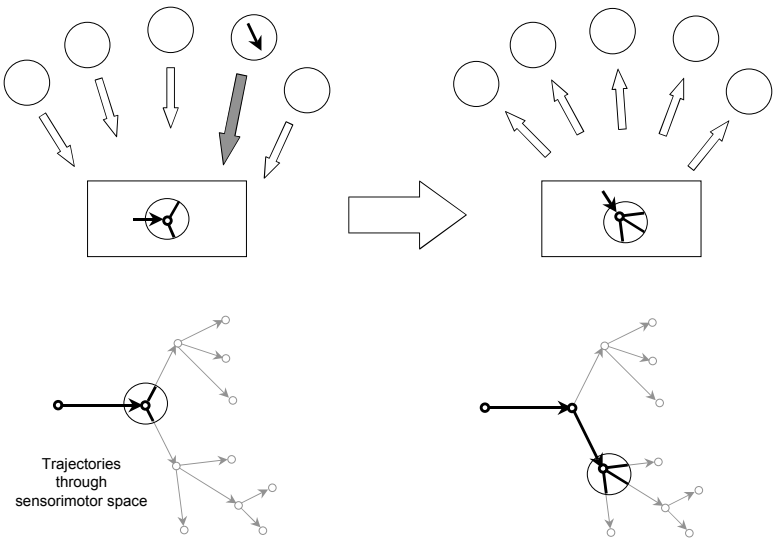


Fig. 6: Fringe-Borne Awareness of Possible Sensorimotor Trajectories

through 360º takes it back to its initial configuration).

In a similar vein, the fringe may sustain our awareness of the personhood of both ourselves and of others, hinting at material available for conscious rehearsal that pertains to our or their bodies, biographies, likes and dislikes, beliefs, desires, and intentions, skills and abilities, and so on. In the present context, the portion of this fringe-borne material of most interest relates to the way the content of the individual's consciousness unfolds. As the fringe-borne awareness of an object's solidity implies awareness of a systematic set of spatial constraints, so the fringe-borne awareness of personhood implies awareness of a systematic set of constraints on consciousness, such as its unity, its identity over time, and its indexical relationship to the body. Furthermore, in the same way that spatial constraints govern conscious thinking about solid objects, so these phenomenological constraints govern reflexively conscious thought. Insofar as we become consciously aware of ourselves as conscious beings, perhaps we do so thanks to our capacity to entertain reflexive thoughts combined with a fringe-borne awareness of the laws governing the way conscious thought unfolds.

References

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

Baars, B.J. (1997). *In the Theater of Consciousness: The Workspace of the Mind*. Oxford University Press.

Baars, B.J. (2002). The Conscious Access Hypothesis: Origins and Recent Evidence. *Trends in Cognitive Science* 6 (1), 47–52.

Bressler, S.L. & Kelso, J.A.S. (2001). Cortical Coordination Dynamics and Cognition. *Trends in Cognitive Science 5* (1), 26–36.

Cotterill, R. (1998). *Enchanted Looms: Conscious Networks in Brains and Computers*. Cambridge University Press.

Hesslow, G. (2002). Conscious Thought as Simulation of Behaviour and Perception. *Trends in Cognitive Science* 6 (6), 242–247.

Mangan, B. (1993). Taking Phenomenology Seriously: The "Fringe" and its Implications for Cognitive Research. *Consciousness and Cognition* 2 (2), 89–108.

Mangan, B. (2001). Sensation's Ghost: The Non-Sensory "Fringe" of Consciousness. *PSYCHE* 7 (18), http://psyche.cs.monash.edu.au/v7/psyche-7-18-mangan.html.

Rosenthal, D. (1986). Two Concepts of Consciousness. *Philosophical Studies* 49 (3), 329–359.

Shanahan, M.P. & Baars, B.J. (2005). Applying Global Workspace Theory to the Frame Problem. *Cognition* 98 (2), 157–176.

Shanahan, M.P. (2005). Global Access, Embodiment, and the Conscious Subject. *Journal of Consciousness Studies* 12 (12), 46–66.

Shanahan, M.P. (2006). A Cognitive Architecture that Combines Internal Simulation with a Global Workspace. *Consciousness and Cognition*, in press.

Von der Malsburg, C. (1999). The What and Why of Binding: A Modeler's Perspective. *Neuron* 25, 95–104.

How to experience the world: some not so simple ways

Aaron Sloman
School of Computer Science, University of Birmingham,
Edgbaston, Birmingham, B15 2TT, UK
A.Sloman@cs.bham.ac.uk
http://www.cs.bham.ac.uk/~axs/

Extended Abstract:

I believe the best way to extend our scientific understanding of consciousness is to stop using the noun and investigate all the many mental processes that can and do occur in humans and other animals and future robots in very great detail and explain how they are possible. Then everything of substance about consciousness will have been covered, and the vacuous, incoherent unanswered questions generated in philosophical discussions will remain unanswered as they should be, because they are unanswerable.

My talk is an illustration of a small part of this project, starting from a comment made by Wittgenstein when discussing the experience of ambiguous figures. He wrote:

> The substratum of this experience is the mastery of a technique.

I don't really know what he meant by that, but those words slightly modified thus:

> The substratum of an experience is mastery of a large collection of techniques available and ready to be deployed if required, possibly in new combinations.

could be used to express a theory I am trying to develop in the context of trying to understand how to give a robot human-like (to be more precise, child-like) capabilities in the context of perceiving and manipulating 3-D objects.

The idea is that an infant-toddler-child-youth (and future domestic robot) develops by constantly actively and creatively exploring many aspects of the environment and thereby learning a very large number (possibly many thousands, certainly many hundreds) of different facts about the environment including facts about different kinds of stuff things are made of, different kinds of surface fragments that can occur, different kinds of ways things can be combined or decomposed, different kinds of relationships that can occur between simple and complex objects, different ways collections of relations can change, different kinds of actions that can be produced, and of course different consequences of all the above.

These facts are not expressed as propositions using what we would call a human language, but they must be somehow represented internally in a usable form, and in particular, for creative experiments to be performed and novel problems to be solved by combining prior knowledge the information must be recombinable in novel ways for some uses.

So, a child or future intelligent domestic robot is constantly learning orthogonal, recombinable, competences. (Actually, not totally orthogonal since independent variation of phenomena is limited in many ways, that have to be learnt.) It seems that precocial species either cannot do this or do it to a much more limited extent: they start off with the vast majority of what they need to know about the world and how to act in it pre-programmed by evolution (contradicting familiar arguments about the requirements for 'symbol grounding'). Altricial species that develop very complex and diverse cognitive competences probably evolved these powerful information acquiring, restructuring, mechanisms because (a) genetic mechanisms lacked the space to encode them and (b) evolutionary history did not provide all the opportunities that would have been needed to derive them.

Because they evolved for dealing with a world that is not only complex, but is also constantly changing, these abilities to cope with novel processes (i.e. perceive, represent, and use information about them) at very short notice had to be implemented in architectures that made them

readily available to be invoked on demand in different combinations. I suggest that that fact determines requirements for the design and implementation of visual systems that have not yet been fully articulated. Moreover, the implementation will use mechanisms that have not yet been thought of by neuroscientists, psychologists or AI researchers.

One of the requirements for an organism that may need to monitor, evaluate, modulate and perhaps extend its own mental states and processes (e.g. improving its reasoning, problem-solving, learning, capabilities) is that it should be able to learn not only about the environment but also about its internal states. As with exploration of the environment, this could use a self-organising mechanism that adapts to what it encounters by chunking things and inventing labels for reusable chunks.

This could include labels for aspects of the contents of various sensory manifolds. Because of the manner of their development, such concepts will have a feature referred to as 'causal indexicality', i.e. their intension is intimately connected with their conditions of use. But because they are used for categorising states and processes in virtual machines that are not accessible by anyone else, these concepts will be inherently incommunicable: accounting for one aspect of what people who discuss qualia are trying to say.

When we have designed or discovered appropriate mechanisms for acquiring and using all these different competences, and the kind of architecture required to accommodate them I conjecture that this will explain a wide range of familiar phenomena including the variety of ways in which an individual can experience the world and some of the ways in which things can be experienced as ambiguous, flipping between different interpretations that make use of different competences (or 'techniques').

Some half-baked explorations of these ideas can be found in the html file referenced here

http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0601

COSY-DP-0601 Orthogonal Competences Acquired by Altricial Species (Blankets, string, and plywood)

One problem with the theory is that nothing I have learnt about brain mechanisms (on which I am no expert) seems to be capable of explaining how these competences are acquired, stored and recombined on demand.

For example, the kinds of models of neural nets that I am aware of just do not seem to be capable of meeting those requirements, though perhaps networks of networks could? Chemical information processing systems have more of the right features, but would probably be too slow, and could not easily be coupled with the processes that acquire and use the information.

It is possible that there are such mechanisms, but they have not been found because nobody was looking for them. They may be implemented in subtle ways as high level virtual machines on lower level physiological machines that seem to be doing something more mundane or something mysterious.

Note that the recombinability of orthogonal competences seems to require some sort of internal syntax. This could have been a crucial precursor to the development of external social language. It could not be based on human language because the learning and creativity I am talking about occur in prelinguistic children and some other animals.

These ideas have some echoes of global-workspace theory, though I think there are several workspaces of different sorts, supporting different kinds of concurrent processes in the architecture I envisage.

This work is partly inspired by collaboration with Jackie Chappell who studies animal cognition, especially New Caledonian Crows and Parrots/Parakeets.

Machine Consciousness and Machine Ethics

Steve Torrance
Institute for Social and Health Research
Middlesex University
Queensway, Enfield, Middlesex EN3 4SA UK
s.torrance@mdx.ac.uk

Abstract

Questions about the possibility of genuine consciousness existing in future artificial humanoids are closely tied up with ethical considerations. I discuss how the assumed presence or absence of consciousness in artificial persons might make a difference to our ethical attitudes towards them.

Questions about the possibility of genuine consciousness existing in future artificial humanoids are closely tied up with ethical considerations. How might the assumed presence or absence of consciousness in artificial persons make a difference to our ethical attitudes towards them?

For simplicity we will limit ourselves to considering electronic person simulations (EPersons) rather than organic replicates. EPersons could develop to have very rich behavioural and functional properties. Might we expect EPersons to have any ethical responsibilities, or to be subjects of ethical appraisal in any way? And could EPersons have genuine moral interests, or genuine demands on our moral concern? I will consider how the answers to these questions may vary as we consider (a) a condition where EPersons are assumed to possess a form of phenomenal consciousness (and thus can genuinely experience pleasure and suffering) and (b) a condition where they are not assumed to possess such states?

What might be our ethical obligations to such creatures in either the with-consciousness or the without-consciousness conditions? It may be neither rational nor intelligible to bestow moral concern on beings we consider to lack consciousness Conversely, if the behavioural repertoire of EPersons is sufficiently rich and varied, and they enter into a sufficiently wide range of social relations with us, it may be difficult in everyday practice to avoid perceiving or taking EPersons as making legitimate moral claims on us in at least some types of circumstance – even if they are not acknowledged as having phenomenal states.

Could we regard EPersons as having genuine moral responsibility, desert, accountability for their actions or judgments in either of the two conditions? Could they be useful moral advisors? Could they even have a coherent conception of what morality consists of? These questions in part turn on one's view on the role of emotions, and the links between emotions and rationality, in the constitution of a moral agent. It is plausible that our moral conceptions and outlook are derived from our evolutionary inheritance, and are deeply interconnected with a wide range of emotions - anger, envy, compassion, empathy, friendliness, etc.; and that these in turn emanate from biologically-based sentience in natural creatures. EPersons designed around current paradigms of information-processing cognitive architectures, may be incapable of instantiating a deep enough model of emotion and empathetic rationality to support more than a rather impoverished array of moral sentiments at best. This may be true even in the with-consciousness condition; indeed the ability to possess such emotions may be pivotal to the realizability of phenomenal consciousness in EPersons.

I will argue that phenomenal consciousness makes a difference in the cases both of being a genuine bearer of moral responsibilities and of being a worthy recipient of moral treatment. Having fully-fledged or intact phenomenally conscious states is not the sole criterion of moral worth (think of neonates, PVS, dementia); nevertheless the generic property of being the kind of creature that can have phenomenal states is arguably crucial to being a member of the universe of moral concern.

In general we take artefacts to be instruments; so there seems to be an inherently paradoxical quality wrapped up in the idea of extending morality to artificial humanoid beings. However new ways of thinking about moral relationships may forced on us in an era where artificial humanoids live alongside us in significant proportions. A new conception of 'us' may be required.

Motor Development

5th April 2006

Organiser

Luc Berthouze, AIST Neuroscience Research Institute, Japan

Programme Committee

Christian Balkenius, Lund University
Luc Berthouze, AIST Neuroscience Research Institute
Yiannis Demiris, Imperial College London
Eugene Goldfield, Boston Children's Hospital
Brian Hopkins, Lancaster University
Giorgio Metta, Genoa University
Claes Von Hofsten, Uppsala University

Contents

Robot bouncing: The assembly, tuning, and transfer of action systems*

Luc Berthouze*
*Neuroscience Research Institute (AIST)
Tsukuba AIST Central 2, Umezono 1-1-1
Tsukuba 305-8568, Japan
Luc.Berthouze@aist.go.jp

Abstract

The early exploratory behaviours of infants include many rhythmical stereotypies (Thelen, 1979). Their study should therefore yield insights on the mechanisms underlying motor development. Goldfield et al. (1993) observed young infants learning to bounce in a Jolly Jumper. The longitudinal profile of the learning process revealed two developmental stages – the assembly phase and the tuning phase – that may be typical of infants' acquisition of new motor skills. To gain a mechanistic view of those stages, we replicated the study by using a small humanoid robot suspended to a fixed frame by rubber springs (Lungarella and Berthouze, 2004). Since compliance is a key feature of infants' musculoskeletal system, the robot was equipped with passively compliant leg joints (viscoelastic material mounted in series with the actuators) and feet (Meyer et al., in press). Sensory feedback was provided by force sensing resistors (FSR) placed underneath each foot. We interpreted the assembly phase in terms of the self-organization of elementary functional units, i.e., the formation of muscle synergies, and the tuning phase in terms of the adjusting of their respective time constants. The functional units were modeled as Bonhoeffer-Van der Pol (BVP) oscillators (Fitzhugh, 1961), a reduction of the 4-variable Hodgkin-Huxley model to a simpler algebraic form with two variables (excitable and recovery variables). These oscillators display both goal directedness (the ability to reach the desired endpoint), and sensitivity to environmental and mechanical input. They exhibit robust phase locking even in the presence of large delays in the feedback loop, a characteristic that is important given the compliance of the system. Both assembly and tuning phases were embedded in an exploration of the parameter space modelled as a biased random search with a value system based on two qualities of the behaviour: bouncing height and stability. Experiments revealed a longitudinal profile qualitatively similar to that reported by Goldfield et al. The transfer of the newly assembled action system, i.e., its ability to adapt to physical or environmental changes, was verified by changing the lift-off weight of the robot during bouncing. The fact that each perturbation was followed by rapid recovery without reconfiguration suggests that the learned parameters were task-specific.

References

R. Fitzhugh. Impulses and physiological states in theoretical models of nerve membrane. *Biophysical Journal*, 1:445–466, 1961.

P. Foo, E.C. Goldfield, B.A. Kay, and W.H. Warren. Infant bouncing: The assembly, tuning, and transfer of action systems. In *Proceedings of the 9th International Congress on Research in Physical Activity and Sport*, 2001.

E.C. Goldfield, B.A. Kay, and W.H. Warren. Infant bouncing: the assembly and tuning of action systems. *Child Development*, 64:1128–42, 1993.

M. Lungarella and L. Berthouze. Robot bouncing: On the synergy between neural and body-environment dynamics. In *Embodied Artificial Intelligence*, pages 86–97. Springer-Verlag Berlin Heidelberg, 2004.

F. Meyer, A. Spröwitz, and L. Berthouze. Passive compliance for an RC servo-controlled bouncing robot. *Advanced Robotics*, in press.

E. Thelen. Rhythmical stereotypies in normal human infants. *Animal Behavior*, 27:699–715, 1979.

*Title inspired from Foo et al. (2001).

Active learning of probabilistic forward models in visuo-motor development

Anthony Dearden and Yiannis Demiris
Department of Electrical and Electronic Engineering
Imperial College London
Exhibition Road, London, SW7 2BT
E-mail: {anthony.dearden, y.demiris}@imperial.ac.uk

Abstract

Forward models enable both robots and humans to predict the sensory consequences of their motor actions. To learn its own forward models a robot needs to experiment with its own motor system, in the same way that human infants need to babble as a part of their motor development. In this paper we investigate how this babbling with the motor system can be influenced by the forward models' own knowledge of their predictive ability. By spending more time babbling in regions of motor space that require more accuracy in the forward model, the learning time can be reduced. The key to guiding this exploration is the use of probabilistic forward models, which are capable of learning and predicting not just the sensory consequence of a motor command, but also an estimate of how accurate this prediction is. An experiment was carried out to test this theory on a robotic pan tilt camera.

1 Introduction

Forward models enable both robots and humans to predict the sensory consequences of their motor actions [Jordan and Rumelhart, 1992, Wolpert and Flanagan, 2001]. This is extremely useful for robotics as it allows the robot to simulate the effects of its actions internally before physically executing them. Being able to simulate multiple possible actions allows the robot to choose the most appropriate command for a particular task, for example imitation [Demiris and Johnson, 2003]. Practically any environment a robot operates in will change, or have properties which cannot be modelled beforehand. Even if the environment is assumed to be completely predictable, endowing the robot with this knowledge may be beyond the ability or desire of its programmer. A truly autonomous robot, therefore, needs to be able to learn and adapt its own forward models.

The idea of learning a model of an unknown system is explored extensively in the field of system identification [Ljung, 1987]. In system identification, the task of choosing experiments and interventions to perform on the unknown system is the role of the human designing the control system. Here, however, we want this process to be automated - the robot should essentially design its own experiments. The idea of a robot as a scientist provides some interesting analogies with the theories of learning in human infants presented by Gopnik [Gopnik et al., 2004], who uses Bayesian networks to model how infants actively form and test causal models of the world. Meltzoff discussed 'body babbling' as a method used by human infants to learn and adapt control of their motor system [Meltzoff and Moore, 1997].

By using as little prior information as possible, we want the robot to learn about its own motor system. This knowledge it gains about its motor system is stored in the form of a forward model. Previous work on learning forward models has looked at how a robot can develop an internal representation of the state of the world with information from it's vision system [Dearden and Demiris, 2005]. The forward model for predicting the state was learnt and represented probabilistically using a Bayesian network. The training data was provided by random babbling of motor commands to produce the corresponding set of sensor data to train the model. The work here expands on this by allowing the exploration, or babbling, of the motor system to be driven by the estimated prediction accuracy of multiple competing forward models. By spending more time babbling with motor commands which the forward models are worse at predicting, the forward models can be more rapidly learnt and used.

Active exploration of the environment by a robot to learn or adapt models has been attempted previously, in [Lipson and Bongard, 2004]. Using multiple inter-

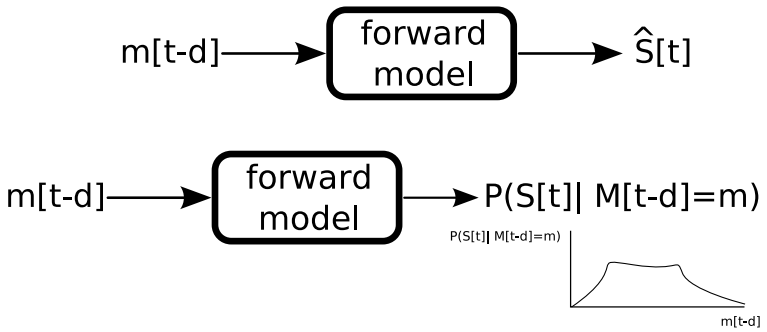


Figure 1: *A forward model and a probabilistic forward model for motor command* m[t]*, sensor prediction,* $\hat{S}$*, and motor delay d.*

nal models generated and adapted using a genetic algorithm, their exploration-estimation algorithm uses a two phase process of choosing motor commands to best discriminate between potential models. The exploration is not driven by the prediction error as in this paper, but by choosing interventions which will maximally differentiate between the different internal models. The idea of 'adaptive curiosity' is used in [Oudeyer et al., 2005] to guide a robot to learn how to interact with its environment. The robot is made to focus on situations that are progressively harder for it to predict.

2 An architecture for learning and representing forward models

The system proposed here for learning the model of a robot's motor system is based on using multiple probabilistic forward model 'primitives'. Active learning is used to decide how motor commands should be chosen by each individual forward model primitive, and selected from the multiple possible commands requested by the forward models.

2.1 Why *probabilistic* forward models?

All forward models are wrong, but some are useful[1]. A forward model will not be able to completely accurately model a robot's motor system - errors will occur in predictions because of insufficient or noisy training data or the necessarily simplified internal representations of the model. The system which is being modelled may itself be stochastic. To overcome this inaccuracy, it makes sense for a forward model to include information regarding not just its prediction, but how accurate it expects that prediction to be. This inaccuracy can be modelled by having the forward model learn not just a prediction for a given motor command and state, but a joint probability distribution across the inputs to the forward model and its predicted outputs. The output of a probabilistic forward model is thus a conditional probability distribution for a particular motor command, *m*, and state, *s*, at time t : $P\left(S\left[t\right] \mid M\left[t-d\right]=m\right)$, as shown in Figure 1. The other parameter, *d*, is used to model the delay in the motor system - in any real system, there will be a delay between a motor command being executed and its effects being measured at its sensors. For a forward model to be useful in this situation it must model (and learn) this delay.

The advantage of a probabilistic representation of prediction is that, instead of predicting a specific outcome, the prediction will be of a range of possible outcomes, each weighted with a particular likelihood. The forward model essentially has knowledge about its own ability to predict. Any control system using the forward model will receive not just one prediction, but a probability distribution. This provides the control system with more information about the predicted consequences of its actions. This extra information is also useful for guiding the motor control during the learning process. The disadvantage of using a probabilistic representation is that more training data may be required. This is not as much of a disadvantage as it would be in a typical machine learning situation because the data set is not limited - the robot has active control over the system it is trying to model, so can easily acquire training data.

To overcome the trade-off between the complexity of the modelled conditional probability distribution and the amount of training data - and therefore time - required to train it, the normal distribution was

[1] A modification of a quote attributed to George EP Box

used. The forward model therefore needs to learn and represent two functions: $\hat{S}(m)$ - the estimated mean of the sensor value as a function of the motor input(s) $\hat{\sigma}_S(m)$, the estimated standard deviation. The output distribution as a function of the motor command, *m*, is therefore $P(S[t] \mid M[t-d]=m) \sim N\left(\hat{S}(m), \hat{\sigma}(m)\right)$. Both these functions can be estimated with any appropriate function approximator that can be learnt online. In the experiments here, both radial basis functions and conditional probability tables were used.

2.2 Why *multiple* forward models?

The idea of using multiple forward models has been used in both robotics for imitation [Demiris, 2002], and in neuroscience to model motor skill learning in humans [Wolpert et al., 2003]. In these architectures, the multiple forward models are used together with inverse models to achieve higher level control. In this work, however, we are just interested in learning the forward model that can be used by these systems.

Using multiple primitive forward models to model a system is similar to the mixture of experts idea introduced by [Jacobs et al., 1991]. As the forward models are probabilistic and represent causal connection between the random variables for motor command, *M*, and predicted output, *S*, the forward models make up a Bayesian network [Pearl, 1988]. The forward models are the conditional probability distributions connecting random variables. Splitting the forward model into a distributed system using multiple, simpler forward models has numerous advantages over using a single forward model

- The learnt structure represents causal structure of the robot's motor system. This means the learning process requires less data (and is therefore faster) because unnecessary connections between motors and sensors are not learnt. The robot also has an internal representation of the higher level causal structure of its motors system.
- Robots have different kinds of motor commands and sensors (e.g. discrete or continuous). The appropriate internal representation for the forward model may be different depending on the nature of these. Using multiple forward models allows several different types of function approximators to be used simultaneously.

Figure 2 shows a comparison between a single and multiple primitive forward models.

2.3 Active learning and babbling

In a typical machine learning situation, it is assumed that a set of data representing samples from an underlying function or probability distribution is available. The task is to learn a function or distribution which approximates this distribution. The situation with a robot is different in two ways. Firstly, the process is performed online as opposed to in a batch - data is continuously received and the learnt forward models should be continuously adapted. Secondly, and most importantly, the robot has active control over the inputs it can send to its as yet unknown motor system. The situation where the learner has the ability to select some of the data is referred to as active learning [Hasenjager and Ritter, 2002, Tong and Koller, 2000]. The principal benefit of this is that the data can be selected either to speed up the learning process, or to optimise the learnt model to be most useful for a particular task. For example a robot could concentrate on learning particular forward models that would be needed to imitate a specific task.

The use of multiple competing forward models fits well into the concept of active learning, as each primitive forward model can now compete not just to offer the best prediction, but also to get control over the motor system to provide itself with training data. This does, however, complicate the situation somewhat. As well as the problem of how each forward model chooses a motor command or set of motor commands to be sent to the motor system, there is the important issue of how to choose which of the forward models should be given control of the motor system at any particular time.

This problem has many similarities to attention mechanisms studied in robotics [Khadhouri and Demiris, 2005, Demiris and Khadhouri, to appear], which investigate the allocation of processing resources. In contrast, the task here is to control the allocation of *motor* resources. In this paper the approach taken to guide the babbling is to allow each forward model to suggest a particular motor degree of freedom and value to babble with. The probability of a particular motor command being chosen by a forward model is proportional to the estimated standard deviation of the forward model in that region of motor space, $\hat{\sigma}_S(m)$. Therefore, the forward model is more likely to pick a motor command that it estimates has high prediciton error. Several motor commands will be requested simultaneously, one for each forward model. The learning system currently chooses a forward model at random to ensure that each forward model is given the opportunity to control the motor system.

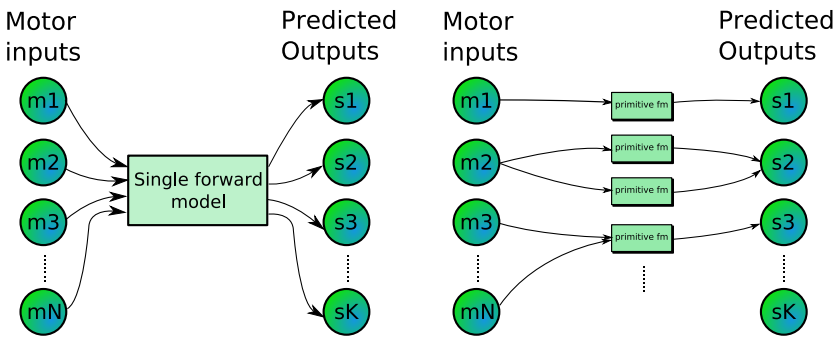


Figure 2: *Using a single forward model (a) or multiple primitive forward models (b)*

3 Stages in learning the forward model

The learning of the forward model needs to be divided into distinct phases. This simplifies learning a complex forward model by learning different aspects of the model's structure or parameters sequentially. The developmental stages represent increasing complexity in the learning and adaptation of forward models, from establishing a causal connection, calculating the time delay, and finally adapting more and more precisely to the causal relationship. The developmental stages used to learn forward models are as follows:

1. Observe and learn a steady state model of the sensors

   In this first stage of learning, the robot does not actually interact with the environment - it simply learns the statistics of the sensor data $P(S)$ as a normal distribution. This an important preliminary stage to learning any forward model because the robot cannot model how its different motor commands are influencing particular sensors until it has modelled how its sensors behave without any intervention.

2. Try impulse commands to learn time delay, and basic causal structure of the network

   In previous work, the time delay in the motor system was learnt by simultaneously learning multiple forward models with different time delays. The correct time delay was found from the forward model which could best predict the data [Dearden and Demiris, 2005]. Here the time delay is estimated directly by using the learnt models of the sensors. Impulse motor commands are issued to the motor system at time *T*, one degree of freedom at a time. The likelihood of the incoming sensor data, $s[T+t]$ given the sensor model learnt in step one is calculated - i.e. $P(S = s[T+t])$. If this likelihood falls to a low value then it is likely that this motor degree of freedom is influencing this sensor, and that the delay for the influence to occur is *t* discrete time-steps; the threshold likelihood used in the experiments here was 0.001. Thus not only can the motor delay be learnt, but some initial information about the causal structure of the forward model is learnt - if a motor command does not reduce the likelihood of a sensor model, it is unlikely it can influence it, and therefore this relationship does not have to be modelled.

3. Completely random babbling to learn the range of values for the sensor data

   Function approximators generally need the data to be scaled within a set range, e.g. [0,1]. When sensor data is being received online, and no prior information about it is available, this cannot be done. Therefore, a stage of experimenting with extremes of motor commands to find the extremes in the range of sensor data is necessary. Once this stage of adaptation is complete, the sensor data can be scaled between the calculated minimum and maximum values.

4. Learn steady state model between motor commands and sensors, using guided babbling

   The guided babbling in this stage happens as described in section 2.3. Because we are currently only interested in learning steady state models, learning is paused from the issuing of a motor until the sensor system has reached a steady state.

4 Experiment & results

The experiments here were carried out using the pan-tilt unit on an Activmedia Peoplebot[2]. The sensor data used were the properties of the most salient coloured object in the scene - its position, width, height and angle of rotation. The object is located from the thresholded camera image in hue space, and tracked using the Camshift algorithm [Bradski, 1998]. The first and second stage of the learning process identified the delay for both the pan and tilt motor commands to be 5 time-steps, or 333ms. As shown in Figure 3 , it also learnt that, whilst the co-

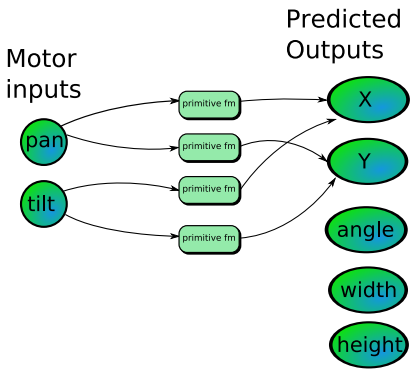


Figure 3: *The primitive forward model structures learnt from sending impulse commands to the motor system.*

ordinates of the object in the scene were affected by the pan and tilt commands, the size and angle of the object were not affected. By learning this causal relationship early on, the robot has thus reduced the number of models it has to initially learn from ten to four.

The evolution of the prediction errors, from a typical experiment, for the four forward models using guided babbling are shown in Figure 4a . The prediction error is as the sum of the variance estimation over all motor commands, $\int \hat{\sigma_s}(m)\, dm$. This can be compared with the evolution of the prediction errors when random motor commands are chosen, as shown in Figure 4b. Converging to accurate models takes significantly longer in this case.

If the camera had been mounted perfectly straight then the *pan* motor command would have no effect on the y-coordinate of the object, and similarly for the *tilt* command and the x-coordinate. However, since the camera is at a slight angle, there is a slight dependence between these. It is interesting to note that whilst the forward models linking the *pan* command to the *y*-coordinate and the *tilt* command to the *x*-coordinate do converge to a particular model, they are much less accurate at predicting than the *pan*-to-*x*-coordinate and *tilt*-to-*y*-coordinate forward models, as one would expect.

Part of the evolution of the *pan*-to-*x*-coordinate forward model's mean prediction, $\hat{S}(m)$, is shown in Figure 5. As expected, the model learnt is a linear one - the position of the object, *X*, is proportional the the *pan* command. Of particular interest is the prediction of a low valued motor command, as shown by the bold line. Because the estimated error in prediction is initially high for this motor command, more time is spent babbling in this region, and hence it converges to a more accurate model.

5 Conclusions and future work

In this paper, we investigated how the learning of forward models for a robot could be made faster by allowing the forward models to guide the exploration of the motor space with guided babbling. The results show that accurate models can be learnt more quickly if the errors in the predictions of the forward models are used to guide which region of motor space is explored. Future work will involve investigating this idea further, by looking at how the motor requests from each individual model should be allocated. Important factors in this decision include:

- How to cope with many more degrees of freedom
- How well a forward model is predicting
- How much data the forward model has previously been allowed
- What is the goal of the babbling - to learn a model as fast as possible or as accurate as possible for a particular task?
- How many primitive forward models want access to the same region of motor space

We are also investigating how the primitive forward models can be improved to represent and adapt to dynamic environments by adding another stage to the learning process.

References

G. Bradski. Real time face and object tracking as a component of a perceptual user interface. In *4th*

[2]http://www.activmedia.com

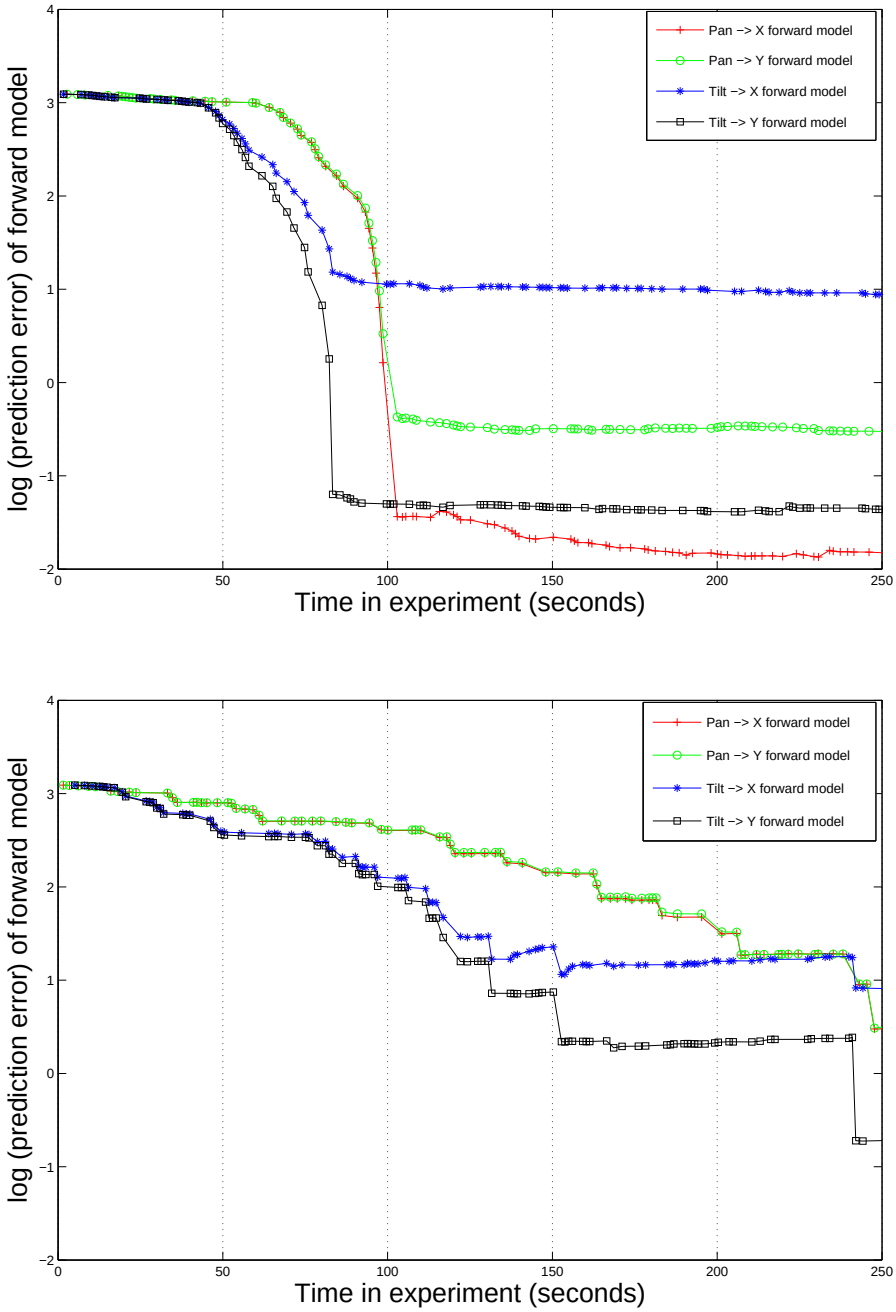


Figure 4: *The evolution of the prediction errors of the four of the forward models created when using guided babbling (a) and random babbling (b).*

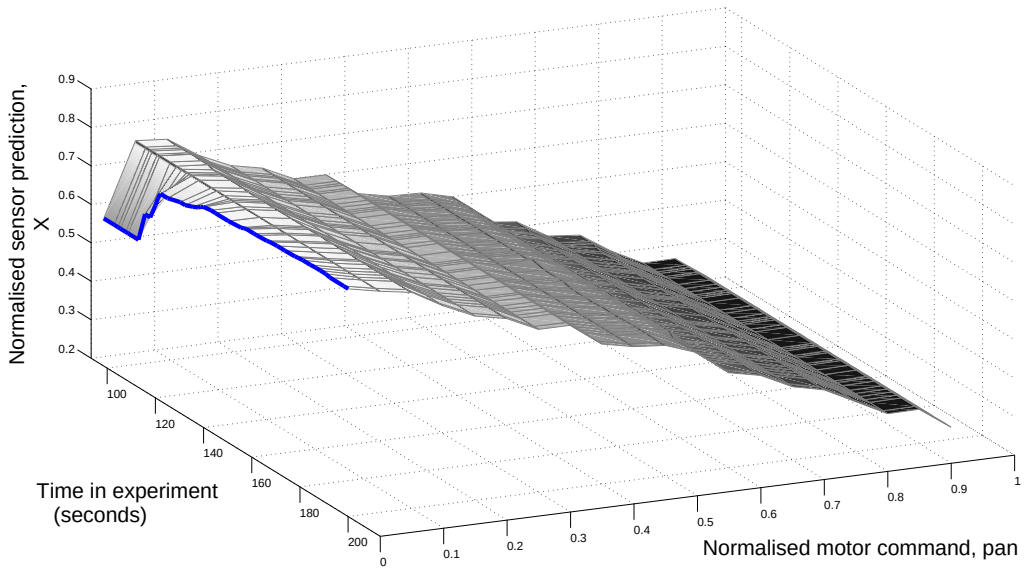


Figure 5: *Evolution of the pan -> x forward model during the experiment. The bold line shows the evolution of the motor command prediction where there is high error. Because more time is spent babbling with this motor command, the prediction converges quickly to a more correct one.*

IEEE Workshop on Applications of Computer Vision (WACV'98), page 214. IEEE, 1998.

A. Dearden and Y. Demiris. Learning Forward Models for Robotics. In *Proceedings of IJCAI 2005*, pages 1440–1445, 2005.

Y. Demiris. Imitation, Mirror Neurons, and the Learning of Movement Sequences. In *Proceedings of the International Conference on Neural Information Processing (ICONIP-2002)*, pages 111–115. IEEE Press, 2002.

Y. Demiris and M. Johnson. Distributed, predictive perception of actions: a biologically inspired robotics architecture for imitation and learning. *Connection Science*, 15(4), 12 2003.

Y. Demiris and B. Khadhouri. Hierarchical attentive multiple models for execution and recognition (hammer). *Robotics and Autonomous Systems Journal*, to appear.

A. Gopnik, C. Glymour, D. Sobel, and D. Danks. A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review*, pages 1–30, 1 2004.

M. Hasenjager and H. Ritter. Active learning in neural networks. *New learning paradigms in soft computing*, pages 137–169, 2002.

R. Jacobs, M. Jordan, S. Nowlan, and G. Hinton. Adaptive Mixtures of Local Experts. *Neural Computation*, 3:79–87, 1991.

M. Jordan and D. E. Rumelhart. Forward models: Supervised learning with a distal teacher. *Cognitive Science*, (16):307–354, 1992.

B. Khadhouri and Y. Demiris. Compound effects of top-down and bottom-up influences on visual attention during action recognition. In *Proceedings of IJCAI-2005*, pages 1458–1463, 2005.

H. Lipson and J. Bongard. An exploration-estimation algorithm for synthesis and analysis of engineering systems using minimal physical testing. In *Proceedings of DETC04*, 2004.

L. Ljung. *System Idenetification, Theory for the user*. Prentice Hall, 1987.

A. N. Meltzoff and M. K. Moore. Explaining facial imitation: A thoretical model. *Early Development and Parenting*, (6):179–192, 6 1997.

P. Oudeyer, F. Kaplan, V. Hafner, and A. Whyte. The playground experiment: Task-independent development of a curious robot. In *proceedings of the*

AAAI Spring Symposium Workshop on Developmental Robotics, 2005.

J. Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, 1988.

S. Tong and D. Koller. Active learning for parameter estimation in Bayesian networks. *Advances in Neural Information Processing Systems (NIPS)*, 12 2000.

D. M. Wolpert and J. R. Flanagan. Motor Prediction. *Current Biology*, 11(18), 2001.

D. M. Wolpert, K. Doya, and M. Kawato. A unifying computational framework for motor control and social interaction. *Phil. Trans. of the Royal Society of London B*, 358:593–602, 2003.

A Procedural Learning Mechanism for Novel Skill Acquisition

Sidney D'Mello, Uma Ramamurthy, Aregahegn Negatu, & Stan Franklin
Department of Computer Science & Institute for Intelligent Systems
University of Memphis, Memphis, Tennessee 38152, USA
{sdmello|franklin|urmmrthy|asnegatu}@memphis.edu

Abstract

In this paper we attempt to develop mechanisms for procedural memory and procedural learning for cognitive robots on the basis of what is known about the same facilities in humans and animals. The learning mechanism will provide agents with the ability to learn new actions and action sequences with which to accomplish novel tasks.

1 The LIDA Model

The Learning Intelligent Distribution Agent (LIDA) architecture provides a conceptual and computational model of cognition. She is the partially conceptual, learning extension, of the original IDA system implemented computationally as a software agent. IDA 'lives' on a computer system with connections to the Internet and various databases, and does personnel work for the US Navy, performing all the specific personnel tasks of a human (Franklin, 2001).

The major components of the LIDA architecture are perceptual associative memory, working memory, episodic memory, functional consciousness, procedural memory, action selection, and sensory-motor memory, with the last three being of interest to this paper. LIDA's mechanisms for procedural memory, action selection, and action realization (execution) are inspired by variants of models originally conceived by Drescher's schema mechanism (1991), Maes' behavior network (1989), and Brooks' subsumption architecture (1986) respectively.

Procedural memory in LIDA is a modified and simplified form of Drescher's schema mechanism (1991), the scheme net. The scheme net is a directed graph whose nodes are (action) schemes and whose links represent the 'derived from' relation. Built-in primitive (empty) schemes directly controlling effectors are analogous to motor cell assemblies controlling muscle groups in humans. A scheme consists of an action, together with its context and its result. The context and results of the schemes are represented by perceptual symbols (Barsalou, 1999) for objects, categories, and relations in perceptual associative memory (not described here). The perceptual symbols are grounded in the real world by their ultimate connections to various primitive feature detectors having their receptive fields among the sensory receptors. The action of a scheme is connected to an appropriate network in sensory-motor memory (described later) that directly controls actuators.

Each scheme also maintains two statistics, a *base-level activation* and a *current activation*. The base-level activation (used for learning) is a measure of the scheme's overall reliability in the past. It estimates the likelihood of the result of the scheme occurring by taking the action given its context. The current activation is a measure of the relevance of the scheme to the current situation (environmental conditions, goals, etc,). At the periphery of the scheme net lie empty schemes (schemes with a simple action, but no context or results), while more complex schemes consisting of actions and action sequences are discovered as one moves inwards.

The LIDA architecture employs an enhancement of Maes' behavior net (1989) for high-level action selection in the service of feelings and emotions. The behavior net is a digraph (directed graph) composed of behaviors codelets (a single action), behaviors (multiple behavior codelets operating in parallel), and behavior streams (multiple behaviors operating in an ordered sequence) and their various links. These three entities all share the same representation in procedural memory (i.e., a scheme).

Once an action has been selected, it triggers a suitable sub-network of the sensory-motor memory, modeled after Brook's subsumption architecture (Brooks, 1986). With sensors directly driving effectors, this sub-network effects the selected action.

2 Procedural Learning

Our model of procedural learning is based on functional consciousness, implemented in adherence to Global Workspace Theory (Baars, 1988), and reinforcement learning. Reinforcement is provided via a sigmoid function such that initial reinforcement becomes very rapid but tends to saturate. The inverse of this same sigmoid function serves as the decay curve. Therefore, schemes with low base level activation decay rapidly, while schemes with high (saturated) base level activation values tend to decay at a much lower rate.

For learning to proceed initially, the behavior network must first select the instantiation of an empty scheme for execution. Before executing its action, the instantiated scheme (activated behavior codelet) spawns a new expectation codelet (a codelet that tries to bring the results of an action to consciousness). After the action is executed, this newly created expectation codelet focuses on changes in the environment as a result of the action being executed, and attempts to bring this information to consciousness. If successful, a new scheme is created, if needed. If one already exists, it is appropriately reinforced. Conscious information just before the action was executed becomes the context of this new scheme. Information brought to consciousness right after the action is used as the result of the scheme. The scheme is provided with some base-level activation, and it is connected to its parent empty scheme with a link.

Collections of behavior codelets that operate in parallel form behaviors. The behavior codelets making up a behavior share preconditions and post conditions. Certain attention codelets (codelets that form coalitions with other codelets to compete for consciousness) notice behavior codelets that take actions at approximately the same time, though in different cognitive cycles (a cyclical process beginning with perception and ending in an action). These attention codelets attempt to bring this information to consciousness. If successful, a new scheme is created, if it does not already exist. If it does exist, the existing scheme is simply reinforced, that is, its base-level activation is modified. If a new scheme has to be created, its context is taken to be the union of the contexts of the schemes firing together. The result of the new scheme is the union of the results of the individual schemes. Additionally, this new scheme is provided with some base-level activation and is connected by links to the original schemes it includes. If this composite scheme executes in the future it will pass activation along these links.

Collections of behaviors, called behavior streams can be thought of as partial plans of actions. The execution of a behavior in a stream is conditional on the execution of its predecessor and it directly influences the execution of its successor. When an attention codelet notices two behavior codelets executing in order within some small time span, it attempts to bring this information to consciousness. If successful, it builds a new scheme with links from the first scheme to the second, if such a scheme does not already exist, in which case the existing scheme is simply reinforced. If a new scheme has to be created, its context is the union of the contexts of the first scheme and the second, excluding the items that get negated by the result of the first. Similarly the result of the new scheme formed will be the union of both results, excluding the results of the first that are negated by the results of the second. Using such a learning mechanism iteratively, more complex streams can be built.

3 Discussion

With the continually active, incremental, procedural learning mechanism an autonomous agent will be capable of learning new ways to accomplish new tasks by creating new actions and action sequences. Although our model of procedural learning is motivated by Drescher's schema model (1991), the learning mechanism is different in two significant aspects. First, our approach maintains that functional conscious involvement is a necessary condition for supraliminal learning. The second distinction arises from the fact that while learning in Drescher's system relies on each schema maintaining several reliability statistics, we only use a single, computationally more tractable statistic, the base-level activation modeled by a saturating sigmoid function.

References

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

Barsalou, L. W. 1999. Perceptual symbol systems. *Behavioral and Brain Sciences* 22:577–609.

Brooks, R. A. (1986). A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation.*, RA-2, April, 14-23

Drescher, G. 1991. *Made Up Minds: A Constructivist Approach to Artificial Intelligence*, Cambridge, MA: MIT Press.

Franklin, S. 2001. Automating Human Information Agents. *Practical Applications of Intelligent Agents*, ed. Z. Chen, and L. C. Jain. Berlin: Springer-Verlag.

Maes, P. 1989. How to do the right thing. *Connection Science* 1:291-323.

Is a kinematics model a prerequisite to robot imitation?

Bart Jansen*
*Department of Electrical Engineering and Informatics
Vrije Universiteit Brussel
Pleinlaan 2, 1050 Brussel, Belgium
bjansen@(nospam)etro.vub.ac.be

Abstract

In robot imitation it is considered natural that a kinematics model is a prerequisite to imitation. This is supported by the simple observation that imitation is not possible if the robot has no accurate control over its embodiment. In this document we argue that this reasoning can easily be inverted: by imitating each other, the robots learn an increasingly more accurate model of the kinematics of their effectors. Meanwhile, the quality of their imitative attempts improves. After a while, no qualitative difference is experienced between both approaches.

1 Introduction

In robot imitation it is investigated how an observer (a robot) can learn to perform a task simply by observing a demonstrator (either a human or a robot). The skill which is to be transferred can vary over a wide range of complexity: meaningless hand motions (Jansen et al., 2003), object manipulations (Kuniyoshi et al., 1994; Billard and Matarić, 2001; Demiris and Hayes, 1996; Alissandrakis et al., 2002) and even intentional behaviour (Calinon et al., 2005; Lockerd and Breazeal, 2004; Billard et al., 2004; Crabbe and Dyer, 2000; Jansen and Belpaeme, 2005) can be imitated. Learning from observation is important as it yields the promise of extremely easily programmable robots.

Consider the scenario in which a demonstrator performs a sequence of block manipulations (picking up a block, stacking a block on another and the like). The imitating robot must be aware of its forward (what will be the position of the hand after performing a given set of motor commands) and inverse kinematics (which motor commands are required to position my hand in a given configuration) in order to be able to imitate such behaviour. Models of forward and inverse kinematics are often highly complex. In several cases kinematics models can simply not be calculated analytically.

A wide range of work exists on the learning of kinematic models for robots. See among others (Jordan and Rumelhart, 1992; Kawato, 1990; Versino and Gambardella, 1995). The robot typically engages in a kind of motor babbling phase in which several random movements are performed and their outcome is measured somehow.

Besides its importance for the building of robots which can be controlled accurately and intuitively, the study of motor babbling with robots is also important in cognitive sciences and developmental psychology. A vast amount of work exists on the building of computer simulations of human kinematics learning (Oztop et al., 2004; Lee and Meng, 2005). Such models might provide predictions which can be investigated in the human counterpart of the simulation.

Sporns and Edelman (1993) proposed three stages in the process of animal sensory-motor learning: Generation of random movements (1), the detection of effects of the movements and the recognition of their adaptive value (2), and the ability to select movements based on their adaptive value (3). In their computational model of grasp learning Oztop et al. (2004) propose that the "joy of grasping" might be such an adaptive value: We propose "that the sensory feedback arising from the stable grasp of an object, "joy of grasping", is a uniquely positive, motivating reward for the infant to explore and learn actions that lead to grasp-like experiences." (Sporns and Edelman, 1993)[p4].

All systems that are built, either computer simulations or robot implementations, have in common that the motor babbling phase precedes the task which the robots are actually designed for, simply because accurate models are required for performing the task accurately. In this paper we investigate whether a simple imitation task can be used as a mechanism for the learning of kinematics models. This is not trivial since both processes rely on each other: without accurate kinematics models, accurate imitation seems not

to be possible. But also the imitation task influences the kinematics learning process: the agents will learn the kinematics models based on the actions they try to imitate. The variation in those actions influences the quality of the kinematics models.

2 Learning kinematics models

We consider a small population of robotic agents. Each robot is equipped with an arm with N degrees of freedom in which all joints have equal lengths. Locally weighted learning (LWL) (Atkeson et al., 1997a,b) is used as a method for learning models of both forward and inverse kinematics. Locally weighted learning is very simple and assumes that the mapping of actions on observations is linear, in any very small region. Suppose a training set consists of N actions A and corresponding observations O. Further suppose that we want to estimate the observation o_q corresponding to a query action a_q. We can then calculate a weight matrix $W_{ii} = w_i$ with $w_i = \sqrt{K(d(a_i, a_q))}$, where $K(x) = 1/x$. By using the weight vector, points which are located further from the query point are considered less important for calculating the result. Since linearity is assumed on small patches, the distance measures on actions can be the euclidean distance on the action components. Assuming a linear mapping we have $Ax = O$. Take $A' = WA$ and $O' = WO$, thus $A'x = O'$. We then have $o_q = a_q A'^{+} O'$. For stability reasons, the pseudo-inverse A'^{+} is calculated using Singular value decomposition.

Using LWL the robot can learn a model of its kinematics by performing random actions and observing their outcome. The action and its observation can simple be added to the set of known associations. For every new query, its K nearest neighbours are retrieved form the action space, together with the corresponding observations. Only those K associations are used in the LWL algorithm.

Finding nearest neighbours computationally efficient is done using the projection method (Friedman et al., 1975). All data is kept sorted in all dimensions. For every query, a heuristic function indicates the best dimension to find neighbours in.

3 Imitation task

As an imitation task, we use a computer simulation of the learning of simple actions by a robot. The paradigm of "imitation games" is used. The paradigm differs from most others in that it considers a population of agents rather than a single demonstrator and a single imitator. Rather than transferring all knowledge of the demonstrator to the imitator, all agents in the population can act both as a teacher and as a student. As a result, all skills of all agents spread into the population. Moreover, the agents are not created with some set of preprogrammed skills, all skills are invented and transferred during the course of the imitation games. This results in a set of self-organising repertoires of behaviours which are shared among all agents in the population. The repertoires of all agents will converge towards a set of skills which can be imitated easily by all agents in the population.

The paradigm of imitation games was introduced in computer simulations of the learning of human vowel systems (de Boer, 1999, 2000). Later on, the paradigm was used both in the study of imitation of actions (Jansen, 2003; Jansen et al., 2004) and goals (Jansen and Belpaeme, 2005; Jansen, 2005b).

The concept of imitation games requires multiple interactions between all agents of the population, as only local and minor changes to the repertoires of the agents are possible during a single interaction. As a single game is an interaction with only two participants, games are repeated many times with different participants, randomly selected from the population. In every game, the roles of demonstrator and imitator are also assigned at random.

A single imitation game is played as follows (Jansen, 2005a):

1. The initiator randomly selects an action *a* from its repertoire and executes this action. If its repertoire is empty, a new random action is first added.

2. The imitator observes the action, finds the best matching action *a'* from its own repertoire (categorical perception) and executes the action. If its repertoire is empty, a new random action is first added.

3. The initiator observes this action, finds the best matching action *a"* from its own repertoire and compares its initial action *a* with the recognized imitated action *a"*.

4. The initiator compares its initial action *a* with the observed imitation *a"*. If both actions are the same, the game succeeds, otherwise it fails.

5. The initiator announces the outcome of the game to the imitator.

6. Both agents adapt their repertoire using this feedback, such that future games become more successful.

On an abstract level, the game thus consists of three consecutive steps: *interaction*(1–4), *sending feedback*(5) and *learning*(6). Since we do not study learning to imitate, we assume this interaction pattern is simple, fixed, innate and the same for all agents. The interaction pattern was designed such that no external observer is required to judge the success of the game. The initiator decides on the success of the game by comparing its initial action with its best matching action for the observed imitation. Opposed to other approaches, no threshold is required to decide on the success of the game. Due to the categorical perception, the observed action can be compared directly to previously stored actions. So, even if the actions performed by both agents are quite different, imitation can succeed: as long as the demonstrator categorizes the observed imitation and the initial action as the same, the game succeeds. Two observations are categorized as belonging to the same action if they both best resemble the same action in the repertoire of the agent (cf. a prototype).

After the interaction, the initiator sends binary feedback to the imitator. Therefore, we assume a single bit perfect communication channel to exist between the agents. Learning consists of two phases: first the imitator adapts its repertoire: If the game succeeds, the action it used is shifted towards the observed action. If the game fails, the same shift is performed on condition that the action considered was not permanently very successful in past interactions. Since it is of no use to adapt a successful action, a new action, matching the observed action, is created in that case.

This adaptation to the imitators' repertoire is based on the current state of its repertoire, the action observed from the initiator and the feedback it received. This is only local information, i.e. the imitator has no other access to the initiators internal state than by observing its actions and receiving its feedback.

Additionally, both agents perform some general updates on their repertoires:

1. With a small probability the agents can add new random actions to their repertoires.

2. With every action, use and success counters are associated. Whenever an action is performed, its use counter is increased. Whenever imitation succeeds, the success counter of the used action is increased. Actions which have proven to be permanently unsuccessful in the past are removed from the repertoire.

3. Actions which are too similar are merged, such that no confusion can exist between those two actions.

4 Measures

The quality of the imitative attempt is measured by two main monitors: the imitative success and repertoire size. Whether a single imitative attempt fails or succeeds is decided by the agents themselves and is a binary decision as explained above. This allows for a simple measure evaluating the success over a longer time over the population. A running average over a window of 100 games is calculated of the fraction of successful games. Clearly, high imitative success should denote successful imitation. However, if both participants to a game only have learned a single action, the game always succeeds.

In order to investigate how fast the agents succeed in developing a repertoire of actions, the average size of their repertoires is monitored as well. Agents should succeed in building a stable repertoire of behaviours while imitative success should be high.

Besides monitoring the quality of imitation, the motor babbling process can be monitored directly. Therefore we monitor the average associations number (how many action/observation pairs are stored) together with the average forward and inverse predicting error. Those errors are estimated by calculating forward and inverse kinematics for 500 random queries and averaging the prediction results. The errors are scaled to percentages by dividing them by the length of the agents arms. It is also interesting to monitor how the action/observation associations are spread: many associations might be grouped together while other regions are not very crowded.

5 Experiments and results

In the imitation task as specified above, the concept of *action* was not strictly defined. Indeed, the framework of imitation games is a general method for learning categories by means of imitation, no matter what behaviour is precisely imitated. In previous work, we have argued and shown that the precise embodiment of the agent shapes the behaviour which is learned. Rather than simulating an existing robot arm with gripper and the like (as Jansen (2005b) did, a

trivial experimental platform is used here. The platform is similar to the one used by Alissandrakis et al. (2005).

A robot consists of an arm and a camera system, such that it can perform some actions with the arm and observe those actions with the camera system, either performed by himself, either by another agent. The robot arm is two-dimensional and consists of N joints of equal length and there are no constraints on the joint angles. Every configuration of the arm defines a state. Every change from one state to another state defines an action or a sequence of actions. In the work presented here, the robot imitation system itself is not studied. Hence, it is kept as trivial as possible. The actions which are learned are simply arm configurations. In previous work we have shown convincingly that qualitative properties of the simulation results do not depend on the embodiment.

Below, we present results on three experiments in which imitation task which is described above is performed. In the first variant, the agents are endowed with a preprogrammed model of the kinematics of their effector. The agents need those models to map actions on observations and the other way around, for instance when modifying categories in the learning phase. In the second variant of the experiments, all agents learn the models of the kinematics of their effector on an individual basis, prior to imitation. In the third experiment, the agents learn models of the kinematics of their effectors while imitating.

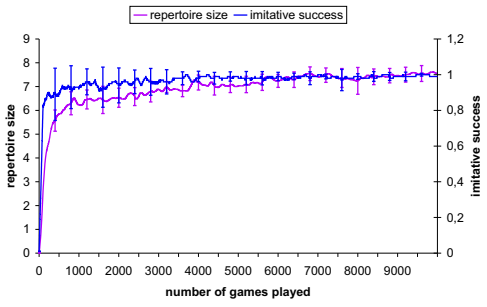


Figure 1: Results for the imitation experiment in which the participants have predefined kinematics models. Imitative success and repertoire size are plotted.

5.1 Imitation with preprogrammed models of kinematics

In this experiment it is assumed that the agents are endowed with a model of the kinematics of their effector. Such a model enables them for instance to predict the outcome of their actions, but also allows them to estimate the actions that should be performed in order to obtain a certain outcome. For the extreme simple case of the 2D arm we proposed, analytical calculations of forward and inverse models are possible, independent of the number of joints in the arms. In many non-trivial cases (exact) analytical calculation of kinematics models is not possible.

In figure 1, results are plotted for an experiment in which two agents engage in 10000 imitation games. Experiments are repeated ten times, such that 95% coinfidence intervals can be drawn easily. Both imitative success and repertoire size are plotted. Results show how the agents succeed in building up a repertoire of actions very fast. The actions in the repertoires of the agents are sufficiently similar to allow for highly successful imitation games. The results here act as a reference to compare results from further experiments with.

5.2 Imitation after learning kinematics models

In the second variant of the experiment, all agents in the population learn models of forward and inverse kinematics on an individual basis, prior to engaging in imitative interactions. In a motor babbling phase, the agent performs random actions and observes their outcome. Associations between actions and observations are stored, such that future queries to the forward or inverse models can use those associations, for instance by using the locally weighted learning method which is explained above.

Rather than adding a fixed amount of associations to the kinematics memory, agents self-evaluate the accuracy of the predicted arm positions and only add associations when the difference between the actual outcome and the outcome which is predicted based on the models learned so far is bigger than a given threshold. After this motor babbling phase, the agents play imitation games as in the previous experiment. Results are plotted in figure 2 at the left and clearly show that the agents are capable of learning the kinematics model as their performance in the imitation game is hardly not affected: the agents succeed in learning a stable repertoire of actions which can be imitated with high success. Results also show that a fixed amount of action/observation pairs is added to the agents' repertoires: less than 250 associations allow them to predict the kinematics models accurately enough to allow for successful imitation.

A plot is provided which shows the spread of the

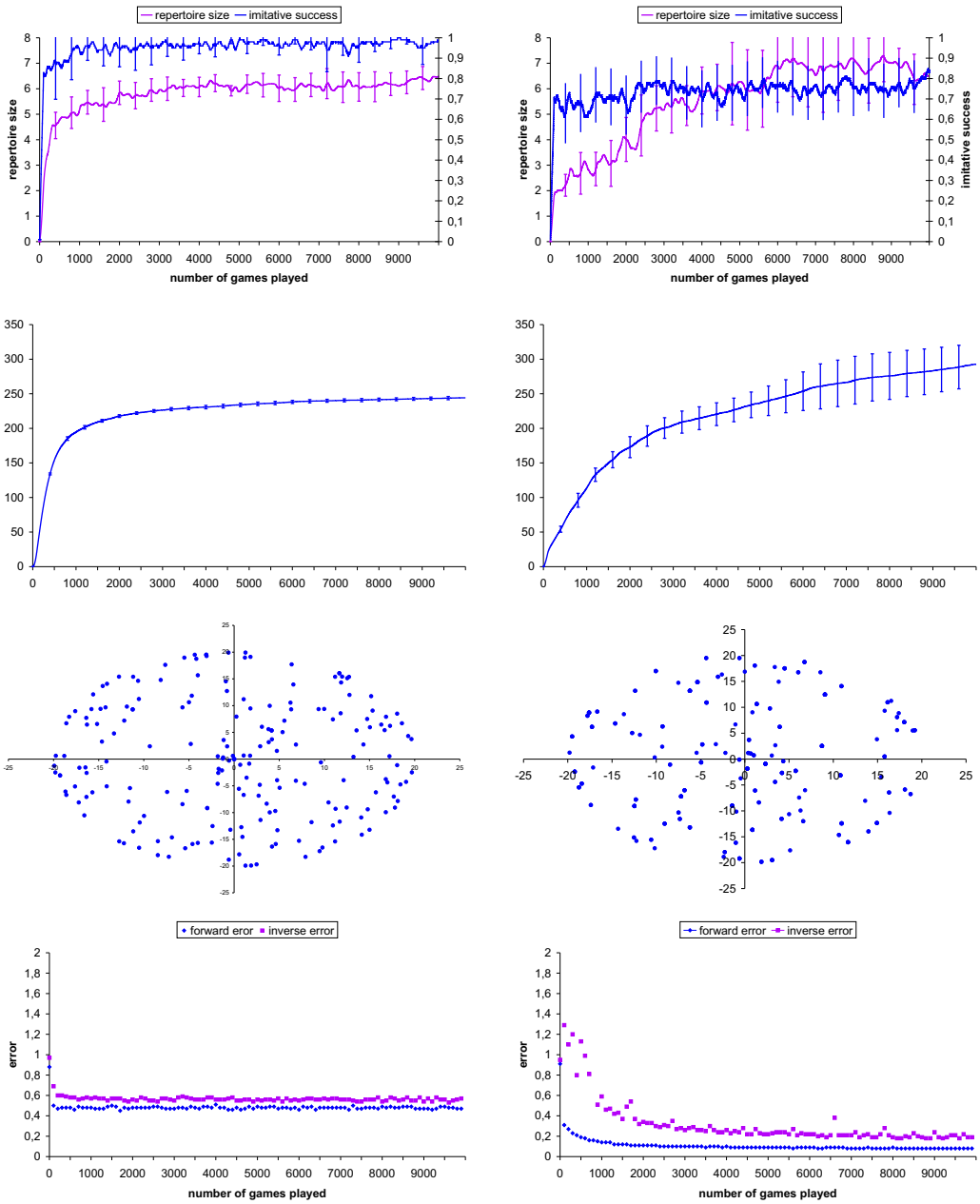


Figure 2: Results of the second (left plots) and third (right plots) variant of the experiment. Top row: imitative success and repertoire size. Second row: average amount of action-observation associations. Third row: Visual representation of the action space of the agents. Bottom row: forward and inverse models error.

actions over the action space. From the figure it seems that it is well covered, which is the natural consequence of generating random actions. This plot serves as a reference for comparison to the third experiment.

5.3 Imitation while learning kinematics models

In this third experiment, the agents start without any explicit or implicit kinematics model. The models will be learned while playing imitation games. We want to investigate whether kinematics models can be learned successfully by imitating and whether imitation can be successful with an improving kinematics model.

Results of the experiment are shown in figure 2 at the right hand side and clearly show that also in this case successful imitation is possible: a stable repertoire of actions emerges while imitative success is high. At the beginning of the experiment, the agents are equipped with a very inaccurate kinematics model. Nevertheless, successful imitation is possible, due to the categorical perception mechanism. During the first stage of the experiment, the agents have learned very few actions. Even with the inaccurate kinematics models, it is for instance possible to distinguish between as few as two actions. While imitating those two actions, new action - observation pairs are stored and the kinematics models become more accurate. So, all associations which are stored are caused by actions to be performed and imitated. Nevertheless, the entire motor space is well covered with associations (see third figure), meaning that the entire space of actions was well explored in learning actions which are easy to distinguish and to imitate.

6 Conclusions

From the three simple experiments which are reported in this document, several conclusions can be drawn: The agents succeed in imitating each other, whether they have a preprogrammed kinematics model or a model which is learned either prior to imitation, either while imitating. The agents succeed in building a repertoire of actions which is stable. The repertoires of the agents are shared in the sense that they are similar enough to allow for successful imitation.

A more general conclusion is that there are at least specific cases in which a kinematics model is not a prerequisite to imitation as it can be learned while imitating. The categorical perception mechanism allows for successful imitation without accurate kinematics models in the specific imitation paradigm reported here.

Acknowledgment

Part of the research was performed when the author was working at the VUB AI-Lab. He was funded by the IWT (Institute for the Promotion of Innovation by Science and Technology in Flanders). Currently, the author works at the Electrical Engineering and Informatics group (ETRO) of the Faculty of Engineering Sciences of the VUB and is funded by Brucare.

References

Aris Alissandrakis, Chrystopher L. Nehaniv, and Kerstin Dautenhahn. Imitating with alice: Learning to imitate corresponding actions across dissimilar embodiments. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, 32(4):482–496, 2002.

Aris Alissandrakis, Chrystopher L. Nehaniv, Kerstin Dautenhahn, and Joe Saunders. Achieving corresponding effects on multiple robotic platforms: Imitating in context using different effect metrics. In *Proceedings of the Third International Symposium on Imitation in Animals and Artifacts*, pages 10–19. The Society for the Study of Artificial Intelligence and Simulation of Behaviour, 2005.

Chris Atkeson, Andrew Moore, and Stefan Schaal. Locally weighted learning. *AI Review*, 11:11–73, April 1997a.

Chris Atkeson, Andrew Moore, and Stefan Schaal. Locally weighted learning for control. *AI Review*, 11:75–113, April 1997b.

Aude Billard, Yann Epars, Sylvain Calinon, Stefan Schaal, and Gordon Cheng. Discovering optimal imitation strategies. *Robotics and autonomous systems*, 47:69–77, 2004.

Aude Billard and Maja J. Matarić. Learning human arm movements by imitation: Evaluation of biologically inspired connectionist architecture. *Robotics and Autonomous Systems*, (941):1–16, 2001.

S. Calinon, F. Guenter, and A. Billard. Goal-directed imitation in a humanoid robot. In *Proceedings of*

the IEEE Intl Conference on Robotics and Automation (ICRA), Barcelona, Spain, April 18-22 2005.

F. L. Crabbe and M. G. Dyer. Observation and imitation: Goal sequence learning in neurally controlled construction animats: VI-MAXSON. In *Proceedings of the Sixth International Conference on Simulation of Adaptive Behavior (SAB)*, 2000.

Bart de Boer. *Self organisation in vowel systems*. PhD thesis, Artificial Intelligence Lab, Vrije Universiteit Brussel, 1999.

Bart de Boer. Self organization in vowel systems. *Journal of Phonetics*, 28:441–465, 2000.

Yiannis Demiris and Gillian M. Hayes. Imitative learning mechanisms in robots and humans. In Volker Klingspor, editor, *Proceedings of the 5th European Workshop on Learning Robots*, pages 9–16, Bari, Italy, 1996.

J. H. Friedman, F. Baskett, and L. J. Shustek. An algorithm for finding nearest neighbors. *IEEE Trans. Comput.*, C-24:1000–1006, 1975.

Bart Jansen. An imitation game for emerging action categories. In Wolfgang Banzhaf, Thomas Christaller, Peter Dittrich, Jan Kim, and Jens Ziegler, editors, *Proceedings of the 7th European Conference on Artificial Life, Lecture Notes in Artificial Intelligence*, pages 800–809, Berlin, 2003. Springer.

Bart Jansen. Imitation of actions between differently embodied agents, without kinematic models and without feedback. In *Proceedings of TAROS2005 (to appear)*, 2005a.

Bart Jansen. *Robot imitation - The emergence of shared behaviour in a population of agents*. PhD thesis, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel, Belgium, 2005b.

Bart Jansen and Tony Belpaeme. A computational model of intention reading in imitation. *Robotics and Autonomous Systems*, (To appear), 2005.

Bart Jansen, Bart De Vylder, Bart de Boer, and Tony Belpaeme. Emerging shared action categories in robotic agents through imitation. In *Proceedings of the Second International Symposium on Imitation in Animals and Artifacts.*, pages 145–152, 2003.

Bart Jansen, Tom ten Thij, Tony Belpaeme, Bart De Vylder, and Bart de Boer. Imitation in embodied agents results in self-organization of behavior. In *Proceedings of the Simulation of Adaptive Behavior conference (SAB 2004)*, Los Angeles, 2004.

M. Jordan and D. Rumelhart. Forward models: Supervised learning with a distal teacher. *Cognitive Science*, 16:307–354, 1992.

M. Kawato. Computational schemes and neural network models for formation and control of multijoint arm trajectories. In *Neural networks for control*, pages 197–228. MIT Press, Cambridge, 1990.

Y. Kuniyoshi, M. Inaba, and H. Inoue. Learning by watching: Extracting reusable task knowledge from visual observation of human performance. *IEEE Transactions on robotics and automation*, 10:799–822, 1994.

M.H. Lee and Q. Meng. Growth of motor coordination in early robot learning. In *Proceedings of IJCAI 2005*, page 1732, 2005.

A. Lockerd and C. Breazeal. Tutelage and socially guided robot learning. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2004)*, 2004.

Erhan Oztop, Nina S. Bradley, and Michael A. Arbib. Infant grasp learning: a computational model. *Experimental Brain Research*, 158:480–503, 2004.

O Sporns and GM Edelman. Solving bernsteins problem: a proposal for the development of coordinated movement by selection. *Child Dev*, 64:960–981, 1993.

C. Versino and L.M. Gambardella. Learning the visuomotor coordination of a mobile robot by using the invertible kohonen map. In *From natural to artificial neural computation, international workshop on artificial neural networks*, pages 84–91, Germany, 1995. Springer-Verlag.

Adaptive combination of motor primitives

F. Nori* G. Metta* L. Jamone* G. Sandini*
*LIRA-Lab, DIST
University of Genova, ITALY
{iron, pasa, lorejam, giulio}@liralab.it

Abstract

Recently there has been a growing interest in modeling motor control systems with modular structures. Such control structures have many interesting properties, which have been described in recent studies. We here focus on some properties which are related to the fact that specific set of contexts can themselves be modeled modularly.

1 Introduction

Humans exhibit a broad repertoire of motor capabilities which can be performed in a wide range of different environments and situations. From the point of view of control theory, the problem of dealing with different environmental situations is nontrivial and requires significant adaptive capabilities. Even the simple movement of lifting up an object, depends on many variables, both *internal and external* to the body. All these variables define what is generally called the context of the movement. As the context of the movement alters the input-output relationship of the controlled system, the motor command must be tailored so as to take into account the current context. In everyday life, humans interact with multiple different environments and their possible combinations. Therefore, a fundamental question in motor control concerns how the control system adapts to a continuously changing operating context.

Recently, there has been a major interest in modeling motor control by means of combinations of a finite number of elementary modules. Within this modular approach, multiple controllers co-exist, with each controller suitable for a specific context. If no controller is available for a given context, the individual controllers can be combined to generate an appropriate motor command. Among the features of this model, two are extremely relevant:

- **Modularity of contexts.** The contexts within which the model operates can be themselves modular. Experiences of past contexts and objects can be meaningfully combined; new situations can be often understood in terms of combinations of previously experienced contexts.
- **Modularity of motor learning.** In a modular structure only a subset of the individual modules cooperate in a specific context. Consequently, only these modules have a part in motor learning, without affecting the motor behaviors already learned by other modules. This situation seems more realistic than a global structure where a unique module is capable of handling all possible contexts. Within such a global framework, motor learning in a new context possibly affects motor behaviors in other (previously experienced) contexts.

Remarkably, Mussa-Ivaldi and Bizzi (2000) have proposed an interesting experimental evidence supporting the idea that biological sensory-motor systems are organized in modular structures. At the same time, Shadmehr and Mussa-Ivaldi (1994) and Brashers-Krug et al. (1996) have shown the extreme adaptability of the human motor control system. So far, adaptation has been proven to depend on performance errors (see Shadmehr and Mussa-Ivaldi (1994)) and context related sensory information (see Shelhamer et al. (1991)).

Based on these findings, there has been recently a growing interest in investigating the potentialities of *adaptive and modular* control schemes (refer to Wolpert and Kawato (1998); Mussa-Ivaldi (1997)). Within these investigations, the modular structure is often formalized in terms of multiple inverse models[1]. Motor commands are usually obtained by combining these elementary inverse models. Given this formalization, two fundamental questions must be faced:

1. Is there a way to choose the elementary inverse

[1]Here an inverse model is considered to be a map from desired movements to motor commands.

models so as to cover all the contexts within a specified set?

2. Given a set of inverse models which appropriately cover the set of contexts which might be experienced, how is the correct subset of inverse models selected for the particular current context?

Both questions have been already investigated in Wolpert and Kawato (1998) and in Mussa-Ivaldi and Giszter (1992) within the function approximation framework. Recently, the same two questions have been considered by Nori and Frezza (2005) within a control theoretical framework. So far this innovative approach has been proven to provide new interesting results in answering the first question (see Nori and Frezza (2004a) and Nori (2005)). In the present paper, we proceed along the same line to answer the second question. Specifically, we propose a strategy to adaptively select a given set of inverse models. The selection process is based on the minimization of performance errors. Context related sensory information (which is related to a different cognitive process) is instead not considered here. The key features of the proposed control scheme are the following:

- **Minimum number of modules.** Previous works Nori (2005) have established the minimum number of modules which are necessary to cover all the contexts in a specified set. The present paper will describe how this minimality result can be fitted in the adaptive selection of the modules.

- **Linear combination of modules.** The theory of adaptive control has been widely studied since the early seventies. Interesting results have been obtained, especially in those situations where some linearity properties can be proven and exploited. In our case, linearity will be a property of the considered set of admissible contexts.

2 Reaching in different contexts

To exemplify the ideas presented in the introduction, we consider a specific action, nominally the action of reaching a target with the hand. In order to immerse the same action into different contexts, we consider the movement of reaching while holding objects with different masses and inertias. Within this framework, a successful execution of the reaching movements requires a control action which should adapt to the current context. Since the controlled system[2] changes its properties with the context, suitable changes should be imposed on the control action.

2.1 Model of the arm

We model the dynamics of the arm as a fully actuated kinematic chain with n degrees of freedom corresponding to n revolute joints. It is well known in literature that such model can be expressed as follows:

$$M(\mathbf{q})\ddot{\mathbf{q}} + C(\mathbf{q}, \dot{\mathbf{q}})\dot{\mathbf{q}} + g(\mathbf{q}) = \mathbf{u}, \tag{1}$$

where $\mathbf{q}$ are the generalized coordinates which describe the pose of the kinematic chain, $\mathbf{u}$ are the control variables (nominally the forces applied at the joints) and the quantities M, C and g are the inertia, Coriolis and gravitational components.

2.2 Model of the contexts

In this paper, we consider the problem of controlling (1) within different contexts. The different contexts affect the arm in terms of modifying its dynamical parameters. The considered parameters are the mass, the inertia and the center of mass position of each of the n links which compose the controlled arm. The vector with components represented by these parameters is:

$$\mathbf{p} = \begin{bmatrix} m_i & I_1^i & \dots & I_6^i & {\mathbf{c}^i}^\top \end{bmatrix}_{i=1\dots n}^\top, \tag{2}$$

where m_i is the mass of the i^{th} link, $I_1^i, \dots, I_6^i$ represent the entries of the symmetric inertia tensor, and $\mathbf{c}^i = \begin{bmatrix} c_x^i, c_y^i, c_z^i \end{bmatrix}^\top$ is the center of mass position. The system to be controlled is therefore:

$$M_{\mathbf{p}}(\mathbf{q})\ddot{\mathbf{q}} + C_{\mathbf{p}}(\mathbf{q}, \dot{\mathbf{q}})\dot{\mathbf{q}} + g_{\mathbf{p}}(\mathbf{q}) = \mathbf{u}, \tag{3}$$

where $\mathbf{p}$ identifies by the specific context. Note that the considered class of contexts is suitable for modeling an arm which holds objects with different masses and inertias. Therefore, the model is appropriate for the proposed reaching scenario.

Note: the proposed set of contexts is itself modular. It can indeed be proven that (see Kozlowski (1998)):

$$M_{\mathbf{p}}(\mathbf{q})\ddot{\mathbf{q}} + C_{\mathbf{p}}(\mathbf{q}, \dot{\mathbf{q}})\dot{\mathbf{q}} + g_{\mathbf{p}}(\mathbf{q}) = \mathbf{u}, \tag{4}$$

can be rewritten as:

$$\sum_{j=1}^{J} \Psi_j(\mathbf{p}) Y^j(\mathbf{q}, \dot{\mathbf{q}}, \ddot{\mathbf{q}}) = \mathbf{u}. \tag{5}$$

[2] composed of the arm *and* the held object.

This is a crucial property which is fundamental to prove the results which will be claimed in the rest of this paper.

3 Modular control action

In this section we formalize our concept of modular control action. The proposed formalization is biologically inspired and has been originally proposed by Mussa-Ivaldi and Bizzi (2000). Specifically, experiments on frogs and rats have shown that their motor systems is organized into a *finite number of modules*. Each module has been described in terms of the muscular synergy evoked by the microstimulation of specific interneurons in the spinal cord. These modules can be *linearly combined* to achieve a wide repertoire of different movements. A mathematical model of these findings has been proposed again by Mussa-Ivaldi and Bizzi (2000):

$$\mathbf{u} = \sum_{k=1}^{K} \lambda_k \Phi^k(\mathbf{q}, \dot{\mathbf{q}}). \tag{6}$$

Practically, the system (1) is controlled by means of a linear combination of a finite number of modules $\Phi^k(\mathbf{q}, \dot{\mathbf{q}})$, which can be seen as elementary control actions. The control signals are no longer the forces $\mathbf{u}$ but the vector $[\lambda_1, \ldots, \lambda_K]^\top = \lambda$ used to combine the modules.

3.1 Modules synthesis problem

Remarkably, a modular structure requires a major attention in selecting the modules themselves. In this section it is pointed out that only a suitable choice of the modules allow to generate a wide repertoire of movements (Section 3.1.1) while handling different contexts (Section 3.1.2).

3.1.1 Modules for reaching admissible configurations

Obviously, the individual modules Φ^k need to be carefully chosen in order to preserve the capability of reaching any admissible configuration[3]. Simple examples can demonstrate that, in general, this capability may be easily lost. As to this concern, the following problem has been formulated:

Problem 1: *find a set of modules* $\{\Phi^1, \ldots, \Phi^K\}$ *and a continuously differentiable function* $\lambda(\cdot)$*, such that for every desired final state* $\mathbf{q}_f$ *the input:*

$$\mathbf{u} = \sum_{k=1}^{K} \lambda_k(\mathbf{q}_f) \Phi^k(\mathbf{q}, \dot{\mathbf{q}}) \tag{7}$$

steers the system (1) to the configuration $\mathbf{q}_f$*.*

Nori and Frezza (2004b) have proposed a solution to the problem above with the use of $n + 1$ modules. This solution was proven to be composed by a minimum number of modules (see Nori (2005)).

3.1.2 Modules for handling admissible contexts

In this paper, we consider the problem of solving problem 1 in different contexts. Practically, we face the following problem where instead of controlling (1) we want to control (3) which is context dependent.

Problem 2: *find a set of modules* $\{\Phi^1, \ldots, \Phi^K\}$ *and a continuously differentiable function* $\lambda(\cdot, \cdot)$*, such that for every desired final state* $\mathbf{q}_f$ *and for every possible context* $\mathbf{p}$ *the input:*

$$\mathbf{u} = \sum_{k=1}^{K} \lambda_k(\mathbf{q}_f, \mathbf{p}) \Phi^k(\mathbf{q}, \dot{\mathbf{q}}) \tag{8}$$

steers the system (3) to the configuration $\mathbf{q}_f$*.*

Obviously the proposed problem is related to the question posed in the introduction: is there a way to choose the elementary (inverse) models so as to cover all the contexts within a specified set? The answer turns out to be 'yes'. Specifically, a complete procedure for constructing a solution of problem 2 has been proposed in Nori (2005). The solution turns out to have the following structure:

$$\mathbf{u} = \sum_{i}^{I} \sum_{j}^{J} \lambda_i(\mathbf{q}_f) \mu_j(\mathbf{p}) \Phi^{i,j}(\mathbf{q}, \dot{\mathbf{q}}), \tag{9}$$

where $\{\Phi^{1,j}, \ldots, \Phi^{I,j}\}$ is a solution to problem 1 for a specific context $\mathbf{p}^j$.

3.2 Adaptive modules combination

In many situations, the context of the movement is not known *a priori*. Within our formulation, if the context $\mathbf{p}$ is unknown, we cannot compute the way the modules have to be combined. This is a consequence of the fact that the way modules are combined depends not only on the desired final position $\mathbf{q}_f$ but

[3] in control theory the capability of the system of reaching any admissible configuration is called *controllability* (see Nori and Frezza (2005)).

also on the current context $\mathbf{p}$. A possible solution consists in adaptively choosing μ_j (which are context dependent) on the basis of available data. When the only information available is the performance error[4], we can reformulate the estimation problem in terms of an adaptive control problem. It can be proven that a way to successfully reach the desired final position $\mathbf{q}_f$ consists in adaptively adjusting μ_j according to the following differential law:

$$\frac{d}{dt}\mu_j = -\mathbf{s}^\top \left[\sum_i^I \lambda_i(\mathbf{q}_f)\Phi^{i,j}(\mathbf{q}, \dot{\mathbf{q}}) \right], \qquad (10)$$

where $\mathbf{s}$ is the performance error (see Kozlowski (1998) for details). A mathematical proof of the system stability properties is out of the scope of the present paper and is therefore omitted. It suffices to say that, in fact, it can be proven that (10) leads to a stable system.

4 Future works

In the framework of motor control, this paper proposes a method for performing on-line learning of reaching movements. The proposed control structure is not only biologically compatible, but turns out to be very useful when dealing with modular contexts. A crucial step in our future work will be the implementation of the system in a robot capable of adapting itself to different contexts determined, for instance, by manipulating/holding different objects. The underlying idea is that a modular control structure should reveal useful for handling objects which are themselves modular.

5 Conclusions

Modular control structures are appealing since there exist contexts which can be modular as well. In the present paper we have considered a simple movement (moving the arm towards a target) within different contexts (handling different objects). Intuitively, a modular control structure is best suited to operate within modular contexts. In the specific problem of moving the arm while holding different objects, we have shown that the system dynamics are modular themselves. Taking advantage of this property we have shown that a modular control structure is capable of handling multiple contexts. Finally, a way to adaptively combine the modules has been proposed.

[4]The performance error $\mathbf{s}$ measures the difference between the desired reaching trajectory $\mathbf{q}^d$ and the actual trajectory q. Further details can be found in Kozlowski (1998)

Acknowledgements

The work presented in this paper has been supported by the project ROBOT CUB (IST-2004-004370) and by the project NEUROBOTICS (IST-2003-511492). Both projects are funded by the European Union through the Sixth Framework Programme for Research and Technological Development (FP6).

References

T. Brashers-Krug, R. Shadmehr, and E. Bizzi. Consolidation in human motor memory. *Nature*, 382:252–255, 1996.

K. Kozlowski. *Modelling and Identification in Robotics*. Springer-Verlag, 1998.

F. A. Mussa-Ivaldi. Nonlinear force fields: A distributed system of control promitives for representing and learning movements. In *CIRA '97, Montrey California, USA*, 1997.

F. A. Mussa-Ivaldi and E. Bizzi. Motor learning through the combination of primitives. *Philosophical Transactions of the Royal Society: Biological Sciences*, 355: 1755–1769, 2000.

F.A. Mussa-Ivaldi and S.F. Giszter. Vector field approximation: A computational paradigm for motor control and learning. *Biological Cybernetics*, 37:491–500, 1992.

F. Nori. *Symbolic Control with Biologically Inspired Motion Primitives*. PhD thesis, Università degli studi di Padova, 2005.

F. Nori and R. Frezza. A control theory approach to the analysis and synthesis of the experimentally observed motion primitives. *Biological Cybernetics*, 93(5):323–342, 2005.

F. Nori and R. Frezza. Biologically inspired control of a kinematic chain using the superposition of motion primitives. In *Proceedings of 2004 Conference on Decision and Control*, pages 1075–1080, 2004a. CDC'04.

F. Nori and R. Frezza. Nonlinear control by a finite set of motion primitives. In *Proceedings of 2004 Nonlinear Control Systems Design*. IFAC, 2004b.

R. Shadmehr and F. A. Mussa-Ivaldi. Adaptive representation of dynamics during learning of a motor task. *Journal of Neuroscience*, 74(5):3208–3224, 1994.

M. Shelhamer, D.A. Robinson, and H.S. Tan. Context-specific gain switching in the human vestibuloocular reflex. *Annals of the New York Academy of Sciences*, 656 (5):889–891, 1991.

D.M. Wolpert and M. Kawato. Multiple paired forward and inverse models for motor control. *Neural Networks*, 11: 1317–1329, 1998.

Experimental Comparison of the van der Pol and Rayleigh Nonlinear Oscillators for a Robotic Swinging Task

Paschalis Veskos* and Yiannis Demiris*
*Biologically-inspired Autonomous Robots Team (BioART)
Department of Electrical and Electronic Engineering
Imperial College London
South Kensington, London SW7 2BT, UK
{paschalis.veskos, y.demiris}@imperial.ac.uk

Abstract

In this paper, the effects of different lower-level building blocks of a robotic swinging system are explored, from the perspective of motor skill acquisition. The van der Pol and Rayleigh oscillators are used to entrain to the system's natural dynamics, with two different network topologies being used: a symmetric and a hierarchical one. Rayleigh outperformed van der Pol regarding maximum oscillation amplitudes for every morphological configuration examined. However, van der Pol started large amplitude relaxation oscillations faster, attaining better performance during the first half of the transient period. Hence, even though there are great similarities between the oscillators, differences in their resultant behaviours are more pronounced than originally expected.

1 Introduction

Various neural oscillators have been used in the past to implement several rhythmic motor control tasks. Mutually-inhibiting neurons (Matsuoka 1985) have been used to entrain humanoid arms with a slinky toy and turn a crank (Williamson 1998), bipedal walking (Taga 1991; Taga 1995), swinging (Lungarella and Berthouze 2002; Matsuoka, Ohyama et al. 2005), and bouncing (Lungarella and Berthouze 2004), while the van der Pol and Rayleigh oscillators have been utilised for the purposes of planar bipedal walkers (Zielinska 1996; Dutra, de Pina Filho et al. 2003; de Pina Filho, Dutra et al. 2005). In motor control studies, systems are often treated at a more abstract level of behaviour and less attention is paid to the impact the low level components have on the overall functioning of the system. In a previous study (Veskos and Demiris 2005a) we investigated the use of the van der Pol oscillator for a robotic swinging task. In this paper, we implement an additional oscillator, known as the Rayleigh oscillator. The two oscillators have a similar mathematical structure, thus allowing us to make direct comparisons between them and the resultant behaviours. We are specifically interested in determining whether this similar basic building block alters the higher-level behaviours of the system. Furthermore, we also wish to investigate the influence of different oscillatory network topologies. We therefore experimented with a hierarchical network structure, in addition to the previously used symmetric one.

2 Experimental Setup

We utilise two similar nonlinear oscillators to build the neural control system for our experiments: van der Pol and Rayleigh. Additionally, we connect these in two different manners, using a symmetric and a hierarchical topology.

2.1 Nonlinear Oscillators

The equations of the van der Pol (vdP) oscillator, as used in our experiments, are of the form:

$$\ddot{x}_i + \mu\left(x_i^2 - 1\right)\cdot \dot{x}_i + \omega^2 x_i = G_{in}\cdot fb + G_{i-j}\cdot x_j \quad (1)$$

where $i, j = \{\text{hip}, \text{knee}\}$, $\mu \geq 0$ is a parameter controlling the damping term, ω is the natural frequency of the oscillator, fb is the feedback from the vision system, G_{in} is the feedback gain, while $G_{hip-knee}$ and $G_{knee-hip}$ are the cross-coupling term gains. The final output given to the position-controlled motors activating the joints, is:

$$\theta_i = G_{out}\cdot sign\left(\dot{x}_i\right),\ i = \{hip, knee\} \quad (2)$$

where G_{out} is the output gain.

For the Rayleigh oscillator, $\dot{x}$ is inserted in the $\left(x_i^2 - 1\right)$ term to yield:

$$\ddot{x}_i + \mu\left(\dot{x}_i^2 - 1\right)\cdot \dot{x}_i + \omega^2 x_i = G_{in}\cdot fb + G_{i-j}\cdot x_j \quad (3)$$

This difference alters the response of the two oscillators to changes in their natural frequency. For the vdP, increasing ω increases the oscillator's output *frequency*, while for Rayleigh has the effect of increasing output *amplitude*. Given that we only make use of the timing information and discard the amplitude in equation (2), it should be easier for Rayleigh to achieve entrainment to mechanical systems as its own natural dynamics are less pronounced. Furthermore, simulations of a planar bipedal walker task have shown Rayleigh to recover from random perturbations faster than van der Pol (Roy and Demiris 2005).

2.2 Neural Topologies

Two different neural topologies were investigated by altering the values of the cross-coupling gains. By equating them, the topology is symmetric, where both degrees of freedom affect each other and strong neural entrainment takes place. This is shown in Figure 1.

To arrive at a hierarchical topology where only the hip oscillators directly receive the feedback signal, the vision feedback is not forwarded to the knee oscillator. Additionally, the intra-neural connection sending information back to the hip oscillator is severed. The knee oscillator can then entrain to the mechanical system solely by means of the hip-knee connection. This is better illustrated in Figure 2.

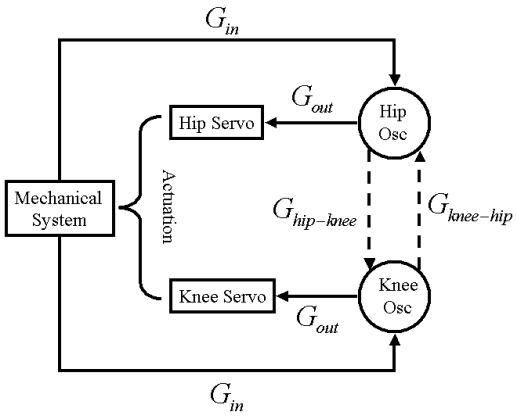


Figure 1: A functional overview of our experimental system for the symmetric neural topology. Both oscillators receive vision feedback and strong neural entrainment is facilitated by the intra-neural connections (shown with the dashed line).

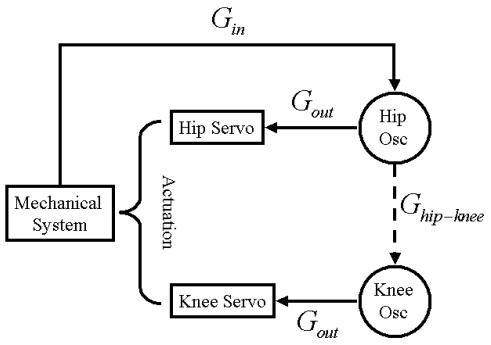


Figure 2: In the hierarchical topology, only the proximal (hip) oscillator directly receives vision feedback, which is then propagated to the distal (knee) oscillator, as shown by the dashed line.

2.3 Mechanical Setup

Experiments were performed on the robotic platform previously described in (Veskos and Demiris 2005b), shown in Figure 3.

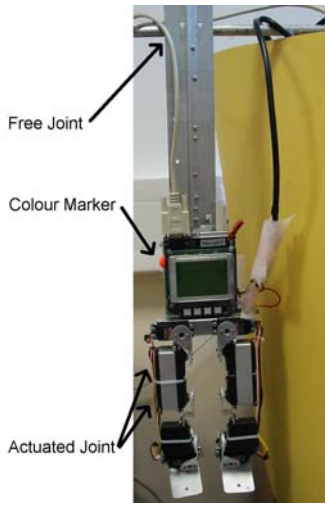


Figure 3: The robotic setup

The robot can be thought of as an underactuated triple pendulum with the top joint being free while the bottom two joints are totally forced to the output of the nonlinear oscillator. A coloured marker on the robot is tracked by a webcam viewing the setup from the side. The x coordinate of this marker is then used as feedback for the neural oscillator (*fb* term). In this study, only the hip and knee joints were actuated, while all others on the robot were held stiff.

3 Experiments

Actuating the hip joint with the van der Pol and Rayleigh oscillators result in the phase plots of Figure 4. The maximum amplitude of oscillation of the robot is 39% larger using Rayleigh (179 instead of 129 units), for the same value of the natural frequency, $\omega^2 = 3.0$. Although the van der Pol oscillator has a more consistent neural limit cycle with less variation in amplitude, the mechanical system operates more smoothly with the Rayleigh oscillator. This is evidenced by the more even mechanical system plot; the phase portrait (a projection of the 3D plot on 2D, by removing the time axis) resembles a circle rather than an hourglass-like shape. The irregularities distorting the uniformity of the limit cycle occur at the point corresponding to the robot's flight phase past the midway position. Its speed there should be maximum, but the van der Pol shows a relative reduction in the value of the derivative, thus causing this "dent". While this improves as the system reaches the steady state, it does not disappear and is an indication of task suboptimality.

In the symmetric neural topology, strong neural entrainment takes place and due to the symmetry of the feedback system, the hip and knee oscillators essentially identical outputs, completely in phase. Again, Rayleigh was capable of producing larger amplitude oscillations than van der Pol, given the same system parameters: 214 versus 182 units, an 18% difference. These results are illustrated in Figure 5.

Results for the hierarchical topology in terms of maximum oscillation amplitudes were very similar, with the corresponding values being 214 and 181 units (Figure 6). In terms of the timing however, the symmetry in the neural topology makes the coupling between the two joint oscillators weaker and allows for delays to be introduced between the hip and knee. Rayleigh, however is much more resilient to this effect, as illustrated in Figure 7.

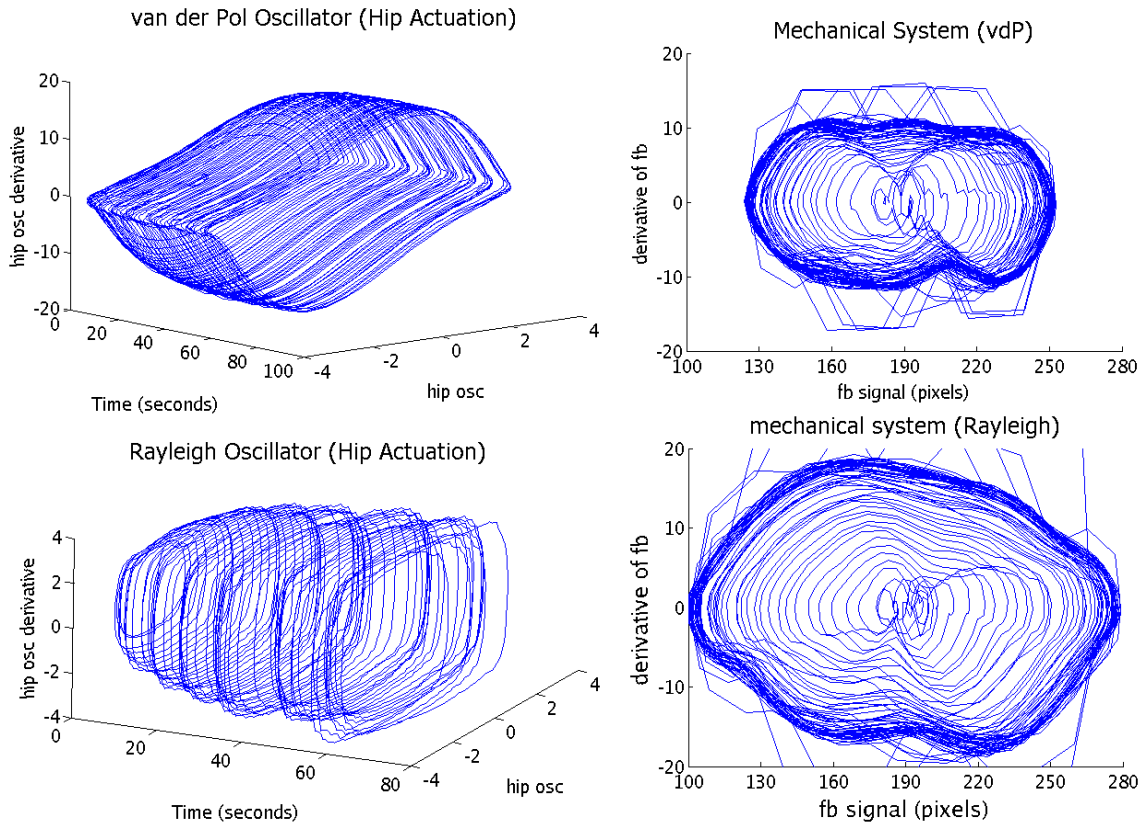


Figure 4: Phase plots for 1-DOF proximal (hip) actuation. The van der Pol oscillator is shown on the top row: the neural system on the left and a phase portrait of the mechanical system on the right. The plots for the Rayleigh oscillator are shown in the bottom row.

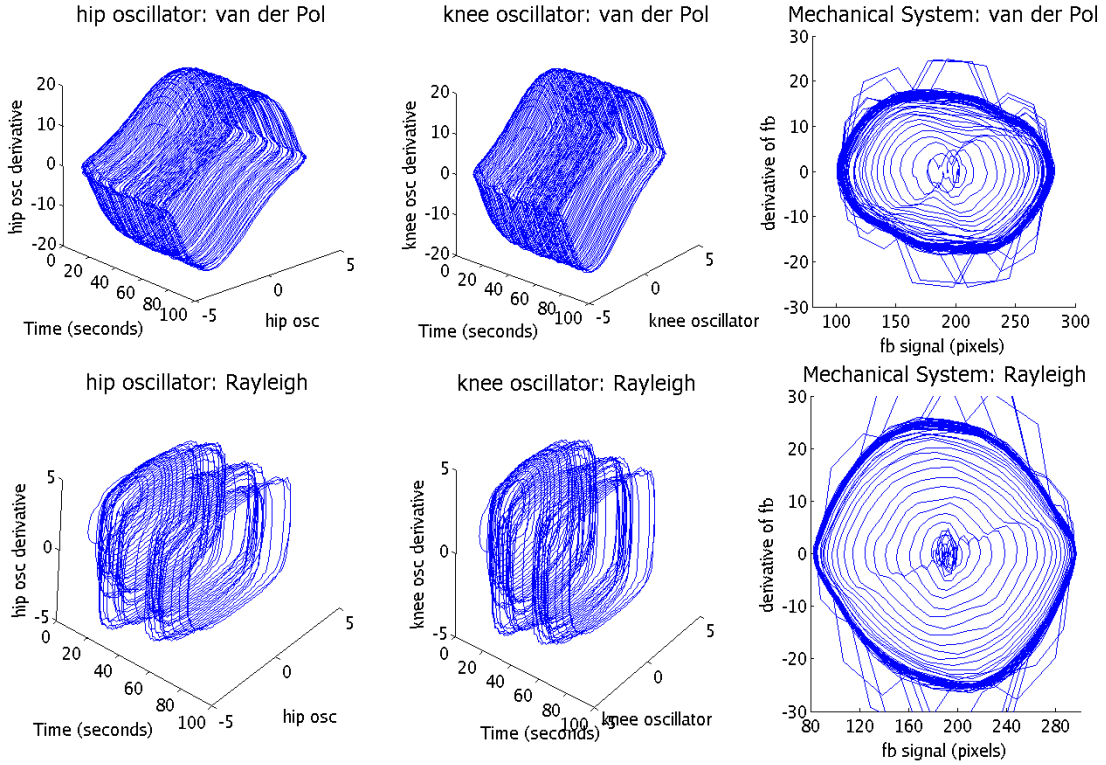


Figure 5: Phase plots for 2-DoF actuation with the symmetric neural topology. Results for van der Pol oscillator are shown on the top row and for Rayleigh on the bottom. From left to right, the columns are: hip oscillator phase plot, knee oscillator phase plot and mechanical system phase portrait.

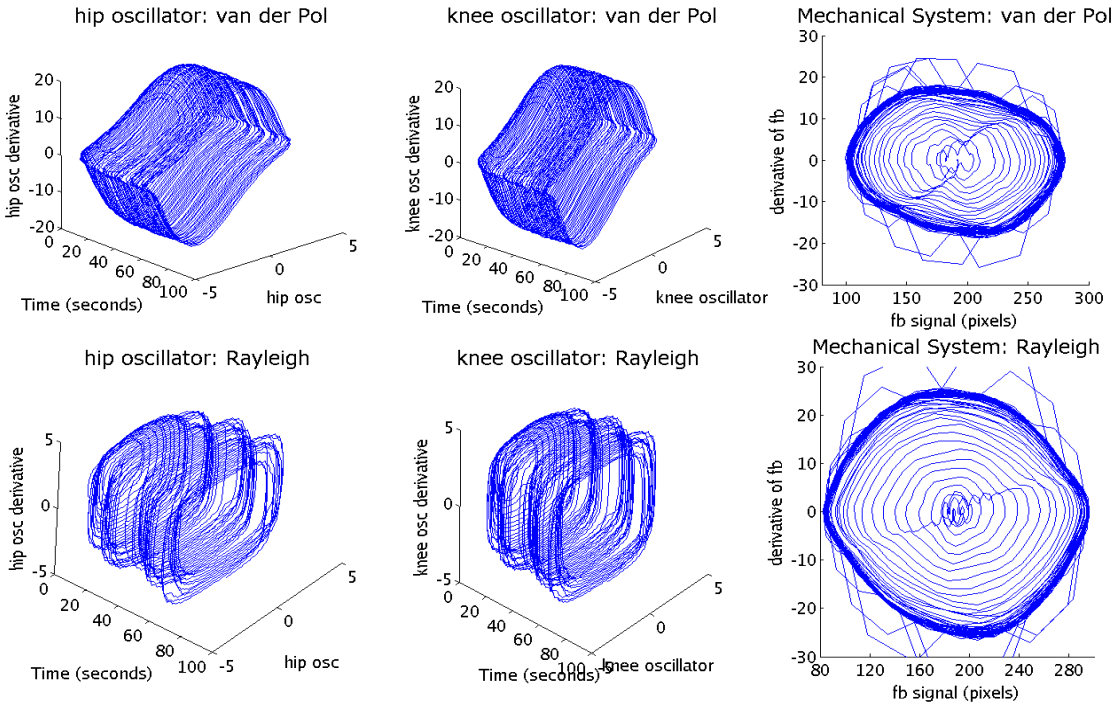


Figure 6: Phase plots for the hierarchical network topology. The asymmetry in the neural topology makes the coupling between the two joint oscillators weaker and introduces a small phase difference. This way the limit cycles are not identical for both joints as in the symmetric case.

4 Discussion

Larger oscillation amplitudes were always achieved when both degrees of freedom were actuated. Injecting more energy in the system also made limit cycles smoother, eliminating the suboptimal speed drops observed for the hip-only actuation scheme.
To analyse the transient behaviour of the two different oscillators, a comparison of the envelopes of oscillation for the entire experiments was made. This is shown in Figure 9. Something that should be noted is that van der Pol started producing relaxation oscillations earlier, thus giving it an advantage over the first ten seconds of the trial. This phenomenon is more pronounced in the case of both degrees of freedom being activated. Another two trials were performed where the second degree of freedom (knee) was released at t=5s. This moment was chosen as an 'early' release point where the system was still in its transient state. The Rayleigh oscillator's behaviour is almost identical to the 1-dof case until t=9s and only manages to reach the performance of the vdP at t=12s.

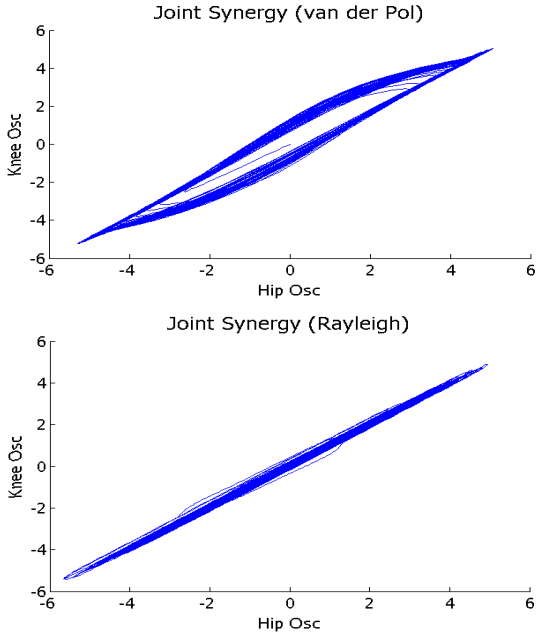


Figure 7: Hip-knee joint correlation plots for the van der Pol (top) and Rayleigh (bottom) oscillators in the hierarchical topology experiments. Rayleigh manages to maintain a 1:1 timing relationship between the two joints, while vdP introduces a phase difference.

Additionally, to compare the oscillators' frequency adaptation speed, the instantaneous period during the above experiments was plotted in Figure 8. Rayleigh consistently forces the mechanical system to oscillate at a lower frequency than van der Pol. This phenomenon is especially pronounced for the 2-DoF configurations. The difference in topology seems to have little effect on this matter; period of oscillation remains unaffected for a given oscillator.

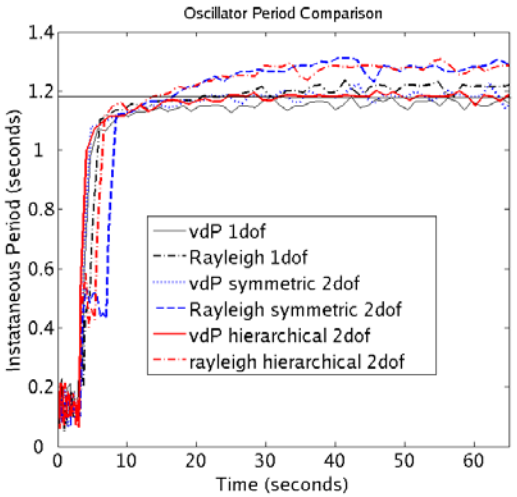


Figure 8: The instantaneous period of the mechanical system as driven by the different neural configurations. The Rayleigh oscillator consistently drives the system at a lower frequency than van der Pol, especially for the 2-DoF regimes. The mechanical system's natural period is denoted by the horizontal line at $T = 1.181\text{Hz}$.

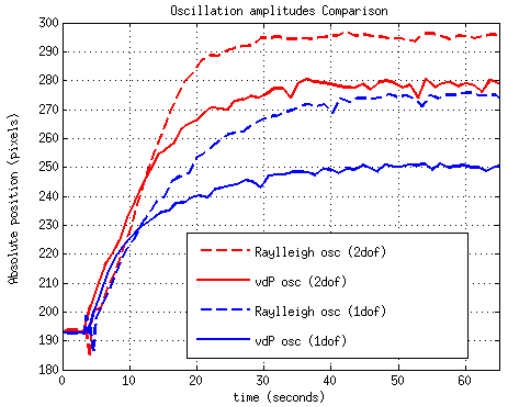


Figure 9: The envelopes of oscillation for four trials. The Rayleigh oscillator has a longer rise time than the vdP, but consistently reaches a larger oscillation amplitude given the same parameters, for both the 1-dof and staged-release 2-dof cases.

5 Conclusions

Our experiments have shown that even though the differences between the two oscillators studied are small, the nature of their dynamics altered the high-level behaviour of the system. Given the same experimental parameters, Rayleigh attained larger oscillation amplitudes for the mechanical system, at each morphological configuration. It also consistently forced the system to oscillate at frequencies lower than van der Pol. However, van der Pol starts large amplitude relaxation oscillations faster, attaining better performance during the first half of the transient period. This trade-off however is of limited scope, as it is of a fixed-offset nature; once Rayleigh has matched vdP's amplitude, it maintains its superior performance.

These experiments have shown that the effect of different oscillators, despite their great similarities are more pronounced than originally expected. Conversely, differing topologies that were expected to lead to stronger suboptimality had less effect.

References

A. C. de Pina Filho, M. S. Dutra, et al. (2005). Modeling of a bipedal robot using mutually coupled Rayleigh oscillators. *Biological Cybernetics* **92**: 1-7.

M. S. Dutra, A. C. de Pina Filho, et al. (2003). Modeling of a bipedal locomotor using coupled nonlinear oscillators of Van der Pol. *Biological Cybernetics* **88**(4): 286-292.

M. Lungarella and L. Berthouze (2002). On the Interplay Between Morphological, Neural, and Environmental Dynamics: A Robotic Case Study. *Adaptive Behavior* **10**(3/4): 223-242.

M. Lungarella and L. Berthouze (2004). Robot Bouncing: On the Synergy Between Neural and Body-Environment Dynamics. *Lecture Notes in Computer Science*(3139): 86-97.

K. Matsuoka (1985). Sustained oscillations generated by mutually inhibiting neurons with adaptation. *Biological Cybernetics* **52**: 367-376.

K. Matsuoka, N. Ohyama, et al. (2005). Control of a Giant Swing Robot Using a Neural Oscillator. *Lecture Notes in Computer Science*: 274-282.

P. Roy and Y. Demiris (2005). *Analysis of biped gait patterns generated by van der Pol and Rayleigh oscillators under feedback.* 3rd International Symposium on Adaptive Motion in Animals and Machines (AMAM2005), pp. Ilmenau, Germany, in Proceedings Of AMAM2005,

G. Taga (1991). Self-organised control of bipedal locomotion by neural oscillators in unpredictable environment. *Biological Cybernetics* **65**: 147-159.

G. Taga (1995). A model of the neuro-musculo-skeletal system for human locomotion. I. Emergence of basic gait. *Biological Cybernetics* **73**(2): 97.

P. Veskos and Y. Demiris (2005a). *Developmental acquisition of entrainment skills in robot swinging using van der Pol oscillators*. The 5th International Workshop on Epigenetic Robotics (EPIROB-2005), pp. 87-93, Nara, Japan, in Lund University Cognitive Studies, 123.

P. Veskos and Y. Demiris (2005b). *Robot swinging using van der Pol nonlinear oscillators*. 3rd International Symposium on Adaptive Motion in Animals and Machines (AMAM2005), pp. Ilmenau, Germany, in Proceedings Of AMAM2005,

M. M. Williamson (1998). Neural control of rhythmic arm movements. *Neural Networks* **11**(7/8): 1379-1394.

T. Zielinska (1996). Coupled oscillators utilised as gait rhythm generators of a two-legged walking machine. *Biological Cybernetics* **74**(3): 263-273.

Author Index