

Cognitieve
Kunstmatige
Intelligentie

Cognitieve
Kunstmatige
Intelligentie

Cognitieve
Kunstmatige
Intelligentie

Cognitieve
Kunstmatige
Intelligentie

Cognitieve
Kunstmatige
Intelligentie

Cognitieve
Kunstmatige
Intelligentie



Robo Sapiens

Proceedings of the First Dutch Symposium
on Embodied Intelligence

de Back, van der Zant, Zwanepol (eds.)

Preprint nr. 024 April 2001

Artificial Intelligence Preprint Series

© 2000, Onderwijsinstituut CKI - Utrecht University

ISBN 90-393-2730-0

ISSN 1389-5184

Prof.dr. A. Visser, Editor

Onderwijsinstituut CKI

Utrecht University

Heidelberglaan 8

3584 CS Utrecht

The Netherlands

Proceedings of the
First Dutch Symposium on
Embodied Intelligence

ROBO SAPIENS

Walter de Back
Tijn van der Zant
Lars Zwanepol
(Eds.)

April 4th, 2001

TABLE OF CONTENTS

- Table of Contents
- Preface

Section I: ROBO SAPIENS

- Introduction
- Robotics: Philosophy of Mind using a Screwdriver
Inman Harvey
- Embodied Cognitive Science
Rens Kortmann
- Evolutionary Robotics
Dario Floreano
- First Evolution Experiments on a Physical CAM-Brain Machine
Hugo de Garis et al.
- Building Gods or our Potential Exterminators
Hugo de Garis

Section II: ROBO PLAZA

- Introduction
- Groningen RoboCup Team
Erwin Mulder et al.
- Clockwork Orange: the Dutch RoboCup Team
Marco Wiering et al.
- GMD Robots Powered by Dual Dynamics
Hans-Ulrich Kobiak et al.
- Kennistechnologie in Uitvoering
Bolesian
- Adaptive Grounding of Lexicons on Mobile Robots
Paul Vogt
- Light-Switch Problem: Evolution of Adaptive Synapses
Dario Floreano et al.

Proceedings of the
First Dutch Symposium on
Embodied Intelligence

ROBO SAPIENS

Preface

Embodied Intelligence is a relatively new area of interest. In contrast to the traditional artificial intelligence (GOFAI), it approaches intelligence as an emergent property of the interaction of embodied systems with their environment. Rather than applying the 'divide-and-conquer' (modularity) principle to investigate the high-level intelligence of the human mind (like using language, expert knowledge, planning), the embodied intelligence community investigates the low-level intelligence (like avoiding obstacles, reactive/adaptive control) through the construction of complete embodied systems.

Those embodied systems are physical or simulated robots which are set in a complex dynamical environment. These systems are equipped with their own sensors, control system and actuators. They are situated systems because their reactive and/or adaptive movement is due to its own (imprecise) sensation of the environment. This kind of research is inspired by and relevant to research areas that differ from traditional AI. These include, but are not limited to: electronics, mechanics, computer science, artificial intelligence, psychology, bio-robotics, neuro-ethology, evolutionary computation, and ultimately to the philosophy of mind.

A particularly interesting (and new) topic is evolutionary robotics. This approach is heavily inspired by biology and evolutionary computation. It constructs robotic systems that are governed by embedded neural networks. These networks are optimised by way of artificial evolution through the application of genetic algorithms to the network's connections weights and/or topology. This is a radically new way of engineering the behaviour of robots and provides a new perspective for robot engineers, (neuro-)ethologists, people interested in evolutionary computation, and people interested in artificial intelligence in general.

Robo Sapiens is our term for biologically inspired intelligent embodied systems. We use the Linnaeus terminology to indicate that we are interested in biological(ly inspired) systems. Not only in biological systems themselves, but also in synthesizing all kind of life-like properties in robots. Robo Sapiens associates to the name of the modern human species, Homo Sapiens. By this, we indicate that we feel this line of research will eventually result in the construction of robots that are capable of exhibit human-like intelligence, perhaps even beyond this level.

With this symposium, the organisation aims to educate students on this new AI and its related fields and spread the conviction that this approach to intelligence is a promising one. Furthermore, we hope to contribute to the knowledge of the AI researchers present at Robo Sapiens and to promote this line of research in their work.

We feel Embodied Intelligence has been disregarded by (GOF)AI as a 'purely engineering' science and therefore has been underestimated in its potential to explain and synthesize intelligence. With the organisation of this First Dutch Symposium on this topic, we hope to spread some light onto these issues and wish to bring into existence a wide and active Dutch research community on Embodied Intelligence.

Acknowledgements

We thank our chairman, prof.dr.ir. Wim van de Grind, our lecturers, Rens Kortmann, dr. Iman Harvey, prof. Dario Floreano and prof. Hugo de Garis for their time, effort and their contribution to Robo Sapiens and these proceedings. We thank you for the articles and the permissions to print them.

We thank all the people that provided demonstrations on Robo Plaza: dr. Erwin Mulder, dr. Paul Vogt, Jesper Blynell, Claudio Matussi,

Jean-Christophe Zufferey, prof. Dario Florano, dr. Hans-Ulrich Kobiacka, Peter Schöll, dr. Marco Wiering et al. (the Dutch RoboCup Team), the team of students CKI-20b and the people from Bolesian. We acknowledge the great efforts that go into getting a robot to work when you want it to; like most creatures, they usually refuse to cooperate what-so-ever. We also thank the institutes where the development of the demos was conducted for providing the time and environment: AS-RUG, IKAT-UM (actually, the research was conducted at the AILab of the Free University of Brussels), EAST-LAMI-EPFL, AiS-GMD, UU, UvA, TUD, Bolesian.

We thank our commercial sponsors: Bolesian (main sponsor), Cap Gemini - Ernst & Young, ABN-AMRO Bank, and Philips; and the contributors from within our university: the Department of Philosophy, Institute for Information and Computing Sciences and the University Utrecht. Thank you for making this event possible. We thank the Department of Philosophy for their generous offer and for the printing of these proceedings (esp. Maarten Janssen and Albert Visser, series editors of the pre-print series).

We thank the board of USCKI Incognito for their trust and support. Besides the many times we disagreed with each other, we actually agreed most of the time. Thanks for enabling us to use your resources and thanks for the financial support. Some special thanks go out to Bart Zonneveld for the large amount of work he put in designing the webpage, poster and flyers.

We thank all our the volunteers from USCKI Incognito for your help during the day.

We thank the people who made the ICS RoboLab possible, prof. John-Jules Meijer and dr. Wiebe van der Hoek, and the students working in the ICS RoboLab, especially Arne Koopman for his technical support. Also thanks to: Ann Griffith from AAI Systems (permission on printing Harvey and Floreano) and prof. Frans Groen from the University of Amsterdam (who kindly lend us his Aibo).

Organisation

The First Dutch Symposium on Embodied Intelligence: Robo Sapiens has been organised by students CKI - Cognitive Artificial Intelligence. It is a collaboration between the students from the students association USCKI Incognito, at the Department of Philosophy, and the ICS RoboLab, at the Institute for Information and Computing Sciences.

Organising Committee:

Walter de Back
Heleen Boland
Sieuwert van Otterloo
Simone Rauh
Mariette Stam (chair)
Tijn van der Zant
Bart Zonneveld
Lars Zwanepol

Program Committee:

Walter de Back
Tijn van der Zant
Lars Zwanepol

Sponsor Committee:

Heleen Boland
Simone Rauh

Commercial Sponsors:

Bolesian (main sponsor)
Cap Gemini - Ernst and Young
ABM-AMRO Bank
Philips

Other contributors:

Department of Philosophy
Institute for Information and Computing Sciences
University Utrecht

Section I:

ROBO SAPIENS

Inman Harvey
Rens Kortmann
Dario Floreano
Hugo de Garis

Artificial Intelligence is inspired by, related to, and relevant for a wide variety of research areas. This is true for both the traditional as the new approaches. However, these approaches focus on different issues of intelligence, have different sources of inspiration, and rely on different methods, techniques, and technology. Whereas traditional AI is closely related to formal logic, linguistics, cognitive science and computing sciences, New AI is related to neuro-ethology, embodied cognitive science, evolutionary computing and robotics. At the symposium Robo Sapiens, New AI is discussed from some different points of view; the philosophy, the cooperation with biology, new techniques in evolutionary robotics, new technologies, and future possibilities.

Inman Harvey points out that every effort to building robotic systems is underpinned with some philosophical perspective. In contrast to traditional AI systems, modern roboticists do not adhere to the Cartesian split between body and mind. They construct complete systems, where no distinct separation can be made between the physical and non-physical (if there were such a thing). This is coherent with the scientific points of view in biology.

Rens Kortmann draws attention to the new science of Embodied Cognitive Science. This area of interest has an synthetic or constructivistic approach. It can overcome limitations of sciences in which behavior is measured and modelled in animals (like neuro-ethology and psychology) by employing robots to model aspects of animal behaviour. Results of these robotic experiments can serve as feedback to correct or enhance biological and psychological models in a comparative fashion.

Dario Floreano describes the field of Evolutionary Robotics in which robotic neural network controllers (and sometimes the robots morphology) are evolved by way of artificial evolution. This field can be separated in an engineering approach (goal-directed evolution of hand-made controllers), artificial life (evolving life-like properties without designated goal), and synthetic biology (gaining insight in natural mechanisms by evolution of complex controllers). This field is rapidly gaining interest from various sciences and seems to promise the feasibility of very complex robotic systems.

Hugo de Garis pushes the boundaries of current science one step further. Instead of evolving neural networks of several tens of neurons, as is commonly done in Evolutionary Robotics, he builds Artificial Brains. These are neural networks consisting of millions of neurons that are evolved/grown by artificial evolution on cellular automata in thousands of modules using modern field-programmable gate array (FPGA) technology. The brain is evolved (and stored in RAM) in a CAM-Brain Machine. De Garis has recently done his very first test experiments on this revolutionary new computer.

21st century technology will provide ever smaller and faster computers. Assuming this trend will continue and some break-throughs will occur in modern (evolutionary) robotics, robots could get incredibly complex machines. Hugo de Garis, and Moravec and Kurzeil with him, state that these robots could become many times more intelligent than their human inventors. About the question whether this is a desirable development, some foresee major problems for human society as a whole. Some will say creating machines with godlike intelligence is a wonderful thing, while others will fear the annihilation of mankind.

Robotics: Philosophy of Mind using a Screwdriver

Inman Harvey

School of Cognitive and Computing Sciences
University of Sussex
Brighton BN1 9QH, UK

Abstract

The design of autonomous robots has an intimate relationship with the study of autonomous animals and humans — robots provide a convenient puppet show for illustrating current myths about cognition. Like it or not, any approach to the design of autonomous robots is underpinned by some philosophical position in the designer. Whereas a philosophical position normally has to survive in debate, in a project of building situated robots one's philosophical position affects design decisions and is then tested in the real world — “doing philosophy of mind with a screwdriver”.

Traditional Good Old Fashioned Artificial Intelligence (GOFAI) approaches have been based on what is commonly called a Cartesian split between body and mind — though the division goes back at least to Plato. The Dynamical Systems approach to cognition, and to robot design, draws on other philosophical paradigms. We shall discuss how such varied philosophers as Heidegger, Merleau-Ponty or Wittgenstein, in the improbable event of them wanting to build robots, might be tempted to set about the task.

1 Introduction

Car manufacturers need robots that reliably and mindlessly repeat sequences of actions in some well-organised environment. For many other purposes autonomous robots are needed that will behave appropriately in a disorganised environment, that will react adaptively when faced with circumstances that they have never faced before. Planetary exploring robots, such as the Sojourner robot sent to Mars, cannot afford to wait the long time needed for

radio communication with people on earth for consultation on every individual move they make. The user of a semi-autonomous wheelchair should be able to delegate the same sort of decisions that a horse rider delegates to her horse — how to manoeuvre around obstacles and react instinctively to other traffic. We want such robots to behave to some extent intelligently, or adaptively — in fact to behave in some small part as if they had a mind of their own.

It has been tempting to think of this as ‘merely’ a technical, scientific problem - we should study in an objective, scientific fashion the basic requirements for adaptive intelligence, and then systematically engineer into our robots what we have found to be necessary. But like it or not, any approach to the understanding of cognition and adaptive intelligence, and hence to the design of autonomous robots, is inevitably framed within some philosophical position in the scientist or designer. In a project of building situated robots one’s philosophical position affects design decisions and is then tested in the real world — “doing philosophy of mind with a screw-driver”.

We use basic working metaphors to make sense of scientific theories; billiard balls and waves on a pond have been much used in physics. The metaphor of animals or even humans as machines, as comparable to the technical artefacts that we construct, is a powerful one. When we try to build autonomous robots they are almost literally puppets acting to illustrate our current myths about cognition. The word ‘myth’ sounds possibly derogatory as it often implies a fiction or half-truth; it is not intended as such here. I am merely trying to emphasise that our view of cognition is a human-centred view, from the end of the second millennium, some 4 billion years after the origin of life on this planet.

Someone coming from the conventional scientific perspective may suggest that our current cultural context is irrelevant. After all, objectivity in science is all to do with discounting the accidental perspectives of an observer and discovering universal facts and laws that all observers can agree on, from whatever place and time in which they are situated. However, to pursue this line too far leads one into a paradox. What theories (if any) did the organisms of 2 billion years ago have about the cognitive abilities of their contemporaries? Clearly nothing like ours. What theories might our descendants (if any) 2 billion years hence have about the cognition of *their* contemporaries? It would be arrogant, indeed unscientific, to assume that they would be similar to ours.

The Copernican revolutions in science have increased the scope of our objective understanding of the world by recognising that our observations are not context-free. Copernicus and Galileo used their imagination, and

speculated what the solar system might look like if our view from the planet Earth was not a privileged view from the fixed centre of the universe, but merely one possible perspective amongst many. Darwin opened up the way to appreciating that *Homo Sapiens* is just one species amongst many, with a common mode of evolutionary development from one common origin. Einstein brought about a fresh Copernican revolution with Special Relativity, showing how our understanding is increased when we abandon the idea of some unique fixed frame of reference for measuring the speed of any object. The history of science shows a number of advances, now generally accepted, stemming from a relativist perspective that (surprisingly) is associated with an objective stance toward our role as observers.

Cognitive science seems one of the last bastions to hold out against a Copernican, relativist revolution. In this paper I will broadly distinguish between the pre-Copernican views associated with the computationalist approach of classical Good Old Fashioned Artificial Intelligence (GOFAI), and the contextual, situated approaches of *nouvelle AI*. The two sides will be rather crudely portrayed, with little attempt to distinguish the many differing factions that can be grouped under one flag or the other. The different philosophical views will be associated with the direct implications that they have for the design of robots. It is worth mentioning that Brooks' recent collection of his early papers on robotics (Brooks, 1999) explicitly divides the eight papers into four under the heading of 'Technology' and four as 'Philosophy' - though the division is somewhat arbitrary as the two aspects go together throughout.

2 Cartesian or Classical approaches to Robotics

Descartes, working in the first half of the seventeenth century, is considered by many to be the first modern philosopher. A scientist and mathematician as much as philosopher, his ideas laid the groundwork for much of the way we view science today. In cognitive science the term 'Cartesian' has, perhaps rather unfairly to Descartes, come to exclusively characterise a set of views that treat the division between the mental and the physical as fundamental — the Cartesian cut (Lemmen, 1998). One form of the Cartesian cut is the dualist idea that these are two completely separate substances, the mental and the physical, which can exist independently of each other. Descartes proposed that these two worlds interacted in just one place in humans, the pineal gland in the brain. Nowadays this dualism is not very respectable, yet the common scientific assumption rests on a variant of this Cartesian cut: that the physical world can be considered completely

objectively, independent of all observers.

This is a different kind of objectivity from that of the Copernican scientific revolutions mentioned above. Those relied on the absence of any privileged position, on intersubjective agreement between observers, independent of any specific observer. The Cartesian objectivity assumes that there just is a way the world is, independent of any observer at all. The scientist's job, then, is to be a spectator from outside the world, with a God's-eye view from above.

When building robots, this leads to the classical approach where the robot is also a little scientist-spectator, seeking information (from outside) about how the world is, what objects are in which place. The robot takes in information, through its sensors; turns this into some internal representation or model, with which it can reason and plan; and on the basis of this formulates some action that is delivered through the motors. Brooks calls this the SMPA, or sense-model-plan-act architecture (Brooks, 1999).

The 'brain' or 'nervous system' of the robot can be considered as a Black Box connected to sensors and actuators, such that the behaviour of the machine plus brain within its environment can be seen to be intelligent. The question then is, 'What to put in the Black Box?' The classical computationalist view is that it should be computing appropriate outputs from its inputs. Or possibly they may say that whatever it is doing should be *interpretable* as doing such a computation.

The astronomer, and her computer, perform computational algorithms in order to predict the next eclipse of the moon; the sun, moon and earth do not carry out such procedures as they drift through space. The cook follows the algorithm (recipe) for mixing a cake, but the ingredients do not do so as they rise in the oven. Likewise if I was capable of writing a computer program which predicted the actions of a small creature, this does not mean that the creature itself, or its neurons or its brain, was consulting some equivalent program in 'deciding what to do'.

Formal computations are to do with solving problems such as 'when is the eclipse?'. But this is an astronomer's problem, not a problem that the solar system faces and has to solve. Likewise, predicting the next movement of a creature is an animal behaviourist's problem, not one that the creature faces. However, the rise of computer power in solving problems naturally, though regrettably, led AI to the view that cognition equalled the solving of problems, the calculation of appropriate outputs for a given set of inputs. The brain, on this view, was surely some kind of computer. What was the problem that the neural program had to solve? — the inputs must be sensory, but what were the outputs?

Whereas a roboticist would talk in terms of motor outputs, the more

cerebral academics of the infant AI community tended to think of plans, or representations, as the proper outputs to study. They treated the brain as the manager who does not get his own hands dirty, but rather issues commands based on high-level analysis and calculated strategy. The manager sits in his command post receiving a multitude of possibly garbled messages from a myriad sensors and tries to work out what is going on. Proponents of this view tend not to admit explicitly, indeed they often deny vehemently that they think in terms of a homunculus in some inner chamber of the brain, but they have inherited a Cartesian split between mind and brain and in the final analysis they rely on such a metaphor.

3 What is the Computer Metaphor?

The concepts of computers and computations, and programs, have a variety of meanings that shade into each other. On the one hand a computer is a formal system with the same powers as a Turing Machine (... assuming the memory is of adequate size). On the other hand a computer is this object sitting in front of me now, with screen and keyboard and indefinite quantities of software.

A program for the formal computer is equivalent to the pre-specified marks on the Turing machine's tape. For a given starting state of this machine, the course of the computation is wholly determined by the program and the Turing machine's transition table; it will continue until it halts with the correct answer, unless perhaps it continues forever — usually considered a *bad thing*!

On the machine on my desk I can write a program to calculate a succession of co-ordinates for the parabola of a cricket-ball thrown into the air, and display these both as a list of figures and as a curve drawn on the screen. Here I am using the machine as a convenient fairly user-friendly Turing machine.

However most programs for the machine on my desk are very different. At the moment it is (amongst many other things) running an editor or word-processing program. It sits there and waits, sometimes for very long periods indeed, until I hit a key on the keyboard, when it virtually immediately pops a symbol into an appropriate place on the screen; unless particular control keys are pressed, causing the file to be written, or edits to be made. Virtually all of the time the program is waiting for input, which it then processes near-instantaneously. In general it is a *good thing* for such a program to continue for ever, or at least until the exit command is keyed in.

The cognitivist approach asserts that something with the power of a

Turing machine is both necessary and sufficient to produce intelligence; both human intelligence and equivalent machine intelligence. Although not usually made clear, it would seem that something close to the model of a word-processing program is usually intended; i.e., a program that constantly awaits inputs, and then near-instantaneously calculates an appropriate output before settling down to await the next input. Life, so I understand the computationalists to hold, is a sequence of such individual events, perhaps processed in parallel.

4 Time in Computations and in Connectionism

One particular aspect of a computational model of the mind which derives from the underlying Cartesian assumptions common to traditional AI is the way in which the issue of *time* is swept under the carpet — only the sequential aspect of time is normally considered. In a standard computer operations are done serially, and the lengths of time taken for each program step are for formal purposes irrelevant. In practice for the machine on my desk it is necessary that the time-steps are fast enough for me not to get bored waiting. Hence for a serial computer the only requirement is that individual steps take as short a time as possible. In an ideal world any given program would be practically instantaneous in running, except of course for those unfortunate cases when it gets into an infinite loop.

The common connectionist assumption is that a connectionist network is in some sense a parallel computer. Hence the time taken for individual processes within the network should presumably be as short as possible. They cannot be considered as being effectively instantaneous because of the necessity of keeping parallel computations in step. The standard assumptions made fall into two classes.

1. The timelag for activations to pass from any one node to another it is connected to, including the time taken for the outputs from a node to be derived from its inputs, is in all cases exactly one unit of time (e.g. a back-propagation, or an Elman network).
2. Alternatively, just one node at a time is updated independently of the others, and the choice of which node is dealt with next is stochastic (e.g. a Hopfield net or a Boltzmann machine).

The first method follows naturally from the computational metaphor, from the assumption that a computational process is being done in parallel.

The second method is closer to a dynamical systems metaphor, yet still computational language is used. It is suggested that a network, after appropriate training, will when presented with a particular set of inputs then sink into the appropriate basin of attraction which appropriately classifies them. The network is used as either a distributed content-addressable memory, or as a classifying engine, as a module taking part in some larger-scale computation. The stochastic method of relaxation of the network may be used, but the dynamics of the network are thereby made relatively simple, and not directly relevant to the wider computation. It is only the stable attractors of the network that are used. It is no coincidence that the attractors of such a stochastic network are immensely easier to analyse than any non-stochastic dynamics.

It might be argued that connectionists are inevitably abstracting from real neural networks, and inevitably simplifying. In due course, so this argument goes, they will slowly extend the range of their models to include new dimensions, such as that of time. What is so special about time — why cannot it wait? Well, the simplicity at the formal level of connectionist architectures which need synchronous updates of neurons disguises the enormous complexity of the physical machinery needed to maintain a universal clock-tick over distributed nodes in a physically instantiated network. From the perspective advocated here, clocked networks form a particular complex subset of all real-time dynamical networks ones need be, and if anything *they* are the ones that should be left for later (van Gelder, 1992).

A much broader class of networks is that where the timelags on individual links between nodes is a real number which may be fixed or may vary in a similar fashion to weightings on such links¹. A pioneering attempt at a theory that incorporates such timelags as an integral part is given in (Malsburg and Bienenstock, 1986).

In neurobiological studies the assumption seems to be widespread that neurons are passing information between each other 'encoded' in the rate of firing. By this means it would seem that real numbers could be passed, even though signals passing along axons seem to be all-or-none spikes. This assumption is very useful, indeed perhaps invaluable, in certain areas such as early sensory processing. Yet it is perverse to assume that this is true throughout the brain, a perversity which while perhaps not caused by the computational metaphor is certainly aided by it. Experiments demonstrating that the individual timing of neuronal events in the brain, and the temporal coincidence of signals passing down separate 'synfire chains', can

¹For a simple model without loss of generality any time taken for outputs to be derived from inputs within a node can be set to zero, by passing any non-zero value on instead to the links connected to that node.

be of critical importance, are discussed in (Abeles, 1982).

5 What is a Representation?

The concept of symbolic reference, or representation, lies at the heart of analytic philosophy and of computer science. The underlying assumption of many is that a real world exists independently of any given observer; and that symbols are entities that can 'stand for' objects in this real world — in some abstract and absolute sense. In practice, the role of the observer in the act of representing something is ignored.

Of course this works perfectly well in worlds where there is common agreement amongst all observers — explicit or implicit agreement — on the usages and definitions of the symbols, and the properties of the world that they represent. In the worlds of mathematics, or formal systems, this is the case, and this is reflected in the anonymity of tone, and use of the passive tense, in mathematics. Yet the dependency on such agreement is so easily forgotten — or perhaps ignored in the assumption that mathematics is the language of God.

A symbol P is used by a person Q to represent, or refer to, an object R to a person S . Nothing can be referred to without somebody to do the referring. Normally Q and S are members of a community that have come to agree on their symbolic usages, and training as a mathematician involves learning the practices of such a community. The vocabulary of symbols can be extended by defining them in terms of already-recognised symbols.

The English language, and the French language, are systems of symbols used by people of different language communities for communicating about their worlds, with their similarities and their different nuances and clichés. The languages themselves have developed over thousands of years, and the induction of each child into the use of its native language occupies a major slice of its early years. The fact that, nearly all the time we are talking English, we are doing so to an English-speaker (including when we talk to ourselves), makes it usually an unnecessary platitude to explicitly draw attention to the community that speaker and hearer belong to.

Since symbols and representation stand firmly in the linguistic domain, another attribute they possess is that of arbitrariness (from the perspective of an observer external to the communicators). When I raise my forefinger with its back to you, and repeatedly bend the tip towards me, the chances are that you will interpret this as 'come here'. This particular European and American sign is just as arbitrary as the Turkish equivalent of placing the hand horizontally facing down, and flapping it downwards. Different actions

or entities can represent the same meaning to different communities; and the same action or entity can represent different things to different communities. In Mao Tse-Tung's China a red traffic light meant *GO*.

In the more general case, and particularly in the field of connectionism and cognitive science, when talking of representation it is imperative to make clear who the users of the representation are; and it should be possible to at a minimum suggest how the convention underlying the representation arose. In particular it should be noted that where one and the same entity can represent different things to different observers, conceptual confusion can easily arise. When in doubt, always make explicit the Q and S when P is used by Q to represent R to S .

In a computer program a variable `pop_size` may be used by the programmer to represent (to herself and to any other users of the program) the size of a population. Inside the program a variable i may be used to represent a counter or internal variable in many contexts. In each of these contexts a metaphor used by the programmer is that of the program describing the actions of various homunculi, some of them keeping count of iterations, some of them keeping track of variables, and it is within the context of particular groups of such homunculi that the symbols are representing. But how is this notion extended to computation in connectionist networks?

6 Representation in Connectionism

When a connectionist network is being used to do a computation, in most cases there will be input, hidden and output nodes. The activations on the input and output nodes are decreed by the connectionist to represent particular entities that have meaning for her, in the same way as `pop_size` is in a conventional program. But then the question is raised — 'what about internal representations?'.

If a connectionist network is providing the nervous system for a robot, a different interpretation might be put on the inputs and outputs. But for the purpose of this section, the issues of internal representation are the same.

All too often the hidden agenda is based on a Platonic notion of representation — what do activations or patterns of activations represent in some absolute sense to God? The behaviour of the innards of a trained network are analysed with the same eagerness that a sacrificed chicken's innards are interpreted as representing ones future fate. There is however a more principled way of talking in terms of internal representations in a network, but a way that is critically dependent on the observer's decomposition of that

network. Namely, the network must be decomposed by the observer into two or more modules that are considered to be communicating with each other by means of these representations.

Where a network is explicitly designed as a composition of various modules to do various subtasks (for instance a module could be a layer, or a group of laterally connected nodes within a layer), then an individual activation, or a distributed group of activations, can be deemed to represent an internal variable in the same way that i did within a computer program. However, unlike a program which wears its origins on its sleeve (in the form of a program listing), a connectionist network is usually deemed to be internally 'nothing more than' a collection of nodes, directed arcs, activations, weights and update rules. Hence there will usually be a large number of possible ways to decompose such a network, with little to choose between them; and it depends on just where the boundaries are drawn just who is representing what to whom.

It might be argued that some ways of decomposing are more 'natural' than others; a possible criterion being that two sections of a network should have a lot of internal connections, but a limited number of connecting arcs between the sections. Yet as a matter of interest this does not usually hold for what is perhaps the most common form of decomposition, into layers. The notion of a distributed representation usually refers to a representation being carried in parallel in the communication from one layer to the next, where the layers as a whole can be considered as the Q and S in the formula " P is used by Q to represent R to S ".

An internal representation, according to this view, only makes sense relative to a particular decomposition of a network chosen by an observer. To assert of a network that it contains internal representations can then only be justified as a rather too terse shorthand for asserting that the speaker proposes some such decomposition. Regrettably this does not seem to be the normal usage of the word in cognitive science, yet I am not aware of any well-defined alternative definition.

7 Are Representations Needed?

With this approach to the representation issue, then any network can be decomposed (in a variety of ways) into separate modules that the observer considers as communicating with each other. The interactions between such modules can *ipso facto* be deemed to be mediated by a representation. Whether it is useful to do so is another matter.

Associated with the metaphor of the mind (or brain, or an intelligent

machine) as a computer go assumptions of functional decomposition. Since a computer formally manipulates symbols, yet it is light waves that impinge on the retina or the camera, surely (so the story goes) some intermediate agency must do the necessary translating. Hence the traditional decomposition of a cognitive system into a perception module, which takes sensory inputs and produces a world model; this is passed onto a central planning module which reasons on the basis of this world model; passing on its decisions to an action module which translates them into the necessary motor actions. This functional decomposition has been challenged, and an alternative behavioural decomposition proposed, by Brooks in, e.g., (Brooks, 1999).

In particular, the computationalist or cognitivist approach seems to imply that communication between any such modules is a one-way process; any feedback loops are within a module. Within for instance back-propagation, the backward propagation of errors to adjust weights during the learning process is treated separately from the forward pass of activations. This helps to maintain the computational fiction, by conceptually separating the two directions, and retaining a feed-forward network. But consider the fact that within the primate visual processing system, when visualised as a network, there are many more fibres coming 'back' from the visual cortex into the Lateral Geniculate Nucleus (LGN) than there are fibres going from the retina to the LGN in the 'correct' direction. How does the computationalist make sense of this?

Marr (in (Marr, 1977), reprinted in (Boden, 1990)) classifies AI theories into Type 1 and Type 2, where a Type 2 theory can only solve a problem by the simultaneous action of a considerable number of processes, *whose interaction is its own simplest description*. It would seem that type 2 systems can only be decomposed arbitrarily, and hence the notion of representation is less likely to be useful. This is in contrast to a Type 1 theory, where a problem can be decomposed into a form that an algorithm can be formulated to solve, by *divide and conquer*. Type 1 theories are of course the more desirable ones when they can be found, but it is an empirical matter whether they exist or not. In mathematics the 4-colour theorem has been solved in a fashion that requires a large number of special cases to be exhaustively worked out in thousands of hours of computation (Appel and Haken, 1989). It is hoped that there were no hardware faults during the proof procedure, and there is no way that the proof as a whole can be visualised and assessed by a human. There is no *a priori* reason why the workings of at least parts of the brain should not be comparably complex, or even more so². This

²For the purposes of making an intelligent machine or robot, it has in the past seemed obvious that only Type 1 techniques could be proposed. However evolutionary techniques

can be interpreted as: there is no *a priori* reason why all parts of the brain should be in such a modular form that representation-talk is relevant. The answer to the question posed in the title of this section is *no*. This does not rule out the possibility that in some circumstances representation-talk *might* be useful, but it is an experimental matter to determine this.

8 Alternatives to Cartesianism

Hubert Dreyfus came out with a trenchant criticism of the classical AI computationalist approach to cognition in the 1960s. He came from a very different set of philosophical traditions, looking for inspiration to twentieth century philosophers such as Heidegger, Merleau-Ponty and Wittgenstein. Initially he produced a report for the RAND Corporation provocatively entitled 'Alchemy and Artificial Intelligence' in 1965 (Dreyfus, 1965). Later, more popular, books are (Dreyfus, 1972) and with his brother (Dreyfus and Dreyfus, 1986). These are amongst the easiest ways for somebody with a conventional background in computer science, cognitive science or robotics to approach the alternative set of philosophical views. Nevertheless, the views are sufficiently strange to those brought up into the Cartesian way of thinking that at first sight Dreyfus appears to be mystical, or anti-scientific. This is not the case.

Another view of cognition from a phenomenological or Heideggerian perspective is given in (Winograd and Flores, 1986); Winograd was instrumental in some of the classical early GOFAL work, before coming around to a very different viewpoint. A different Heideggerian perspective is given in (Wheeler, 1996) within (Boden, 1996). The relevance of Merleau-Ponty is drawn out in (Lemmen, 1998). A different perspective that is similarly opposed to the Cartesian cut is given in (Maturana and Varela, 1987; Varela et al., 1991). A much more general textbook on robotics that is written from a situated and embodied perspective is (Pfeifer and Scheier, 1999).

Heidegger rejects the simplistic objective view, that the objective physical world is the primary reality that we can be certain of. He similarly rejects the opposite idealistic or subjective view, that our thoughts are the primary reality. Instead, the primary reality is our experience of the world, that cannot exist independently of one or the other. Our everyday practical lived experience, as we reach for our coffee or switch on the light, is more fundamental than the detached theoretical reflection that we use as rational scientists. Though Heidegger himself would not put it this way, this makes sense from a Darwinian evolutionary perspective on our own species. From

need not restrict themselves in this fashion (Harvey et al., 1997).

this perspective, our language using and reasoning powers probably arrived in *H. sapiens* over just the last million or so years in our 4 billion year evolutionary history, and is merely the thin layer of icing on the cake. From both a phylogenetic and an ontogenetic view, we are organisms and animals first, reasoning humans only later.

As humans, of course, detached theoretical reasoning is one of our hallmarks; indeed the Darwinian view presented in the previous paragraph is just one such piece of reasoning. However practical know-how is more fundamental than such detached knowing-that. This is a complete reversal of the typical approach of a Cartesian cognitive scientist or roboticist, who would attempt to reduce the everyday action of a human (or robot) reaching for the coffee mug into a rational problem-to-be-solved: hence the Cartesian sense-model-plan-act cycle.

The archetypal Heideggerian example is that of hammering in a nail. When we do this normally, the arm with the hammer naturally and without thought goes through its motions, driving the hammer home. It is only when something goes wrong, such as the head of the hammer flying off or the nail bending in the wood, that we have to concentrate and start reflecting on the situation, rationalising what the best plan of action will be. Information processing, knowing-that, is secondary and is built on top of our everyday rhythms and practices of practical know-how - know-how which cannot be reduced to a set of rules that we implement. This is true, Wittgenstein suggests, even for our language skills:

In general we don't *use* language according to strict rules - it hasn't been taught us by means of strict rules either. (Wittgenstein 1960:25)

For the roboticist, this anti-Cartesian alternative philosophy seems at first sight negative and unhelpful. For everyday robot actions this implies that we should do without planning, without the computational model, without internal representations, but nothing has yet been offered to replace such methods. The two lessons that need to be learnt initially is that cognition is

- Situated: a robot or human is always already in some situation, rather than observing from outside
- and Embodied: a robot or human is a perceiving body, rather than a disembodied intelligence that happens to have sensors.

One nice example of a situated embodied robot is the simple walking machine of McGeer which uses 'passive dynamic walking' (McGeer, 1990;

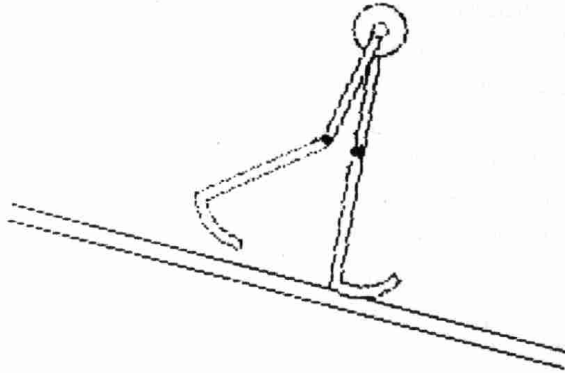


Figure 1: *McGeer's 'passive dynamic walker'.*

McGeer, 1993). This is a two-legged walking robot with each leg made of just an upper and lower limb connected by a knee joint. This knee acts as a human knee, allowing bending freely in one direction but not in the other. At the bottom of each lower limb is a foot shaped in an arc, and the two legs are hinged together at the waist. The dimensions are carefully worked out, but then that is the complete walking robot with no control system needed at all. When it is started off on a gently sloping incline, the legs will walk along in a remarkably natural human-like motion — all the control has come from the natural dynamics, through being situated and embodied.

The walking robot does not take in information through sensors, and does not compute what its current position is and what its next move should be. The designers, of course, carefully computed the appropriate dimensions. In the natural world, organisms and animals have their bodily dimensions designed through natural evolution.

9 The Dynamical Systems alternative

This last example is a robot designed on non-Cartesian principles, using an alternative view that has gained favour in the last decade within AI circles, though its origins date back at least to the early cybernetics movement. One description of this is the Dynamical Systems view of cognition:

... animals are endowed with nervous systems whose dynamics are such that, when coupled with the dynamics of their bodies and environments, these animals can engage in the patterns of

behavior necessary for their survival. (Beer & Gallagher 1992: 91)

At this stage we downgrade the significance of *intelligence* for AI in favour of the concept of *adaptive behaviour*. Intelligence is now just one form of adaptive behaviour amongst many; the ability to reason logically about chess problems may be adaptive in particular refined circles, but the ability to cross the road safely is more widely adaptive. We should note the traditional priorities of AI: the computationalists' emphasis on reasoning led them to assume that everyday behaviour of sensorimotor coordination must be built on top of a reasoning system. Sensors and motors, in their view, are 'merely' tools for information-gathering and plan-execution on behalf of the central executive where the real work is done. Many proponents of an alternative view, including myself, would want to turn this on its head: logical reasoning is built on top of linguistic behaviour, which is built on prior sensorimotor abilities. These prior abilities are the fruit of billions of years of evolution, and language has only been around for the last few tens of thousands of years.

A dynamical system is formally any system with a finite number of state variables that can change over time; the rate of change of any one such variable depends on the current values of any or all of the variables in a regular fashion. These regularities are typically summed up in a set of differential equations. A Watt governor for a steam engine is a paradigmatic dynamical system (van Gelder, 1992), and we can treat the nervous system plus body of a creature (or robot) as one also. The behaviour of a dynamical system such as the governor depends also on the current value of its external inputs (from the steam engine) which enter the relevant differential equations as parameters. In a complementary way, the output of the governor acts as a parameter on the equations which describe the steam engine itself as a dynamical system. One thing that is very rapidly learnt from hands-on experience is that two such independent dynamical systems, when coupled together into (e.g.) steam-engine-plus-governor treated now as a single dynamical system, often behave in a counterintuitive fashion not obviously related to the uncoupled behaviours.

Treating an agent — creature, human or robot — as a dynamical system coupled with its environment through sensors and motors, inputs and outputs, leads to a metaphor of agents being *perturbed* in their dynamics through this coupling, in contrast to the former picture of such agents *computing* appropriate outputs from their inputs. The view of cognition entailed by this attitude fits in with Varela's characterisation of cognition as *embodied action*:

By using the term *embodied* we mean to highlight two points: first, that cognition depends upon the kinds of experience that come from having a body with various sensorimotor capacities, and second, that these individual sensorimotor capacities are themselves embedded in a more encompassing biological, psychological and cultural context. By using the term *action* we mean to emphasise once again that sensory and motor processes, perception and action, are fundamentally inseparable in lived cognition. Indeed, the two are not merely contingently linked in individuals; they have also evolved together. (Varela et al., 1991: 172–173)

10 Evolutionary Robotics and Behaviourism

Moving from natural agents to artificial robots, the design problem that a robot builder faces is now one of creating the internal dynamics of the robot, and the dynamics of its coupling, its sensorimotor interactions with its environment, such that the robot exhibits the desired behaviour in the right context. Designing such dynamical systems presents problems unfamiliar to those who are used to the computational approach to cognition.

A primary difference is that dynamics involves time, real time. Whereas a computation of an output from an input is the same computation whether it takes a second or a minute, the dynamics of a creature or robot has to be matched in timescale to that of its environment. A second difference is that the traditional design heuristic of *divide and conquer* cannot be applied in the same way. It is not clear how the dynamics of a control system should be carved up into smaller tractable pieces; and the design of any one small component depends on an understanding of how it interacts in real time with the other components, such interaction possibly being mediated via the environment. This is true for behavioural decomposition of control systems (Brooks, 1999) as well as functional decomposition. However, Brooks' subsumption architecture approach offers a different design heuristic: first build simple complete robots with behaviours simple enough to understand, and then incrementally add new behaviours of increasing complexity or variety, one at a time, which subsume the previous ones. Before the designer adds a new control system component in an attempt to generate a new behaviour, the robot is fully tested and debugged for its earlier behaviours; then the new component is added so as to keep to a comprehensible and tractable minimum its effects on earlier parts.

This approach is explicitly described as being inspired by natural evolu-

tion; but despite the design heuristics it seems that there is a practical limit to the complexity that a human designer can handle in this way. Natural Darwinian evolution has no such limits, hence the more recent moves towards the artificial evolution of robot control systems (Harvey et al., 1997).

In this work a genetic encoding is set up such that an artificial genotype, typically a string of 0s and 1s, specifies a control system for a robot. This is visualised and implemented as a dynamical system acting in real time; different genotypes will specify different control systems. A genotype may additionally specify characteristics of the robot 'body' and sensorimotor coupling with its environment. When we have settled on some particular encoding scheme, and we have some means of evaluating robots at the required task, we can apply artificial evolution to a population of genotypes over successive generations.

Typically the initial population consists of a number of randomly generated genotypes, corresponding to randomly designed control systems. These are instantiated in a real robot one at a time, and the robot behaviour that results when placed in a test environment is observed and evaluated. After the whole population has been scored, their scores can be compared; for an initial random population one can expect all the scores to be abysmal, but some (through chance) are less abysmal than others. A second generation can be derived from the first by preferentially selecting the genotypes of those with higher scores, and generating offspring which inherit genetic material from their parents; recombination and mutation is used in producing the offspring population which replaces the parents. The cycle of instantiation, evaluation, selection and reproduction then continues repeatedly, each time from a new population which should have improved over the average performance of its ancestors. Whereas the introduction of new variety through mutation is blind and driven by chance, the operation of selection at each stage gives direction to this evolutionary process.

This evolutionary algorithm comes from the same family as Genetic Algorithms and Genetic Programming, which have been used with success on thousands of problems. The technique applied to robotics has been experimental and limited to date. It has been demonstrated successfully on simple navigation problems, recognition of targets, and the use of minimal vision or sonar sensing in uncertain real world environments (Harvey et al., 1997; Thompson, 1995). One distinguishing feature of this approach using 'blind' evolution is that the resulting control system designs are largely opaque and incomprehensible to the human analyst. With some considerable effort simple control systems can be understood using the tools of dynamical systems theory (Husbands et al., 1995). However, it seems inevitable that, for the same reasons that it is difficult to design *complex* dynamical systems, it is

also difficult to analyse them.

This is reflected in the methodology of Evolutionary Robotics which, once the framework has been established, concerns itself solely with the behaviour of robots: "if it walks like a duck and quacks like a duck, it is a duck". For this reason we have sometimes been accused of being 'the New Behaviourists'; but this emphasis on behaviour assumes that there are significant internal states³, and in my view is compatible with the attribution of adaptive intelligence. A major conceptual advantage that Evolutionary Robotics has over classical AI approaches to robotics is that there is no longer a mystery about how one can 'get a robot to have needs and wants'. In the classical version the insertion of a value function `robot_avoid_obstacle` often leaves people uncomfortable as to whether it is the robot or the programmer who has the desires. In contrast, generations of evolutionary selection that tends to eliminate robots that crash into the obstacle produces individual robots that do indeed avoid it; and here it seems much more natural that it is indeed the *robot* which has the desire.

11 Relativism

I take a Relativist perspective, which contrary to the naive popular view does not imply solipsism, or subjectivism, or an anything-goes attitude to science. The history of science shows a number of advances, now generally accepted, that stem from a relativist perspective which (surprisingly) is associated with an objective stance toward our role as observers. The Copernican revolution abandoned our privileged position at the centre of the universe, and took the imaginative leap of wondering how the solar system would look viewed from the Sun or another planet. Scientific objectivity requires theories to be general, to hold true independently of our particular idiosyncratic perspective, and the relativism of Copernicus extended the realm of the objective. Darwin placed humans amongst the other living creatures of the universe, to be treated on the same footing. With Special Relativity, Einstein carried the Copernican revolution further, by considering the viewpoints of observers travelling near to the speed of light, and insisting that scientific objectivity required that their perspectives were equally privileged to ours. Quantum physics again brings the observer explicitly into view.

³Not 'significant' in the sense of representational — internal states are mentioned here to differentiate evolved dynamical control systems (which typically have plenty of internal state) from those control systems restricted to feedforward input/output mappings (typical of 'reactive robotics').

Cognitive scientists must be careful above all not to confuse objects that are clear to them, that have an objective existence for them, with objects that have a meaningful existence for other agents. A roboticist learns very early on how difficult it is to make a robot recognise something that is crystal clear to us, such as an obstacle or a door. It makes sense for us to describe such an object as 'existing for that robot' if the physical, sensorimotor, coupling of the robot with that object results in robot behaviour that can be correlated with the presence of the object. By starting the previous sentence with "It makes sense for us to describe ..." I am acknowledging our own position here acting as scientists observing a world of cognitive agents such as robots or people; this objective stance means we place ourselves outside this world looking in as godlike creatures from outside. Our theories can be scientifically objective, which means that predictions should not be dependent on incidental factors such as the nationality or location or star-sign of the theorist.

When I see a red sign, this red sign is an object that can be discussed scientifically. This is another way of saying that it exists for me, for you, and for other human observers of any nationality; though it does not exist for a bacterium or a mole. We construct these objects from our experience and through our acculturation as humans through education⁴. Just as our capacity for language is phylogenetically built upon our sensorimotor capacities, so our objects, our scientific concepts, are built out of our experience. But our phenomenal experience itself cannot be an objective thing that can be discussed or compared with other things. It is primary, in the sense that it is only through having phenomenal experience that we can create things, objective things that are secondary.

12 Conclusion

Like it or not, any approach to the design of autonomous robots is underpinned by some philosophical position in the designer. There is no philosophy-free approach to robot design — though sometimes the philosophy arises through accepting unthinkingly and without reflection the approach within which one has been brought up. GOFAI has been predicated on some version of the Cartesian cut, and the computational approach has had enormous success in building superb tools for humans to use — but

⁴It makes no sense to discuss (...for *us humans* to discuss ...) the existence of objects in the absence of humans. And (in an attempt to forestall the predictable objections) this view does *not* imply that we can just posit the existence of any arbitrary thing as our whim takes us.

it is simply inappropriate for building autonomous robots.

There is a different philosophical tradition which seeks to understand cognition in terms of the priority of lived phenomenal experience, the priority of everyday practical know-how over reflective rational knowing-that. This leads to very different engineering decisions in the design of robots, to building *situated* and *embodied* creatures whose dynamics are such that their coupling with their world leads to sensible behaviours. The design principles needed are very different; Brooks' subsumption architecture is one approach, Evolutionary Robotics is another. Philosophy does make a practical difference.

Acknowledgments

I thank the EPSRC and the University of Sussex for funding, and Shirley Kitts for philosophical orientation.

References

- Abeles, M. (1982). *Local Cortical Circuits, an Electrophysiological Study*. Springer-Verlag.
- Appel, K. and Haken, W. (1989). Every planar map is four colorable. *American Mathematical Society, Contemporary Mathematics*, 98.
- Beer, R. D. and Gallagher, J. C. (1992). Evolving dynamic neural networks for adaptive behavior. *Adaptive Behavior*, 1(1):91-122.
- Boden, M. A. (1990). *The Philosophy of Artificial Intelligence* Oxford University Press.
- Boden, M. A. (1996). *The Philosophy of Artificial Life* Oxford University Press.
- Brooks, R. (1999). *Cambrian Intelligence: the Early History of the New AI*. MIT Press, Cambridge MA.
- Dreyfus, H. L. (1965). *Alchemy and Artificial Intelligence*. RAND Corporation Paper P-3244, December 1965.
- Dreyfus, H. L. (1972). *What Computers Can't Do: a Critique of Artificial Reason*. Harper and Row, New York.
- Dreyfus, H. L. and Dreyfus, S. E. (1986). *Mind Over Machine*.

- Harvey, I., Husbands, P., Cliff, D., Thompson, A., and Jakobi, N. (1997). Evolutionary robotics: the Sussex approach. *Journal of Robotics and Autonomous Systems* v. 20 (1997) pp. 205-224.
- Husbands, P., Harvey, I., and Cliff, D. (1995). Circle in the round: State space attractors for evolved sighted robots. *Journal of Robotics and Autonomous Systems. Special Issue on "The Biology and Technology of Intelligent Autonomous Agents"*, 15:83-106.
- Lemmen, R. (1998). Towards a Non-Cartesian Cognitive Science in the light of the philosophy of Merleau-Ponty. DPhil Thesis, University of Sussex.
- Malsburg, C. von der, and Bienenstock, E. (1986). Statistical coding and short-term plasticity: a scheme for knowledge representation in the brain. In Bienenstock, E., Fougelman-Soulie, F., and Wiesbuch, G. (eds.) *Disordered Systems and Biological Organization*. Springer-Verlag.
- Marr, D. C. (1977). Artificial Intelligence: a personal view, *Artificial Intelligence*, 9:37-48.
- Maturana, H. and Varela, F. (1987). *The Tree of Knowledge: The Biological Roots of Human Understanding*. Shambhala Press, Boston.
- McGeer, T. (1990). Passive dynamic walking. *Int. J. Robotics Research*, 1990(9)2:62-82.
- McGeer, T (1993). Dynamics and Control of Bipedal Locomotion. *J. Theor. Biol.*, 1993(163), 277-314.
- Pfeifer, R., and Scheier, C. (1999). *Understanding Intelligence*. MIT Press.
- Thompson, A. (1995). Evolving electronic robot controllers that exploit hardware resources. In Morán, F., Moreno, A., Merelo, J.J. and Chacón, P. (eds.) *Proceedings of Third European Conference on Artificial Life*. Springer-Verlag, pp. 640-656.
- van Gelder, T. (1992). What might cognition be if not computation. Technical Report 75, Indiana University Cognitive Sciences. Reprinted in *Journal of Philosophy* 92:345-381 (1995).
- Varela, F., Thompson, E., and Rosch, E. (1991). *The Embodied Mind*. MIT Press.

- Wheeler, M. (1996). From Robots to Rothko: The Bringing Forth of Worlds.
In Boden, M. (ed.) *The Philosophy of Artificial Life*. Oxford University Press.
- Winograd, T. and Flores, F. (1996). *Understanding Computers and Cognition: A New Foundation for Design*. Ablex Publishing Corporation, New Jersey.
- Wittgenstein, L. (1960) *The Blue and Brown Books* Basil Blackwell, Oxford.

Embodied cognitive science

Rens Kortmann

Universiteit Maastricht, IKAT
Neural networks and adaptive behaviour group
P.O. Box 616
NL-6200 MD Maastricht
The Netherlands
kortmann@cs.unimaas.nl

Abstract

The paper provides an introduction to the field of embodied cognitive science from a biological and behavioural perspective. We show how the field of neuro-ethology can help transform cognitive science from a representational to an embodied perspective. The transformation is necessary to introduce a bottom-up approach to understanding cognition in order to resolve some fundamental problems with classical cognitive science. We give examples of current research by which we characterise the key idea of embodied cognitive science: study cognition by starting with the low-level behaviour of simple animals.

1 Introduction

Embodied cognitive science studies how *complete agents* cope with the challenges of their environment (Clark, 1996; Pfeifer and Scheier, 1999). Complete agents are natural or artificial systems (animals or machines) that possess a body (Varela, Thompson, and Rosch, 1991) and are situated (Clancey, 1997) in their environment, i.e., they perceive the environment only through their own sensors. Moreover, complete agents possess characteristics such as autonomy, self-sufficiency, and adaptivity (Pfeifer, 1996). The ultimate aim of embodied cognitive science is to understand how high-level cognitive processes such as reasoning and language arose from low-level interactions with the environment such as object manipulation and corrective eye movements. The main motivation for this approach is the argument that during natural evolution, high-level cognitive capacities and the associated brain areas (neo-cortex in humans) developed from low-level sensori-motor couplings in the deep parts of the brain (e.g., the limbic system in humans) (Kalat, 2001).

Embodied cognitive science, therefore, contrasts traditional cognitive science (e.g., Stillings *et al.*, 1995) and classical artificial intelligence (e.g., Winston, 1977). The latter fields aim at understanding and synthesising high-level cognition using a computational theory of mind (Pylyshyn, 1984; Sterelny, 1990), i.e., interpreting human mental processes as computations on symbolic representations.

Embodied cognitive science builds upon the large scientific heritage of fields such as neuro-ethology (Guthrie, 1980; Camhi, 1984; Hoyle, 1984), cybernetics (Wiener, 1948), and ecological psychology (Gibson, 1979). These areas of research share two features in their scientific method: the bottom-up approach and comparative study.

In this paper, we identify the relation of embodied cognitive science and neuro-ethology. The following sections first treat the essence of traditional cognitive science and classical artificial intelligence (section 2) followed by the pith of neuro-ethology (section 3). We indicate that traditional cognitive science ignores particular aspects of intelligent behaviour, which leads to fundamental problems with the field. Furthermore, neuro-ethology focusses exactly on the aspects that are disregarded by traditional cognitive science (Beer, 1990). In section 4, we show how the research questions of neuro-ethology help transform traditional cognitive science into embodied cognitive science, which *does* address the issues disregarded by traditional cognitive scientists. In addition, we elucidate how embodied cognitive science relieves some of the problems faced by neuro-ethologists. In section 5 we discuss the scope and limitations of embodied cognitive science. Finally, we draw conclusions in section 6.

2 Traditional cognitive science and classical artificial intelligence

Cognitive science is traditionally a science of the intelligent mind (Stillings *et al.*, 1995). Some cognitive scientists restrict their focus to human intelligence, whereas others also include the abstract theory of intelligent processes and computer intelligence (Simon and Kaplan, 1989). In the 1960s, psychologists, philosophers, computer scientists, linguists, and neuroscientists joined forces in order to understand complex mental tasks such as thinking, remembering, and language (Gardner, 1985). For example, the interpretation of speech can be studied from a neuroscientific perspective (sound reception and neural processing in the brain), a linguistic perspective (parsing, word morphology, etc.), a computer science perspective (automatic speech recognition, natural language processing), a philosophical perspective (e.g., semantics), and a psychological perspective (intention of a speech act, etc.).

Pivotal in the approach of traditional cognitive science is the role of

symbolic representations. Pylyshyn (1984) states that the distinctive feature of *cognisers* – the subjects of study for cognitive scientists – is that their behaviour is based on representations. The idea is based on the representational theory of mind (Fodor, 1981; Sterelny, 1990) that states that all mental processes can be described in terms of computations on symbolic representations. In other words, mental activity is equivalent to the execution of an algorithm. For example, traditional cognitive science describes a person's reaction to a visual stimulus as follows. First, the mind transforms the incoming image into some symbolic input representation. The form of this representation could be a logical proposition, a feature vector, or some other description. Second, the input representation is compared to representations stored in memory. Third, from the outcomes of the comparison an output representation is created that contains the appropriate motor actions. Notice that in the traditional approach to cognitive science, the questions of how an input representation is built and of how an output representation is executed by the motor system, is often ignored. The main focus lies on the algorithmic processing in the middle.

In the 1930s, the conception of mental activity as computations on symbolic representations was supported when Alan Turing presented the universal Turing machine – a hypothetical machine that can compute any possible mathematical function (Haugeland, 1985). From Turing's idea, elaborated by von Neuman's architecture for general-purpose computers (Stillings *et al.*, 1995), the field of artificial intelligence (AI) was originated in 1956 (Gardner, 1985). Together with the traditional cognitive scientists, early researchers in artificial intelligence envisioned the human mind as a machine performing computations on symbolic representations to solve problems (Newell and Simon, 1981). Moreover, given the fact that the Turing machine could compute any mathematical function, the researchers felt they were able to build a computer that emulated human-level intelligence.

The achievements of classical AI lie mainly in the domain of mathematical problem solving and reasoning with explicit knowledge. For instance, in 1997 the computer chess programme Deep Blue beat the world champion Kasparov. Moreover, knowledge-based decision support systems are now wide-spread in companies and other organisations. The applications of classical AI research outperform humans in particular when a computer's processing speed and memory directly compete with that of a human brain. For instance, Deep Blue's success is mainly due to the fact that it could compute the consequences of its moves quicker than Kasparov could. Also, logical decision making is better performed by computers that can handle vast numbers of rules as compared to humans experts that rely on intuition much more than on logical inference – e.g., legal decision making by computers is more consistent than by humans (Van den Herik, 1997).

In other domains, though, the performance of artificial systems does not compare to that of natural systems (animals) by far. These domains

Embodied cognitive science, therefore, contrasts traditional cognitive science (e.g., Stillings *et al.*, 1995) and classical artificial intelligence (e.g., Winston, 1977). The latter fields aim at understanding and synthesising high-level cognition using a computational theory of mind (Pylyshyn, 1984; Sterelny, 1990), i.e., interpreting human mental processes as computations on symbolic representations.

Embodied cognitive science builds upon the large scientific heritage of fields such as neuro-ethology (Guthrie, 1980; Camhi, 1984; Hoyle, 1984), cybernetics (Wiener, 1948), and ecological psychology (Gibson, 1979). These areas of research share two features in their scientific method: the bottom-up approach and comparative study.

In this paper, we identify the relation of embodied cognitive science and neuro-ethology. The following sections first treat the essence of traditional cognitive science and classical artificial intelligence (section 2) followed by the pith of neuro-ethology (section 3). We indicate that traditional cognitive science ignores particular aspects of intelligent behaviour, which leads to fundamental problems with the field. Furthermore, neuro-ethology focusses exactly on the aspects that are disregarded by traditional cognitive science (Beer, 1990). In section 4, we show how the research questions of neuro-ethology help transform traditional cognitive science into embodied cognitive science, which *does* address the issues disregarded by traditional cognitive scientists. In addition, we elucidate how embodied cognitive science relieves some of the problems faced by neuro-ethologists. In section 5 we discuss the scope and limitations of embodied cognitive science. Finally, we draw conclusions in section 6.

2 Traditional cognitive science and classical artificial intelligence

Cognitive science is traditionally a science of the intelligent mind (Stillings *et al.*, 1995). Some cognitive scientists restrict their focus to human intelligence, whereas others also include the abstract theory of intelligent processes and computer intelligence (Simon and Kaplan, 1989). In the 1960s, psychologists, philosophers, computer scientists, linguists, and neuroscientists joined forces in order to understand complex mental tasks such as thinking, remembering, and language (Gardner, 1985). For example, the interpretation of speech can be studied from a neuroscientific perspective (sound reception and neural processing in the brain), a linguistic perspective (parsing, word morphology, etc.), a computer science perspective (automatic speech recognition, natural language processing), a philosophical perspective (e.g., semantics), and a psychological perspective (intention of a speech act, etc.).

Pivotal in the approach of traditional cognitive science is the role of

symbolic representations. Pylyshyn (1984) states that the distinctive feature of *cognisers* – the subjects of study for cognitive scientists – is that their behaviour is based on representations. The idea is based on the representational theory of mind (Fodor, 1981; Sterelny, 1990) that states that all mental processes can be described in terms of computations on symbolic representations. In other words, mental activity is equivalent to the execution of an algorithm. For example, traditional cognitive science describes a person's reaction to a visual stimulus as follows. First, the mind transforms the incoming image into some symbolic input representation. The form of this representation could be a logical proposition, a feature vector, or some other description. Second, the input representation is compared to representations stored in memory. Third, from the outcomes of the comparison an output representation is created that contains the appropriate motor actions. Notice that in the traditional approach to cognitive science, the questions of how an input representation is built and of how an output representation is executed by the motor system, is often ignored. The main focus lies on the algorithmic processing in the middle.

In the 1930s, the conception of mental activity as computations on symbolic representations was supported when Alan Turing presented the universal Turing machine – a hypothetical machine that can compute any possible mathematical function (Haugeland, 1985). From Turing's idea, elaborated by von Neuman's architecture for general-purpose computers (Stillings *et al.*, 1995), the field of artificial intelligence (AI) was originated in 1956 (Gardner, 1985). Together with the traditional cognitive scientists, early researchers in artificial intelligence envisioned the human mind as a machine performing computations on symbolic representations to solve problems (Newell and Simon, 1981). Moreover, given the fact that the Turing machine could compute any mathematical function, the researchers felt they were able to build a computer that emulated human-level intelligence.

The achievements of classical AI lie mainly in the domain of mathematical problem solving and reasoning with explicit knowledge. For instance, in 1997 the computer chess programme Deep Blue beat the world champion Kasparov. Moreover, knowledge-based decision support systems are now wide-spread in companies and other organisations. The applications of classical AI research outperform humans in particular when a computer's processing speed and memory directly compete with that of a human brain. For instance, Deep Blue's success is mainly due to the fact that it could compute the consequences of its moves quicker than Kasparov could. Also, logical decision making is better performed by computers that can handle vast numbers of rules as compared to humans experts that rely on intuition much more than on logical inference – e.g., legal decision making by computers is more consistent than by humans (Van den Herik, 1997).

In other domains, though, the performance of artificial systems does not compare to that of natural systems (animals) by far. These domains

typically are related to the real world and an artificial system's interaction with that real world (Brooks, 1990; Wilson, 1991; Clark, 1996; Clancey, 1997; Arkin, 1998; Pfeifer and Scheier, 1999). For instance, the mobile robot Shakey was built to navigate in a laboratory environment using a camera and a symbolic planning system (Nilsson, 1984). The robot's behaviour is brittle and sensitive to noise in the environment (Brooks, 1986; Arkin, 1998). The problems mentioned here are captured by two important theoretical issues regarding the symbolic representation of the real world: the frame problem (McCarthy and Hayes, 1969; Dennett, 1984) and the symbol grounding problem (Harnad, 1990). A thorough treatment of the problems reaches beyond the scope of this paper. We therefore restrict ourselves to giving a brief description in order to indicate the problems faced by traditional cognitive science and classical AI when applying the representational theory of mind to the real-world domain. For a full discussion of the problems we refer to the literature cited above.

The frame problem addresses the difficulties with representing change. If an intelligent system represents its environment symbolically, then how does it update the representation over time? This might not seem a problem for an abstract domain such as chess: there is a finite number of board positions and chess pieces. In contrast, changes in the real world are so fast and abundant that it is hardly plausible that intelligent systems (animals or machines) could efficiently use symbolic representations of the world for all of their behavioural patterns (Brooks, 1991; Kirsh, 1991). The symbol grounding problem discusses how symbolic representations relate to the real world. Again, in highly abstract domains this problem is less significant as such a system needs not know the meaning of the symbols it uses: a decision-support system need not know the meaning of the rules it operates on in order to deduce a valid conclusion from the input it is supplied with. However, a system operating autonomously in a real-world needs to know the relation between real-world objects (food, predators) and the system's symbolic representations of those objects: the meaning of symbols must be grounded in the system's own interaction with the real world (Pfeifer and Scheier, 1999). Gibson (1979) discusses the grounding of classification tasks in terms of affordances: food can be eaten, predators can be escaped from. Traditional cognitive science focusses on high-level cognitive capacities, not on low-level interactions with the real world. Therefore, for Shakey a goal position is a meaningless goal position, supplied by a human experimenter.

From the previous treatment of classical AI it becomes clear that explanations of intelligent behaviour in terms of the processing of symbolic representations might be appropriate in abstract domains, such as mathematics and logic – in more dynamic domains found in the natural world, the frame problem and symbol grounding problem point out to the need of increased understanding of an intelligent system's interaction with the real world. In other words: "understanding the nature of cognition requires

considering more than the complex problem solving and learning of human experts and their tutees." (Clancey, 1997, p. 6). In the next section we provide an introduction to a scientific field in which the interaction of animals with the environment is cardinal. It will be shown that this field in fact helps transform cognitive science from a representational to an embodied perspective.

3 Neuro-ethology

Neuro-ethology is the study of natural behaviour and the neural mechanisms mediating it (Guthrie, 1980; Camhi, 1984; Hoyle, 1984). The field originated from two other biological disciplines: neuro-biology and ethology that, by themselves, differ significantly in research interest and methods. Neuro-biology investigates the workings of neural structures in animals under controlled circumstances (Kandel, 1976). The experimental set-up allows neuro-biologist to obtain strict stimulus-response characteristics of a brain area. As a drawback, the laboratory conditions might cause the test animals to behave in a non-natural way which in turn places a bias on the experimental results (Huber, Franz, and Bülthoff, 1998). In contrast, ethologists study the behaviour of animals in their natural habitat (field experiments) (Slater, 1985). The approach usually guarantees the display of natural behaviour by the animals, but makes brain recordings virtually impossible. As a result, ethological research usually yields no results beyond the level of descriptions of natural behaviour (Camhi, 1984).

Neuro-ethology aims at combining the explanatory power of neuro-biology with the ecological validity of ethological research. In order to reach the aim, neuro-ethologists first observe the behaviour of an animal in its natural habitat and derive problems faced by the animal's neural system to produce the behaviour. These problems are subsequently studied in neuro-biological experiments under as natural circumstances as possible. Performing the experiments usually involves returning to the first phase (observing natural behaviour) when the neuro-biological results raise questions on the behavioural studies. Therefore, an iterative process (in which behavioural and neuro-biological experiments alternate) leads to the resolution of research questions that include the following topics: signal detection and recognition (e.g., calling songs or olfactory trails); co-ordination (e.g., in flight or when walking); localisation (sound sources, landmarks, etc.); and orientation (e.g., in navigation).

The approach is usually centred around a particular common behaviour faced by many animals, such as finding a mate, or avoiding predation. Subsequently, a target animal is chosen to study the behaviour and its neural underpinnings (Camhi, 1984). Although the particular details of neural mechanisms for similar behaviours can vary significantly amongst different

animal species, the nature of neural signalling is common in most animals and many species are constrained by the same physical limitations of their bodies and of their environment. For instance, different insect species have compound eyes and share the same habitat in which they display similar behaviours, such as feeding or target pursuit. Studying the neural mechanism mediating behaviour in one animal species, therefore is likely to reveal insight into the mechanisms of other species. In other words, neuro-ethology relies on comparative study.

Next to the term 'comparative study' also the phrase 'bottom-up approach' is central to the field of neuro-ethology. The phrase indicates that through studying relatively simple behaviours, such as orientation or localisation, neuro-ethologists eventually aim at explaining high-level behaviour such as decision making and the orchestration of highly complex behavioural patterns. We notice this approach contrasts the 'top-down' approach adopted by traditional cognitive scientists and researchers of classical AI. Whereas neuro-ethologists aim at understanding complex behaviours in terms of low-level interactions of an animal with its environment, the representations-focused cognitive science and AI researchers study high-level behaviour directly, while disregarding the underlying low-level behaviours that support the cognitive processes. As an example we mention the research on fear by LeDoux (1996) and co-workers. In the human brain he identified a cortical and sub-cortical pathway mediating fear-responses. When presented with a fearsome stimulus, a subject's cortical pathway elicits the awareness of fear and is therefore associated with high-level cognition. At the same time, the much faster sub-cortical pathway triggers a non-conscious fight-or-flight reaction which is associated with a low-level reactive response. The low-level behaviour can be overruled by the high-level, cognitive behaviour, but can never be replaced by it: the cognitive high road is too slow to trigger a timely response to a fearsome stimulus. In summary, a low-level, non-cognitive brain area is fundamental to a fear-response, even in humans. This observation supports a bottom-up approach to intelligent behaviour, as adopted in neuro-ethology.

Limitations of the neuro-ethological paradigm As was mentioned before, neuro-biological and thus neuro-ethological experiments need a controlled laboratory set-up in order to allow for meaningful measurements in the animal's brain. Using the existing techniques it is still impossible to measure neural activity in freely moving animals. Instead, to obtain valid stimulus-response characteristics, the animals that are investigated are usually placed in open-loop conditions, i.e., they are fixated and thus do not receive feedback from their actions. For instance, the activity of movement-sensitive neurons in flies is measured while the animal is fixated in a hollow tube with wax (Mastebroek, Zaagman, and Lenting, 1980). While neuro-

ethologists put a lot of effort in emulating as natural experimental conditions as possible, still there exists a fundamental problem: it is unclear whether the behaviour of animals and its neural activity is similar in open-loop and closed-loop conditions, i.e., when freely moving or not.

Moreover, neuro-ethological research can not study multiple behaviours in an animal simultaneously: again, in order to obtain valid stimulus-response characteristics, complex behaviours are decomposed into simpler behavioural patterns such as orientation or flight course stabilisation that are studied separately (Camhi, 1984). The synthesis of neural activity for the complex behaviour from the components is not a straightforward procedure, i.e., neural mechanisms cannot simply be added to account for the explanation of complex behaviour (Huber *et al.*, 1998). Another potential risk in neuro-ethology is assuming the existence of different neural mechanisms for different behavioural patterns, whereas one neural mechanism can mediate several behaviours. For instance, the phonotactic behaviour (sound-seeking) of crickets consists of a sound recognition and localisation component. Webb (1995) found that both behaviours are likely to be mediated by the same neural structure (see also section 4).

In summary, neuro-ethological research faces two methodological problems: lack of ecological validity due to closed-loop experiments and the decomposition of natural behaviour. We propose that embodied cognitive science can relieve the problems by employing robots to model aspects of animal behaviour. Robots can be used in closed-loop conditions and the modelling aspect allows for the synthesis of behaviours.

4 Embodied cognitive science

From the previous sections we conclude that neuro-ethology treats some of the shortcomings of traditional cognitive science. Section 2 showed the focus of traditional cognitive science and classical AI which was to describe and synthesise high-level cognitive processes through computations on symbolic representations. Their major shortcoming is the disregard of low-level behaviour that does not require symbolic representations in its explanation. Section 3 gave an introduction to the field of neuro-ethology that fixates on the neural mechanisms underlying low-level behaviour in animals. Drawbacks of the latter field include artificial experimental set-ups and the decomposition of behaviour.

We argue that embodied cognitive science can combine the best of both worlds. First, employing a neuro-ethological focus on low-level behaviour provides a better understanding of how cognisers (Pylyshyn, 1984) interact with the real-world and how they use their low-level behaviour as a foundation for high-level cognitive processes. Second, the modelling techniques from AI can be used to solve parts of the problems faced by neuro-ethology.

In the following, we shall elaborate on the arguments mentioned and by giving examples of current work in embodied cognitive science and the achievements.

Pfeifer and Scheier (1997) investigated a classification task in terms of sensori-motor co-ordination. Traditionally, classification is viewed as the mapping of an instance to a symbolic class representation (Estes, 1994). For instance, when seeing a particular animal, the viewer creates a symbolic representation of the animal and maps it to the class 'dogs'. Alternatively, Pfeifer and Scheier adopt a more 'Gibsonian' perspective on the task. They interpret classification of objects in terms of the possible actions an agent can perform with them: graspable objects vs non-graspable objects, pushable objects vs non-pushable objects, etc. In an experiment a mobile robot was placed in an arena with cylindrical objects from two classes: objects with either a large or a small diameter. When the robot met an object, it started to circle around the object and measured the angular velocity. Moreover, the robot possessed a 'tail' by which it could grasp the objects with small diameter, but not the large ones. Using Hebbian learning, a neural network was trained that decided for the robot to grasp an object. The decision was based on the angular velocity that was measured while circling the object. In this approach, sensori-motor patterns of activity (time-series of sensor and motor values), rather than symbols, constitute class representations. The approach fits better to the way children and animals learn to classify objects than the classical approach does (Smith, 1994). Moreover, the approach describes how an agent (animal or robot) grounds the meaning of an object class in its behaviour.

As a second example, we mention the work by Webb (1995; 1998) on cricket phonotaxis. The work proposes a neural mechanism of female crickets' classification and localisation of male crickets' calling songs. The mechanism was implemented on a robot model and yielded robust behaviour comparable to that of real crickets. Again, the classification task (distinguishing between different calling songs, belonging to males of different cricket species) was based on a behavioural response (approaching the conspecific male), not on a symbolic classification. The work gives another hint that symbolic representations are not always necessary to explain high-level cognitive tasks (categorisation). Moreover, Webb's approach to studying cricket phonotaxis adds an important aspect to cognitive science and neuro-ethology: robotic modelling. The approach relieves two problems of neuro-ethology: first, robots can be used in closed-loop conditions, i.e., it is possible to record the activity of the artificial neural mechanism that controls the robot, while the robot can move freely in a desirable environment; second, robots can be employed to model the orchestration of behaviours. Again, activity in the robot's artificial neural system can be recorded while the robot performs various behaviours in parallel. For a more complete treatment on using robots to model animals we refer to the work of Webb

(1995; 1998; 1999).

Next we shall mention a few examples of robot models of animal behaviour and the neural mechanisms underlying it. Some of the models were implemented in a physical robot, whereas others were built as computer simulations. Already in 1982 Arbib modelled visuo-motor co-ordination in frogs and toads in a system called *Rana computatrix*. Some ten years later Cliff (1991) modelled visually-guided behaviour of the hoverfly *Syricta pipiens* in a computer simulator called SYCO (*Syricta computatrix*). Other animals that were investigated include cockroaches (Beer and Chiel, 1993), lampreys (Ijspeert, Hallam, and Willshaw, 1999), salamanders (Ijspeert and Arbib, 2000), more flies (Franceschini, Pichon, and Blanes, 1992; Huber *et al.*, 1998), ants (Lambrinos *et al.*, 2000), and bats (Peremans *et al.*, 2000).

Applications Besides using robots in order to investigate animals, the field of embodied cognitive science has produced a wide range of applications by taking inspiration from biology. For instance, the field of behaviour-based robotics (Arkin, 1998) employs principles of animal behaviour to the design of autonomous systems (cars, wheelchairs, space rovers). The field of evolutionary robotics (Harvey *et al.*, 1997; Nolfi and Floreano, 2000) applies the theory of evolution to the development of robot controllers. Finally we mention the use of optic flow in visually-guided behaviour in robots (Weber, Venkatesh, and Srinivasan, 1997; Srinivasan *et al.*, 1997).

5 Discussion

In this paper we have discussed important aspects of traditional cognitive science, classical artificial intelligence and neuro-ethology. We pointed out to a few problems relating to the fields and showed that embodied cognitive science can relieve some of the problems through employing robot modelling. In fact, Clancey (1997) views the researchers of embodied cognitive science as a new generation of cyberneticists (Wiener, 1948) to stress the importance of the perception-action loop in embodied cognitive science research. Cybernetics studied the control of low-level behaviour in animals and machines already in the 1940s and 1950s. However, the artificial systems built by early cyberneticists (e.g., Walter, 1950) used simple controllers based on analogue hardware. In contrast, with present-day computing power, the use of neural network models in robots is facilitated.

Converting from a representational to an embodied perspective on intelligence requires some adaptive power. Modesty is called for when adopting a bottom-up approach to cognition, but people will be astounded by the ingenuity of the neural mechanisms underlying low-level behaviour in humans and other animals.

Limitations Using artificial systems (robots or computer simulations) to study real animals has some limitations of which we mention three. First, the role of noise is different in natural and artificial systems. For instance, photon absorption noise is Poisson distributed (Land, 1981), which does not necessarily have to be the case in artificial visual sensors. Second, the motor system of robots is usually extremely simplified which results in a difference between the sensory feedback experienced by robots and animals moving in the same environment. If, however, work focusses on the motor behaviour instead, the sensory processing is usually limited (Beer and Chiel, 1993; Ijspeert and Arbib, 2000). Third, an animal might use sensory input or environmental clues that are not modelled in the artificial system used. For instance, the role of the ultra-violet component of sunlight plays an important role in insect navigation (Lambrinos *et al.*, 2000). In a similar way, other sensory channels that we are currently unaware of might be of importance for the behaviour. Not implementing these channels in a robot model might lead to wrong explanations of the behaviour. This problem is even larger when computer simulations are used instead of physical robots. Computer simulations might represent a too simple environment and miss important sensory information. Instead, physical robots interact with the same environment as real animals and therefore are theoretically able to receive all sensory information available to real animals. The term ‘comparative study’ therefore also applies to embodied cognitive science.

All of the limitations mentioned above are general problems of the scientific method of modelling – they do not apply just to embodied cognitive science. Modelling involves abstracting away from the real system under investigation. All simplifications in the design of the model should be accounted for. For instance, using two wheels to model the motor system of an animal can be justified when yaw behaviour (rotation in the horizontal plane) is studied. Moreover, modelling is usually performed in an incremental fashion (adding more and more components to the model). In fact, the limitations could serve as the explanatory power of embodied cognitive science. For instance, when reproducing behaviour in a robot fails with the neural mechanism assumed to mediate the behaviour in real animals, this might indicate that more sensory channels are involved, or that the role of the actuators is more important than expected.

6 Conclusions

Altogether, we showed that neuro-ethology helps transform the perspective on cognitive science from one of representations to one of embodiment and situatedness. The method of applying robots to model animal behaviour and the neural mechanisms underlying the behaviour relieves some central methodological problems in traditional cognitive science and neuro-ethology.

First we noticed that traditional cognitive science and classical artificial intelligence underestimate the role of low-level real-world interaction in producing intelligent behaviour, whereas embodied cognitive science adopts a bottom-up approach starting with the low-level aspects of behaviour. Second we stressed that the problems in neuro-ethology regarding recording neural activity from living animals can be prevailed by using robot models. The use of robot models allows for comparative study of animals species. Summarising, embodied cognitive science combines the best of two worlds: traditional cognitive science and neuro-ethology.

Future work In future, we think embodied cognitive science could contribute to the scientific community by focusing on the relation between high-level cognitive tasks and the underlying low-level behaviour supporting it. For instance, the use of eye movements (low-level behaviour) is pivotal in understanding how high-level tasks such as face recognition are accomplished (Ballard, 1991). The grounding of symbolic representations is another important topic. Embodied cognitive science requires a change of mind (bottom-up instead of top-down approach) but offers important new insights to the study of intelligence.

Acknowledgements

I would like to thank Eric Postma for his advice and useful discussions.

References

- Arbib, M. (1982). Rana Computatrix: An Evolving Model of Visuo-motor Coordination in Frog and Toad. *Machine Intelligence*, Vol. 10, pp. 501-517.
- Arkin, R. (1998). *Behavior-based robotics*. MIT Press, Cambridge.
- Ballard, D. (1991). Animate vision. *Artificial intelligence*, Vol. 48, pp. 57-86.
- Beer, R. and Chiel, H. (1993). Simulations of cockroach locomotion and escape. *Biological neural networks in invertebrate neuroethology and robotics* (eds. R. Beer, R. Ritzmann, and T. McKenna), pp. 267-285, Academic Press, Boston.
- Beer, R. (1990). *Intelligence as adaptive behavior, an experiment in computational neuroethology*. Academic Press, Boston.
- Brooks, R. (1986). A robust layered control system for a mobile robot. *IEEE journal of robotics and automation*, Vol. 2, pp. 14-23.

- Brooks, R. (1990). Elephants don't play chess. *Robotics and autonomous systems*, Vol. 6, pp. 3-15.
- Brooks, R. (1991). Intelligence without representation. *Artificial intelligence*, Vol. 47, pp. 139-159.
- Camhi, J. (1984). *Neuroethology*. Sinauer associates, Sunderland.
- Clancey, W. (1997). *Situated cognition: on human knowledge and computer representations*. Cambridge University Press, New York.
- Clark, A. (1996). *Being there: putting brain, body, and world together again*. MIT Press, Cambridge.
- Cliff, D. (1991). The Computational Hoverfly: A Study in Computational Neuroethology. *From Animals to Animats: Proceedings of the First International Conference on the Simulation of Adaptive Behaviour* (eds. J.-A. Meyer and S. Wilson), MIT Press, Cambridge.
- Dennett, D. (1984). Cognitive Wheels: The frame problem of AI. *Minds, Machines and Evolution* (ed. C. Hookway), pp. 129-152, Cambridge University Press, Cambridge.
- Estes, W. (1994). *Classification and cognition*. Oxford University Press, New York.
- Fodor, J. (1981). *Representations: philosophical essays on the foundations of cognitive science*. Harvester Press, Brighton.
- Franceschini, N., Pichon, J.-M., and Blanes, C. (1992). From insect vision to robot vision. *Philos. Trans. R. Soc. Lond. Biol.*, Vol. 337, No. 1281, pp. 283-294.
- Gardner, H. (1985). *The mind's new science*. Basic Books, New York.
- Gibson, J. (1979). *The ecological approach to perception*. Houghton Mifflin, Boston.
- Guthrie, D. (1980). *Neuroethology: an introduction*. Wiley, New York.
- Harnad, S. (1990). The symbol grounding problem. *Physica D*, Vol. 42, pp. 335-346.
- Harvey, I., Husbands, P., Cliff, D., Thompson, A., and Jakobi, N. (1997). Evolutionary robotics: the Sussex approach. *Robotics and autonomous systems*, Vol. 20, pp. 205-224.
- Haugeland, J. (1985). *Artificial intelligence : the very idea*. MIT Press, Cambridge.

- Hoyle, G. (1984). The scope of neuroethology. *The Behavioural and Brain Sciences*, Vol. 7, pp. 367–412.
- Huber, S., Franz, M., and Bülthoff, H. (1998). On robots and flies: modeling the visual orientation behavior of flies. Technical Report 56, Max-Planck-Institut für biologische Kybernetik, Tübingen, Germany.
- Ijspeert, A. and Arbib, M. (2000). Visual tracking in simulated salamander locomotion. *From Animals to Animats, Proceedings of the 6th International Conference on the Simulation of Adaptive Behavior* (eds. J. Meyer, A. Berthoz, D. Floreano, H. Roitblat, and S. Wilson), pp. 88–97, MIT Press, Cambridge.
- Ijspeert, A., Hallam, J., and Willshaw, D. (1999). Evolving swimming controllers for a simulated lamprey with inspiration from neurobiology. *Adaptive Behavior*, Vol. 7, No. 2, pp. 151–172.
- Kalat, J. (2001). *Biological psychology*. Brooks/Cole publishing company, Pacific Grove, 7th edition.
- Kandel, E. (1976). *Cellular basis of behavior*. W.H. Freeman and Co., San Francisco.
- Kirsh, D. (1991). Today the earwig, tomorrow man? *Artificial intelligence*, Vol. 47, pp. 161–184.
- Lambrinos, D., Möller, R., Labhart, T., Pfeifer, R., and Wehner, R. (2000). A mobile robot employing insect strategies for navigation. *Robotics and Autonomous Systems*, Vol. 30, pp. 39–64.
- Land, M. (1981). Optics and vision in invertebrates. *Handbook of sensory physiology* (ed. H. Autrum), Vol. VII/6B, pp. 471–592, Springer, Berlin.
- LeDoux, J. (1996). *The emotional brain : the mysterious underpinnings of emotional life*. Simon and Schuster, New York.
- Mastebroek, H., Zaagman, W., and Lenting, B. (1980). Movement detection: performance of a wide-field element in the visual system of the blowfly. *Vision research*, Vol. 20, pp. 467–474.
- McCarthy, J. and Hayes, P. (1969). Some philosophical problems from the standpoint of artificial intelligence. *Machine intelligence*, Vol. 4, pp. 463–502.
- Newell, A. and Simon, H. (1981). Computer science as empirical inquiry: symbols and search. *Mind design – philosophy, psychology, artificial intelligence* (ed. J. Haugeland), pp. 35–66, MIT Press, Cambridge.
- Nilsson, N. (1984). Shakey the robot. Technical Report 323, AI Center, SRI International, Menlo Park.

- Nolfi, S. and Floreano, D. (2000). *Evolutionary robotics – the biology, intelligence, and technology of self-organizing machines*. MIT Press, Cambridge.
- Peremans, H., Müller, R., Carmenta, J., and Hallam, J. (2000). A biomimetic platform to study perception in bats. *Proceedings of the SPIE international symposia on sensor fusion and decentralized control in robotic systems III*. In press.
- Pfeifer, R. and Scheier, C. (1997). Sensory-motor coordination: the metaphor and beyond. *Robotics and autonomous systems*, Vol. 20, pp. 157–178.
- Pfeifer, R. and Scheier, C. (1999). *Understanding intelligence*. MIT-press, Cambridge.
- Pfeifer, R. (1996). Building “Fungus Eaters”: Design principles of Autonomous Agents”. *From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behaviour*, MIT Press, Cambridge.
- Pylyshyn, Z. (1984). *Computation and cognition: toward a foundation for cognitive science*. MIT Press, Cambridge.
- Simon, H. and Kaplan, C. (1989). Foundations of cognitive science. *Foundations of cognitive science* (ed. M. Posner), pp. 1–47, MIT Press, Cambridge.
- Slater, P. (1985). *An introduction to ethology*. Cambridge University Press, Cambridge.
- Smith, E. T. L. (1994). *A dynamic systems approach to the development of cognition and action*. MIT Press, Cambridge.
- Srinivasan, M., Chahl, J., Nagle, M., and Zhang, S. (1997). Embodying natural vision into machines. *From living eyes to seeing machines* (eds. M. Srinivasan and S. Venkatesh), pp. 249–265, Oxford University Press, Oxford.
- Sterelny, K. (1990). *The representational theory of mind: an introduction*. Blackwell, Oxford.
- Stillings, N., Weisler, S., Chase, C., Feinstein, M., Garfield, J., and Rissland, E. (1995). *Cognitive science – an introduction*. MIT Press, Cambridge, 2nd edition.
- Van den Herik, J. (1997). From chess moves to legal decisions: a position statement. *Proceedings of JURIX '97: the tenth conference on legal knowledge based systems* (ed. A. Oskamp), pp. 107–109, Gerard Noodt Instituut, Nijmegen.

- Varela, F., Thompson, E., and Rosch, E. (1991). *The embodied mind*. MIT Press, Cambridge.
- Walter, W. (1950). An imitation of life. *Scientific American*, Vol. 182, No. 5, pp. 42-45.
- Webb, B. (1995). Using robots to model animals: a cricket test. *Robotics and autonomous systems*, Vol. 16, pp. 117-134.
- Webb, B. (1998). Robots, crickets and ants: models of neural control of chemotaxis and phonotaxis. *Neural Networks*, Vol. 11, pp. 1479-1496.
- Webb, B. (1999). A framework for models of biological behaviour. *International journal of neural systems*, Vol. 9, No. 5, pp. 375-381.
- Weber, K., Venkatesh, S., and Srinivasan, M. (1997). Insect inspired behaviours for the autonomous control of mobile robots. *From living eyes to seeing machines* (eds. M. Srinivasan and S. Venkatesh), pp. 227-249, Oxford University Press, Oxford.
- Wiener, N. (1948). *Cybernetics, or control and communications the animal and the machine*. Wiley, New York.
- Wilson, S. (1991). The animat path to AI. *From Animals to Animats: Proceedings of the First International Conference on Simulation of Adaptive Behavior* (eds. J.-A. Meyer and S. Wilson), pp. 15-21, MIT Press, Cambridge.
- Winston, P. (1977). *Artificial intelligence*. Addison-Wesley, Reading.

EVOLUTIONARY ROBOTS: THE NEXT GENERATION

Dario Floreano and Joseba Urzelai
Laboratory of Microprocessors and Interfaces (LAMI)
Swiss Federal Institute of Technology (EPFL)
CH-1015 Lausanne, Switzerland

February 28, 2000

Abstract

After reviewing current approaches in Evolutionary Robotics, we point to directions of research that are likely to bring interesting results in the future. We then address two crucial aspects for future developments of Evolutionary Robotics: choice of fitness functions and scalability to real-world situations. In the first case we suggest a framework to describe fitness functions, choose them according to the situation constraints, and compare available experiments in the literature on evolutionary robotics. In the second case, we suggest a way to make experimental results applicable to real-world situations by evolving online continuous adaptive controllers. We also give an overview of recent experimental results showing that the suggested approaches produce qualitatively superior abilities, scale up to more complex architectures, smoothly transfer from simulations to real robots and across different robotic platforms, and autonomously adapt in few seconds to several sources of strong variability that were not included during the evolutionary run.

1 Introduction

Evolutionary Robotics is the application of artificial evolution to robotic systems with a sensory-motor interface to the world. Although in the early days artificial evolution was mainly seen as a strategy to develop more complex and performant robot controllers, nowadays the field has become much more sophisticated and diversified. We can identify at least three approaches to Evolutionary Robotics: Automated Engineering, Artificial Life, and Synthetic Biology/Psychology. These three approaches largely overlap with each other, but still have quite different goals that eventually show up in the results obtained.

Automated Engineering is about the application of artificial evolution for automatically developing algorithms and machines displaying complex abilities that are hard to program with conventional techniques. Within this context the desired architectures are well defined and the problem is usually cast in terms of parameter optimization by evolutionary techniques. Artificial evolution can come up with strikingly efficient and surprising solutions that exploit *invariants* and features invisible to an external observer¹. For example, in one of our early experiments [30] we evolved

¹Invariants are constant relationships. Our visual system exploits many invariants. For example, the fact that we perceive objects always the same size irrespective of distance is given by neural detectors that

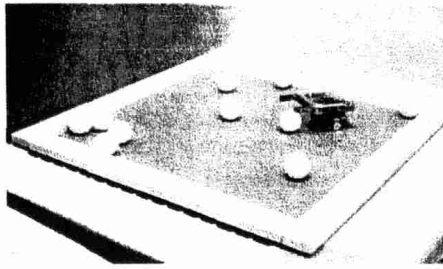


Figure 1: A Khepera robot evolved to find and pick up balls [30].

the control system of a Khepera robot with a gripper for the ability to find the highest number of ping-pong balls scattered around an arena and pick them up (figure 1). This was a difficult task because the resolution of the infrared sensors in front of the robot was not sufficiently good to discriminate between balls and walls. The best evolved robots succeeded using an unexpected strategy. They moved backwards until they detected something and then started to rotate on the spot until they faced the object, picking it up if it was a ball or resuming backward motion otherwise. This was possible because the rotation generated a scan of the object across several neighboring sensors whose combined activations were sufficient to discriminate between objects and assume a correct gripping position if necessary. At the same time, the remarkably simple avoidance strategy, coupled with the geometry of the arena, ensured a good exploration of the full environment.

Artificial Life instead is about evolution of artificial creatures that display life-like properties. In this context the notion of evolutionary goal is not appropriate (living creatures do not evolve towards a prespecified goal) and there are very few constraints that limit the directions that evolution might take. These evolutionary systems are usually self-sufficient, self-contained, and autonomous. The selection criterion is often the energy level of the creature and the environment has an ecological validity in that it includes food sources, mates, predators, a nest, etc. These artificial worlds are easier to implement in computer simulations because simulations give more freedom to experiment life-as-it-could-be. In this context, even evolutionary experiments that end up in complete extinction of one species, or display alternating dynamics such as in competitive co-evolutionary scenarios [35, 5, 12] (figure 2), may be considered important because they reveal interesting patterns of life. The artificial life approach is more interested in the emergent phenomena than in the optimization of a pre-defined strategy.

Synthetic Biology/Psychology attempt to understand the functioning of biological and psychological mechanisms by evolving those mechanisms in a robot put in conditions similar to those of the animals under study. This approach finds its roots in the inspiring booklet *Vehicles* [4] by the neurophysiologist Valentino Braitenberg who showed that apparently complex behaviors and emotions can be reproduced in the eye of an external observer by simple sensory-motor machines.² The evolutionary approach expands this method to more complex mechanisms and environmental conditions. For example, it is commonly assumed that rats use geometric modules

exploit the constant relationship between the size of the retinal projection and the perceived distance of the object. Evolved robots often discover invariants and exploit them to accomplish a task. Notice that since invariants exist mainly from the perspective of the subject perceiving them [18], it is virtually impossible for an external observer to define and incorporate them in a pre-fabricated control software.

²Braitenberg was also the first person to suggest artificial evolution of robots in that same booklet.

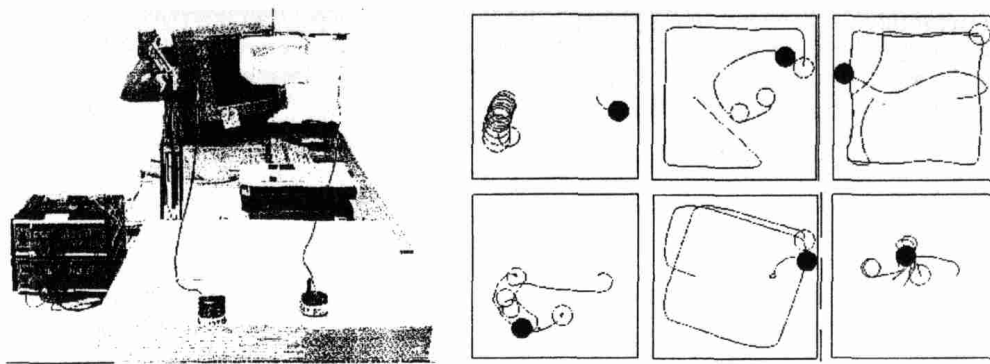


Figure 2: Co-evolutionary predator and prey robots display rapid alternation of diversified strategies caused by continuously changing dynamics. Left: co-evolutionary arena where a predator with vision is co-evolved with a short-sighted but faster prey. Right: Examples of emerging chasing-escape strategies every 10 generations (predator is black disk, prey is empty disk) [14].

and complex cognitive manipulations to find their way to a learned location [17], but behavioral and neurophysiological data are open to different interpretations and are still subject of heated scientific discussion. Orazio Miglino and Henrik Lund [26] investigated this issue by evolving Khepera robots in the same environmental conditions used for rats by ethologists and neurophysiologists. Their evolved robots reported the same performances measured on living animals, but did not use complex cognitive abilities to do so. Instead, these robots displayed simple sensory-motor sequences that capitalized upon geometric invariants of the environment. For example, the layout of long and short walls, combined with certain rotations by the robot, returned a high-probability to end up in the target location, no matter where the robot was initially located. Of course it would be wrong to deduce that rats use similar sensory-motor mechanisms, but evolutionary robotic data show that there is at least one *alternative explanation* for the available biological evidence. Therefore, the power of this approach is that it can be used to debunk strong assumptions based on weak evidence and at the same time suggest additional experiments.

These three approaches generate awareness and draw attention to different aspects of autonomous systems. The overall emerging picture is one where what matters most and differentiates Evolutionary Robotics from other machine learning approaches is that here robots self-organize while freely interacting with their own environment.

2 Research Areas

Evolutionary Robotics build upon several aspects of artificial evolution, as shown in figure 3. We believe that some of them are more promising than other. One way to read figure 3 is to visually organize it in three rows.

The bottom row includes aspects that typically pertain to machine learning and computer science. They include the best way to describe the searching properties of an evolutionary algorithm, how to design and optimally combine evolutionary operators, and how to implement the algorithms in software and hardware so that computational and physical resources are optimally exploited. These aspects are very important for conventional function optimization, but we argue that they are rather secondary for

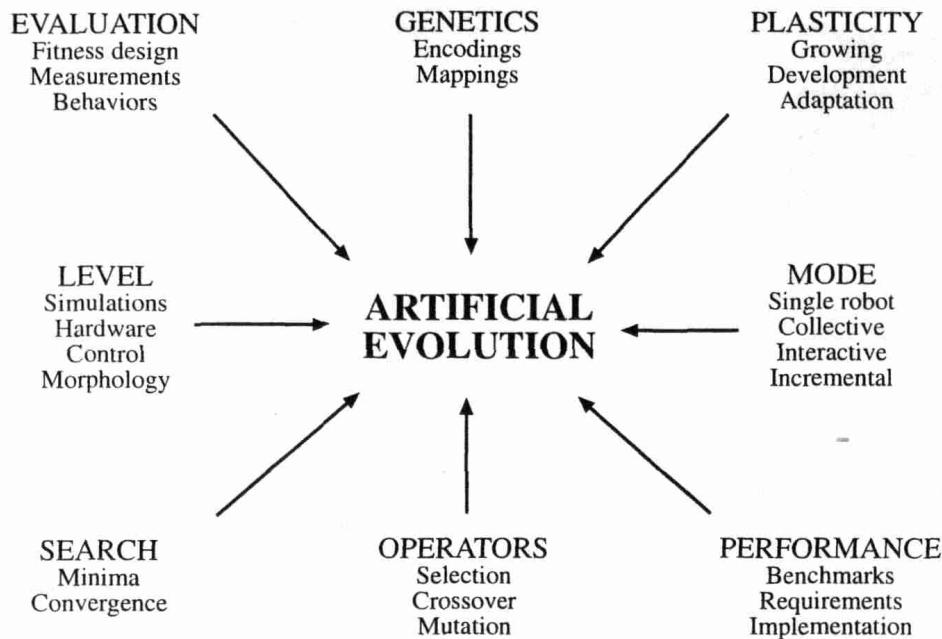


Figure 3: Different aspects of research in Artificial Evolution. The bottom row concerns mainly issues related to conventional evolutionary computation. The middle row displays hot issues that are currently being investigated mainly in the area of Evolutionary Robotics. The top row describes areas that have been only scarcely tackled and represent high potentials for significantly new achievements.

what concerns evolution of autonomous intelligent robots.

Machine Learning	Robot Learning
Learning <i>in vacuum</i>	Embedded Learning
Statistically well-behaved data	Data distribution not homogeneous
Mostly off-line	Mostly on-line
Informative feedback	Qualitative and sparse feedback
Computational time not an issue	Time is crucial
Hardware does not matter	Hardware is a priority
Convergence proof	Empirical proof

Table 1: Some crucial differences between Machine Learning and Learning in Autonomous Robots.

To start with, machine learning and robot learning are quite different, as shown in table 2. From an algorithmic perspective, an evolving robot spends by large most of its time interacting with the environment. Genetic operators and operations that map sensors into motor commands may take less than 5% of the total time. The remaining 95% is taken by mechanical actions, such as move a leg, rotate the camera, update the visual field, transmit signals across various parts of the hardware (and/or to an external workstation), etc.

An experiment could miserably fail if one does not pay sufficient attention to the interaction aspects even if the searching properties and computational efficiency of the evolutionary engine were optimally set. For example, if the robot is started always from the same position, the fitness is quickly maximized but the results may not generalize to different starting positions. On the other hand, by only taking a quick look at the experimental literature it turns out that almost any type of selection strategy (ranking-based, proportional, tournament-based, etc.), crossover and mutation operators, and buggy code can generate interesting results. Some mathematicians exploit this to argue that artificial evolution is nothing more than simulated annealing, implying that there is nothing special about genetics and that it all amounts to random search and chance. I argue that they are wrong because in doing so they automatically dismiss the aspects of artificial evolution shown on the top row of figure 3. But even if they were right, they still miss the point of Evolutionary Robotics. As I said above, what matters most in this approach is to achieve self-organization of an autonomous, situated, and interactive machine. The algorithmic details of how this is achieved come in second place.

The row in the middle of figure 3 shows some of the current and most important aspects of Evolutionary Robotics. The first aspect concerns the Level at which artificial evolution operates. It can be applied to simulated organisms or to physical robots, or to a combinations of both. There is an important ongoing discussion about these issues and several strategies have been suggested to allow transfers across levels (e.g., see [22, 29]) that deserve further efforts. Similarly, one can decide to evolve the control system or some characteristics of the robot body (morphology, sensors, etc.), or co-evolve them both [25]. Finally, one may decide to physically evolve the hardware, such as the electronic circuits [36] and the body shape [33]. All these issues, among others, are likely to define a new engineering methodology.

Another aspect of Evolutionary Robotics concerns the evolutionary Mode. Should one use a single robot and serially test each individual one at a time [8], or is it better to use a population of such robots sharing the same environment [41]? What are the emerging dynamics and how do they affect the results? Within a collective system, one may set up a competitive scenario or a cooperative one. It may even happen that competition and cooperation autonomously develop as an emergent phenomenon. Interactive mode instead is the situation where a robot evolves interactively with a human who manually selects the best individuals. There are only sporadic studies of interactive evolution, but this is going to be a crucial issue for applications related to human assistance and entertainment.

Incremental mode is when one attempts to carry on evolution from previously evolved populations, usually introducing some type of modification to make the system more complex. Incremental evolution is important to tackle complex problems that cannot be evolved from scratch (the bootstrap problem), but only few studies have been dedicated to this topic so far [6, 21, 10].

The top row of figure 3 displays areas of research that have been only scarcely addressed in the literature, but are likely to make significant advancement in Evolutionary Robotics. The Evaluation aspect is concerned with the development of methodologies to set up an evolutionary system, to measure, and assess its development, and to objectively compare it to other evolutionary systems. The current situation is that every one has his own fitness recipes, most describe results in terms of average and best fitness per generation, and nobody compares results with those of other people. Although some authors have attempted to devise new ways of measuring evolutionary dynamics [2, 5], more work in this direction is needed. To this end, in the next section of this paper I will suggest a method to conceive, assess, and compare fitness functions.

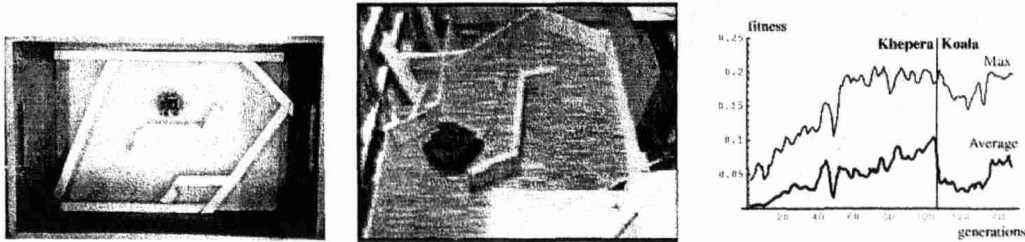


Figure 4: Incremental evolution across different robots. An initial population is evolved for navigation abilities on the miniature mobile robot Khepera (left). After 100 generations the population is transferred on the larger Koala robot and incrementally evolved using the same fitness function (center). The population quickly re-adapts to the new morphology and sensor characteristics, as shown by the fitness data (right) [10].

A similar story holds for the behavioral evaluation of evolved robots. Since these robots are situated systems, one cannot understand their functioning by simply looking at the evolved architectures and parameters. Earlier on we mentioned that evolved robots exploit invariant features, that is subjective constant relationships between the agent and the environment. In order to understand these invariants it will be necessary for roboticists to adopt the same techniques used by psychologists and ethologists to study invariants exploited by animals and humans. There are more than 150 years of well-established psychophysical techniques ready to be adapted to the new generation of intelligent robots.

The Genetics aspect is about what goes into the artificial chromosomes and how these chromosomes are mapped into individuals. Genetic encoding and genotype-phenotype mappings are the key to the evolvability of a system. There are some interesting studies in this area that show the potentials of better understanding this aspects. For example, Harvey's work on Neutral Networks in artificial evolution [20] indicates that some type of genetic mappings are more likely to lead to useful neutral changes (i.e., changes that do not immediately affect the fitness of the individual) that may eventually provide a species with radically new and more adaptive abilities. Several authors have explored different types of encoding and mapping schemes, but none of these strategies has shown a distinctive advantage with respect to vanilla encoding styles.³

Finally, the Plasticity aspect refers to all those processes that contribute to adaptively shape a fully-fledged organism. Structural growth, maturation, and ontogenetic adaptation (often called learning) are largely unexplored factors that complement, improve, and modify the adaptive properties of artificial evolution. For example, it has been shown that combining evolution with generative grammars (rules that recursively unfold into other rules) can effectively produce complex patterns that sometimes resemble life-like structures [19, 23, 35]. However, it is not clear how and what growing rules should be encoded. Nolfi and colleagues have begun to tackle the issue of maturation whereby the control system of a robot adaptively develops in time while the robot interacts with the environment, showing that this results in more efficient behaviors and specialized control architectures [32]. This seems to be the only work available in this area despite the fact that according to biologists and psychologists, maturation plays a major role in the definition of the final organism. Some people have

³I mean that either the same abilities can be evolved with a more straightforward encoding technique or that the evidence presented is by far not conclusive.

addressed the combination of evolution and other types of adaptive mechanisms. In one of the next sections I will propose a different approach that emphasizes evolution of adaptation rather than evolution and adaptation.

Even small advancements in any of the areas displayed on the top row of figure 3 will greatly contribute towards the creation of autonomously evolving machines that display life-like properties. I suspect that in the long run artificial evolution will play a minor-but essential-role in the overall methodology, that providing a medium for the development of powerful adaptive mechanisms such as growth, development, and ontogenetic adaptation.

3 Fitness Space

The fitness function defines which individuals are selected for reproduction. It is therefore a major factor of artificial evolution. If there is no fitness function and individuals are randomly selected, the effects of reproduction amount to genetic drift whereby all individuals become similar to each other with small random variations.

In Evolutionary Robotics, the choice of fitness function has strong consequences for implementation on physical robots, evolvability of the robot, dynamics of the evolutionary process, and eventually for outcomes. Most people struggle with the choice of suitable functions using a trial-and-error strategy. The most affected by this choice are people interested in using artificial evolution for Automated Engineering because they often have well-defined expectations about the final result. Unfortunately, there is not a way to infer a fitness function from the definition of the expected behavior. Typically, one comes up with a function based on one's own experience, tries it out, and then gradually modifies it to accommodate additional constraints. Although conceiving a fitness function suitable for a desired ability is still much easier than designing the corresponding program, the widespread use of Evolutionary Robotics requires better awareness of the decisions involved in setting up a fitness function.

In this section I propose *Fitness Space* as a framework to devise, assess, and compare fitness functions. Fitness Space can be used as a guideline to come up with fitness functions according to one's goals, but it does not provide a recipe to actually define specific functions. Fitness Space is defined by three dimensions.⁴

3.1 Functional-Behavioral Dimension

The first dimension is given by the continuum between Functional and Behavioral fitness. A purely functional fitness is based only on components that directly measure the way in which the system functions. For example, in an early attempt to evolve a neural controller for a walking robot, Lewis and Fagg used a functional fitness that measured the frequency and amplitude of the oscillations of the evolutionary controller [24]. The closer these two components were to the desired pattern, the higher the fitness of the individual.⁵ This implies that the authors knew what type of oscillatory dynamics were required for producing a certain behavior. On the other end, a purely behavioral fitness is based only on components that measure the behavior of the individual's behavior. To stay with the example of the walking robot, a behavioral function would be proportional to the distance covered by the robot in a given amount of time. Another way of describing the difference between these two fitness extremes

⁴Fitness Space should not be confused with Fitness Landscape which instead describes the distribution of fitness values corresponding to all possible combinations of genetic states.

⁵Functional fitness was used mainly in an early stage of evolution. In later stages, the authors added further behavioral components to the fitness.

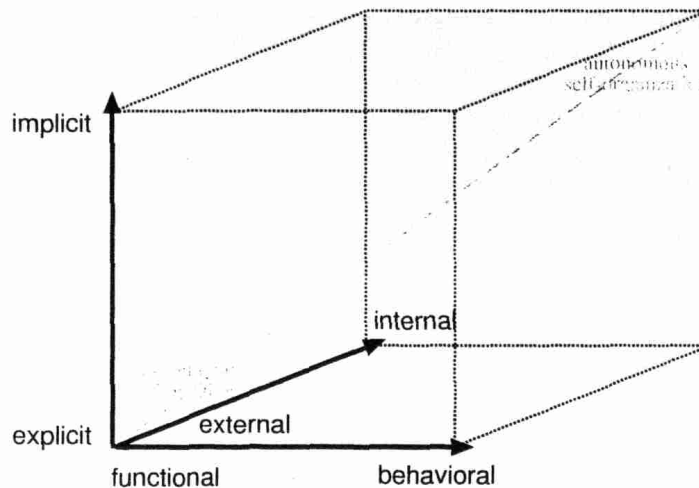


Figure 5: Fitness Space is a framework for defining and comparing fitness functions along three dimensions. It provides a nominal and ordinal scaling of functions. The diagonal between the lower-left corner and the upper-right corner defines a continuum of functions between conventional optimization approaches and generation of autonomous self-organizing systems.

is that functional fitness measures the *causes* of behavior whereas behavioral fitness measures the *effects* of behavior. Either choice has strong implications. A functional fitness can ensure the highest match between evolved and desired behavior, but it is quickly compromised as soon as mechanical or environmental factors do not match any longer the functional aspects of the controller. For example, a wheel may gradually wear out and induce a rotation bias of the robot. On the other hand, behavioral fitness can produce viable results even in the presence of some mechanical and sensory defects because the control system will implicitly accommodate them, but the results may not match the expectations. For example, the evolved robot may crawl instead of walk. The position of a given fitness function along the Functional-Behavioral dimension depends on the number of these two types of components and their relative weights. Later on we shall see in more detail how different functions can be positioned and compared in Fitness Space.

3.2 External-Internal Dimension

The dimension along the External-Internal continuum refers to availability of the fitness measure with respect to the robot. An external fitness component is one that cannot be measured directly by the robot. For example, the exact distance between a robot and an obstacle can be measured only by external positioning devices or by an external observer.⁶ An internal component instead is one that can be measured by the robot itself, such as the energy level or the state of its own sensors. The difference is subtle, but very important in Evolutionary Robotics. For example, external fitness functions are often used in software simulations of robots where all aspects of the system are directly available to the programmer. Here the distinction between internal

⁶Notice that odometry and proximity sensors (such as sonar and active infrared) cannot be reliably used to estimate distances.

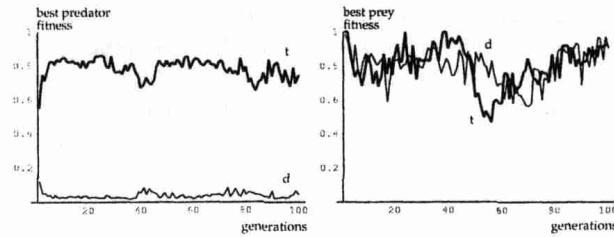


Figure 6: Comparisons between fitness of the best individuals measured as *time to contact* (t) and as *distance*. (d). Both species have been evolved using fitness t. **Left:** Best evolved predators do not attempt to minimize distance. **Right:** Best evolved prey attempt to maximize distance.

and external variables is only formal because both types of variables are readily available to the programmer. External fitness functions are popular because they allow a detailed and precise assessment of the robot performance and also because they naturally conform to the perspective of an external observer (in other words, it is easier to design a function that fits one's perspective of the expected behavior). However, real robots cannot always be evolved using external fitness components and, when feasible, this implies resorting to *ad hoc* and expensive devices, such as a Global Positioning System or an external camera with tracking software. Therefore, the choice of an external function should be carefully evaluated when devising an evolutionary robotic system. For example, external fitness functions are not recommended if one decides to start evolving the robot in simulation and later wishes to incrementally carry on evolution on a physical robot. Similarly, one should be careful about making strong claims on results obtained using external fitness functions because these results might not be easily generalized to many real-world situations where the necessary measuring devices are not available.

In the context of autonomous robots, we think that external fitness functions do not give a higher probability of success than internal fitness functions because they are based on the perspective of an external observer. In other words, the robot may be forced to display a behavior that is too difficult for the characteristics of their sensory-motor apparatus or of their control architecture. Consider for example the case of competitive co-evolution between predator and prey physical robots that we explored in previous work [7]. In that case we used internal fitness functions based on time-to-contact between the two robots measured through the clock of the on-board microcontrollers. The fitness of the prey was *proportional* to the amount of time spent without being touched by the predator, whereas the fitness of the predator was *inversely proportional* to the amount of time spent before catching the prey. In few generations the best individuals of each species were capable of maximizing their own time-based fitness displaying a high variety of behaviors to achieve that. Since in earlier work on simulated predator-prey Cliff and Miller employed an external fitness function based on distance between the robots (proportional for the prey and inversely proportional for the predator) [5], we repeated my experiments in simulation to check whether the robots evolved with time-based fitness reported the same values when measured with distance-based fitness [13]. It turned out that while almost all best prey reported the same fitness values using the two measures, all predator robots reported very low distance-based fitness (figure 6)! The reason was that instead of pursuing the prey, they used other strategies such as waiting for the prey by a wall like a spider or attacking only when the prey was moving in a certain direction and

relative position. This meant that most of the time they were far from the prey (low distance-based fitness value), but were still very successful at catching it (high time-based fitness). This result can be understood if one considers that the prey can go faster than the predator and therefore most strategies that attempt to get closer to the prey all the time will fail to catch it. Instead, a fitness based on time to contact does not necessarily select individuals for their ability to follow prey, giving more freedom to the evolving system. This example does not imply that internal fitness functions always mean less constraints for an evolving system, but they do encourage the engineer to reconsider assumptions deriving from an external perspective.

3.3 Explicit-Implicit Dimension

The Explicit-Implicit dimension refers to the quantity of constraints explicitly imposed by a human person to select individuals for reproduction. An approximate indicator is given by the number of components included in the fitness function. The higher the number of components, the more explicit the fitness function is.

Artificial Life approaches aimed at studying evolution of ecosystems tend to use implicit fitness functions in order to ensure ecological validity because in real life there is not an explicit fitness function. For example, artificial organisms may reproduce only if and when their energy levels reach a certain threshold (these type of ecosystems are also known as *Latent Energy Environments* [28]). Compare this with a situation where the fitness function explicitly rewards—for example—the quantity of food items gathered, distance from predators, and the ability to recognize conspecifics.

Automated Engineering approaches instead tend to resort to more explicit fitness functions in the attempt to actively steer the evolutionary system towards desired behaviors. This may sound reasonable, but in practice it gets out of control very quickly. As the number of constraints increases one is faced with the problem of how to weight and combine them (addition, product, e.g.). Furthermore, a higher number of components can increase the probability of local minima and make the problem too hard for an initial random population (bootstrap problem). Once again, these problems can be partly explained by the fact that fitness constraints are chosen from an external perspective and thus may be hard to meet by the robotic hardware and control architecture.

We think that explicit fitness functions are in contrast with the search of emergent forms of artificial intelligence because although the resulting evolved systems have not been pre-programmed their abilities have largely been decided and constrained by an external observer. Such evolved systems can hardly display unexpected abilities and under some definitions they may not be called emergent (for example, when emergence is defined as the degree of surprise [34]).

3.4 Comparing Evolutionary Experiments

Fitness Space allows us to compare the growing number of experiments available in the literature and make their underlying approach more explicit. In post-Galileian science, there are four methods to compare, or scale, experimental observations. *Categorical or nominal scales*, the most primitive methods, group observations into qualitative classes. For example, one may classify approaches in robotics as “bio-inspired”, “adaptive”, “cartesian”, etc. *Ordinal scales* are possible only when one can assign to each observation a number that reflects some quantity property. Ordinal scale can tell only whether there is a difference between two observations and in what direction the difference goes. For example, we may say one mineral is softer than another because it is damaged when they are scratched together. However, with ordinal scales we do not

know the true magnitude of observed phenomena. *Interval scales* are possible when we can tell precisely the difference between two observations. For example, temperature is expressed in interval scales such as the Celsius scales. If we measure the temperatures of two objects, we can say exactly that one has –for example– 20 more units than the other. However, we cannot say that one object has double temperature than the other. *Ratio scales* allow us to say exactly that. For example, the lengths of two objects is expressed on a ratio scale. In this case we can say that an object is twice as long as the other. Interested readers will find a classic treatment of scaling methods in [38].

Fitness Space supports nominal and ordinal scales to evaluate evolutionary experiments. For example, the diagonal between the lower-left corner and the upper-right corner defines a continuum between conventional optimization approaches and self-organization of autonomous systems (figure 5). Although the point of separation between the two approaches is fuzzy, we can say that experiments falling in the lower-left region are concerned with automatic engineering whereas experiments falling in the upper-right region are concerned with emergent autonomous systems.

We can also order two experiments according to the components of their fitness functions. For example, consider two imaginary experiments aimed at evolving walking controllers for legged robots. The components used in the fitness functions are:

- a =oscillation frequency of leg controller (functional, internal);
- b =distance covered (behavioral, external);
- c =state of motors (behavioral, internal);
- d =state of bump sensor under the belly (behavioral, internal);
- 2 =constant (affects relative position along implicit/explicit dimension).

Fitness function f^1

$$f^1 = (2 * a) + (b * d) \quad (1)$$

is composed of two additive parts. The first part rewards the controller for producing pre-determined oscillation frequencies that correspond to a desired motion pattern for each leg. This part has a strong weight (factor 2). The second part instead adds to it rewards for the distance covered by the robot body multiplied by the state of the sensors. In other words, robots that move longer without creeping over the floor get higher fitness.

Fitness function f^2

$$f^2 = c + d \quad (2)$$

has two parts too. The first part is maximized by the quantity of current sent to the motors and the second is maximized when the robot does not touch the floor. In other words, robots that keep their legs moving but do not stay on the floor will receive higher fitness.

Function f^1 therefore is *less implicit, less internal, and less behavioral* than function f^2 . Assuming that all other conditions are the same, fitness f^1 can generate efficient and specific gaits (depending on the values of a), but it may take more generations and a larger population because the number of constraints are likely to make the fitness landscape very hard. Also, it requires the use of additional hardware to measure the distance covered by the robot. Fitness f^2 instead is not guaranteed to generate efficient gaits, in fact it may well generate robots that dance without covering much forward distance. However, the evolved solutions will be more dependent on the interactions between the robot and its environment. Also, this function does not require additional devices and is portable to a range of different robots with unknown dynamics and kinematics because it does not require functional knowledge of the system.

When comparing evolutionary robots though the fitness function is not the only factor that determines the outcomes and evolvability of a system. Although it plays

an important role, other determinant factors include the type of sensors used, the environmental setup, the type of challenge that is being addressed, and aspects of the controller architecture and genetic encoding.

4 Evolution of Adaptation

In the following sections we shall introduce a methodology to evolve robots that are capable to withstand different sources of environmental changes both during and after evolutionary selection. This is part of our effort to make artificial evolution more applicable to real-world applications without compromising on the autonomy of the robot. The method is based on evolution of mechanisms for adaptation of parameters, instead of evolution of the parameters themselves as in the conventional approach. The fitness function is internal, behavioral, and not too explicit.

4.1 Coping with Change

The situated nature of Evolutionary Robotics is such that often evolved controllers find surprisingly simple –yet efficient– solutions that capitalize upon unexpected invariants of the interaction between the robot and its environment. For example, a robot evolved for the ability to discriminate between shapes can do so without resorting to expensive image processing techniques by simply checking the correlated activity of two receptors located in strategic positions on the retinal surface [21]. Analogously, a robot evolved for finding a hidden location can display the performances similar to those obtained by rats trained under the same conditions without resorting to complex environmental representations by using simple sensory-motor sequences that exploit geometric invariants of the environment [26]. The remarkable simplicity⁷ and efficiency of these solutions is a clear advantage for fast and real-time operation required from autonomous robots, but it raises the issue of robustness when environmental conditions change. Environmental changes can be a problem also for other approaches (programming, learning, e.g.) to the extent in which the sources of change have not been considered during system design, but they are even more so for evolved systems because these often rely on environmental aspects that are often not predictable by an external observer.

Environmental changes can be induced by several factors such as modifications of the sensory appearance of objects (e.g., different light conditions), changes in sensor response, re-arrangement of environment configuration, transfer from simulated to physical robots, and transfer across different robotic platforms.

Some authors have suggested to improve the robustness of evolved systems by adding noise [29, 22] and by evaluating fitness values in several different environments [37]. However, both techniques imply that one knows in advance what makes the evolved solution brittle in the face of future changes in order to choose a suitable type of noise and of environmental variability during evolutionary training. Another approach consists of combining evolution and learning “during life” of the individual (see [3] for a comprehensive review of the combination of evolution and learning). This strategy not only can improve the search properties of artificial evolution, but can also make the controller more robust to changes that occur faster than the evolutionary time scale (i.e., changes that occur during the life of an individual) [31]. This is typically achieved by evolving neural controllers that learn with an off-the-shelf algorithm, such as reinforcement learning or back-propagation, starting from synaptic

⁷This does not imply that evolutionary approaches are restricted to forms of reactive intelligence; see for example [8]

weights specified on the genetic string of the individual [1, 32]. Only initial synaptic weights are evolved. A limitation of this approach is the “Baldwin effect”, whereby the evolutionary costs associated with learning give a selective advantage to the genetic assimilation of learned properties and consequently reduce the plasticity of the system over time [27].

Here we suggest to *evolve the adaptive characteristics* of a controller instead of combining evolution with off-the-shelf algorithms. The method consists of encoding on the genotype a set of four local Hebb rules for each synapse, but *not the synaptic weights*, and let these synapse use these rules to adapt their weights online starting always from random values at the beginning of the life. Since the synaptic weights are not encoded on the genetic string, there cannot be genetic assimilation of abilities developed during life. In other words, these controller can rely less on genetically-inherited invariants and must develop on-the-fly the connection weights necessary to achieve the task. At the same time, the evolutionary cost of adaptation (i.e., the time and energy spent to adapt goes to the detriment of the individual’s fitness) implicitly puts pressure for the generation of fast-adaptive architectures.

In preliminary investigations comparing evolution of genetically-determined weights with evolution of adaptive controllers on a simple navigation task, we have shown that the latter approach generates similarly-good performances in less generations [10] by taking advantage of the combined search methods. Later, we showed that evolution of adaptive controllers significantly alters the performance of robots that must cope with dynamic environments and described an experiment where co-evolutionary adaptive predators adapted on line to co-evolutionary prey robots [11].

Here we describe a new set of experiments designed to further show that this approach can generate more complex controllers and test its robustness to environmental changes that were not included during evolutionary training. For what concerns adaptation to change, here we focus on transfer of evolved controllers across different robotic platforms whose sensory-motor characteristics require partial re-configuration of the control system. In another set of forthcoming papers, we also show that this approach is effective for environmental changes that involve new sensory characteristics and new spatial relationships of the environment [40] and in transfers from simulations to physical robots without additional evolution [39].

In the next sections we give an overview of the evolutionary method and describe its application to a complex sequential task. We then present the results on the transfer of evolved across different robotic platforms. Finally, we discuss the future perspectives of this new evolutionary approach.

4.2 Encoding Mechanisms of Adaptation

The artificial chromosome encodes a set of four modification rules for each component of the neural network (components can be individual synapses or groups of synapses that converge towards the same neuron, as we shall see below), but not the synaptic strengths of the network. Whenever an artificial chromosome is decoded into a neural controller, the synaptic strengths are set to small random values. This means that the robot will initially display random actions both at generation 0 and at later generations. However, as time goes the synapses start to change their value using the genetically specified rules every 100 ms (the time necessary for a full sensory-motor loop on the physical robot). Notice that synaptic adaptation occurs on-line while the robot moves and that the network self-organizes without external supervision and reinforcement signals. The fitness function is evaluated along the whole duration of the robot “life”. This introduces an implicit learning cost [27] that gives selective advantage to individuals that can adapt faster. At the end of the life, the final synaptic

strengths are not “written back” into the artificial chromosome.⁸

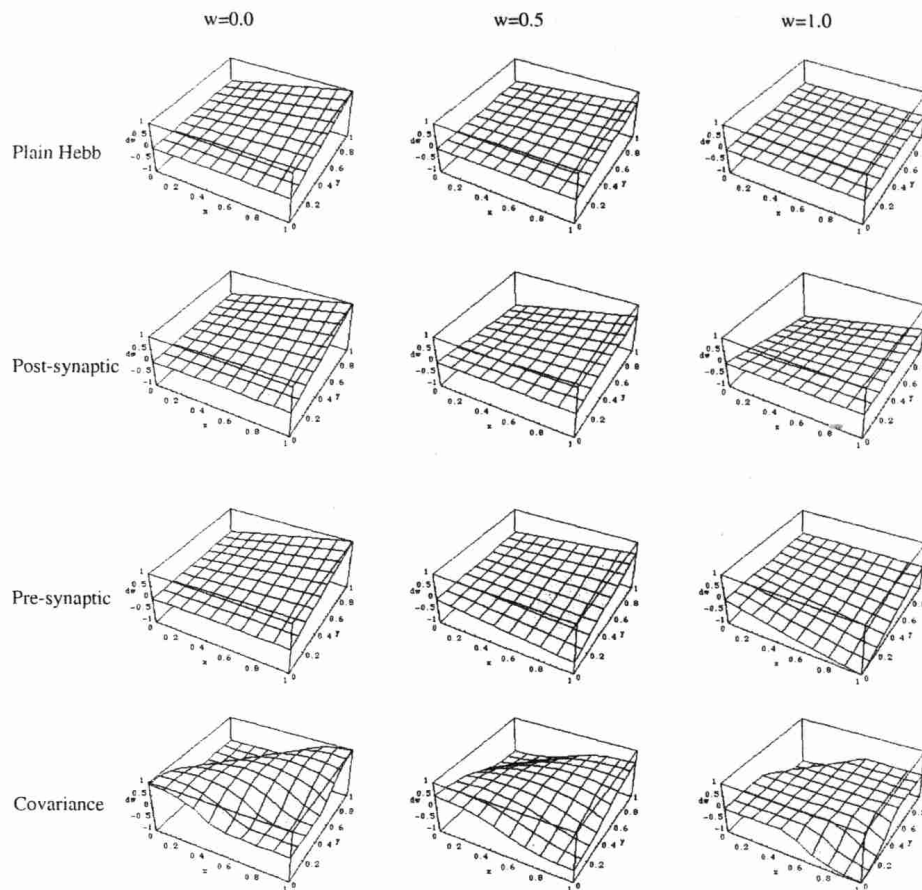


Figure 7: Synaptic change for each of the four Hebb rules. Notice that this is the *amount of change* Δw added to the synapses, not the synaptic strength. Each graph indicates the amount of change as a function of instantaneous presynaptic x and postsynaptic y activity. The amount of change also depends on the current strength w of the synapse so that synapses are always bound between 0 and 1. Three graphs are shown for each rule, in the case of current strength 0.0, 0.5, and 1.0 respectively.

We have selected four types of modification rules (figure 7) to be encoded on the artificial chromosome. The choice has been based on neurophysiological findings and on computational constraints of local adaptation. In other words, these rules capture some of the most common mechanisms of local synaptic adaptation found in the nervous systems of mammals [42]. These rules were modified in order to satisfy the following constraints. Synaptic strength could not grow indefinitely, but was kept in the range $[0, 1]$ by means of a self-limiting mechanism which depended on synaptic strength. Because of this self-limiting factor, a synapse could not change sign, which was genetically specified, but only strength. Each synaptic weight w_{ij} is randomly

⁸In other words, we use Darwinian evolution instead of Lamarckian evolution where the effects of learning are encoded in the artificial chromosome. See [43] for an experimental comparison between these two types of evolution in changing environments.

Encoding	Bits for one synapse / node				
Genotype	1	2	3	4	5
A	sign	strength			
B	sign	Hebb rule		rate	
C	sign	strength		noise	

Table 2: Genetic encoding of synaptic parameters for Synapse Encoding and Node Encoding. In the latter case the sign encoded on the first bit is applied to all outgoing synapses whereas the properties encoded on the remaining four bits are applied to all incoming synapses. A: Genetically determined controllers; B: Adaptive synapse controllers; C: Noisy synapse controllers.

initialized at the beginning of the individual's life and can be updated after every sensory-motor cycle (100 ms),

$$w_{ij}^t = w_{ij}^{t-1} + \eta \Delta w_{ij},$$

where $0.0 < \eta < 1.0$ is the learning rate and Δw_{ij} is one of the four modification rules specified in the genotype:⁹

1. *Plain Hebb rule*: can only strengthen the synapse proportionally to the correlated activity of the pre- and post-synaptic neurons.

$$\Delta w = (1 - w) xy \quad (3)$$

2. *Postsynaptic rule*: behaves as the plain Hebb rule, but in addition it weakens the synapse when the postsynaptic node is active but the presynaptic is not.

$$\Delta w = w(-1 + x)y + (1 - w)xy \quad (4)$$

3. *Presynaptic rule*: weakening occurs when the presynaptic unit is active but the postsynaptic is not.

$$\Delta w = wx(-1 + y) + (1 - w)xy \quad (5)$$

4. *Covariance rule*: strengthens the synapse whenever the difference between the activations of the two neurons is less than half their maximum activity, otherwise the synapse is weakened. In other words, this rule makes the synapse stronger when the two neurons have synchronous activity.

$$\Delta w = \begin{cases} (1 - w)\mathcal{F}(x, y) & \text{if } \mathcal{F}(x, y) > 0 \\ (w)\mathcal{F}(x, y) & \text{otherwise} \end{cases} \quad (6)$$

where $\mathcal{F}(x, y) = \tanh(4(1 - |x - y|) - 2)$ is a measure of the difference between the presynaptic and postsynaptic activity. $\mathcal{F}(x, y) > 0$ if the difference is bigger or equal to 0.5 (half the maximum node activation) and $\mathcal{F}(x, y) < 0$ if the difference is smaller than 0.5.

The genes are composed of five bits. The first bit represents the sign of the synapse. What is encoded on the remaining four bits depends on the evolutionary condition chosen (table 4.2), namely:

⁹These four rules co-exist within the same network.

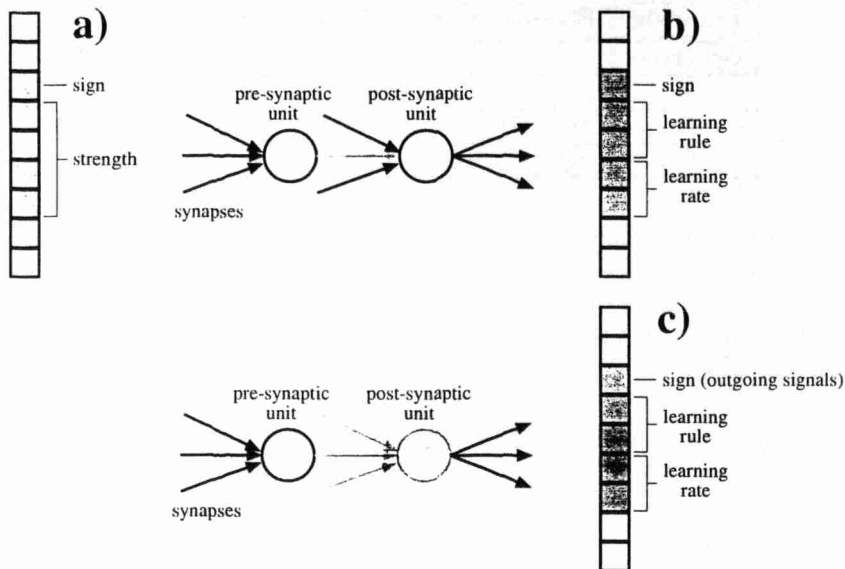


Figure 8: Different type of Genetic Encoding. **a)** Synapse Encoding for genetically-determined networks: The sign and strength of each synapse is encoded on the genotype. **b)** Synapse Encoding for adaptive networks: The sign, four learning rules, and the learning rate of each synapse is encoded on the genotype. **c)** Node Encoding for adaptive networks: all outgoing signals have the same sign and all incoming synapses have the same learning rate and learning rule. In cases **b** and **c**, the initial strength of the synapse is always set to small random values. Node Encoding cannot be applied to for genetically-determined synapses.

1. *Genetically-determined*: 4 bits encode the synaptic strength. This value is constant during "life".
2. *Adaptive synapses*: 2 bits encode 4 adaptive rules and 2 bits the learning rate. Synaptic weights are always randomly initialized at the beginning of an individual's life and then updated according to their own adaptation rule.
3. *Noisy synapses*: 2 bits encode the weight strength and 2 bits a noise range. The synaptic strength is genetically determined at birth, but a random value extracted from the noise range is freshly computed and added after each sensory motor cycle. This latter condition is used as a control condition to check whether the effects of Hebbian adaptation (condition above) are equivalent to random synaptic variability.

Two types of genetic encoding have been considered (figure 8). In the simplest case, known as *Synapse Encoding*, each synapse has its properties (one of the three described above) specified in the artificial chromosome. In *Node Encoding* the properties describe are attributes of each node: the sign bit corresponds to the sign of all the outgoing synapses while the remaining 4 bits apply to all incoming synapses for that node. Synapse Encoding allows a detailed definition of the controller, but for a fully connected network of N neurons the genetic length is proportional to N^2 . Instead Node Encoding requires a much shorter genetic length (proportional to N), but it al-



Figure 9: A mobile robot Khepera equipped with a vision module gains fitness by staying on the gray area only when the light is on. The light is normally off, but it can be switched on if the robot passes over the black area positioned on the other side of the arena. The robot can detect ambient light and the color of the wall, but not the color of the floor.

lows only a rough definition of the controller. In recent work [15] we showed that our evolutionary adaptive approach does not need a lengthy direct representation because the actual weights of the synapses are always shaped at run-time by the genetically specified rules. However, this is not possible in the traditional approaches where it is necessary to assign good initial weights to the controller. Therefore, the experiments reported in this paper compare evolution of genetically-determined networks using Synapse Encoding with evolution of adaptive networks using Node Encoding.

4.3 A Sequential Task: The “Light-Switching” Problem

In this set of experiments, we have compared the performance of evolutionary adaptive controllers with respect to evolution of synaptic weights and evolution of noisy synapses in a sequential task.

A mobile robot Khepera equipped with a vision module is positioned in the rectangular environment shown in figure 9. A light bulb is attached on one side of the environment. This light is normally off, but it can be switched on when the robot passes over a black-painted area on the opposite side of the environment. A black stripe is painted on the wall over the light-switch area. Each individual of the population is tested on the same robot, one at a time, for 500 sensory motor cycles, each cycle lasting 100 ms. At the beginning of an individual’s life, the robot is positioned at a random position and orientation and the light is off.

The fitness function is given by the number of sensory motor cycles spent by the robot on the gray area beneath the light bulb *when the light is on* divided by the total number of cycles available (500). In order to maximize this fitness function, the robot should find the light-switch area, go there in order to switch the light on, and then move towards the light as soon as possible, and stand on the gray area. Since this sequence of actions takes time (several sensory motor cycles), the fitness of a robot will never be 1.0. Also, a robot that cannot manage to complete the entire sequence will be scored with 0.0 fitness. A light sensor placed under the robot is used to detect the color of the floor—white, gray, or black— and passed to a host computer in order to switch on the light bulb and compute fitness values. The output of this sensor is *not* given as input to the neural controller. After 500 sensory motor cycles, the light is switched off and the robot is repositioned by applying random speeds to the wheels for 5 seconds.

Notice that the fitness function does not explicitly reward this sequence of actions (which is based on our external perspective), but only the final outcome of the sequence

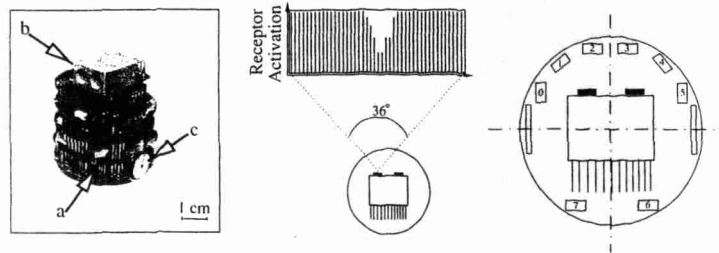


Figure 10: The Khepera robot used in the experiments. Infrared sensors (a) measure object proximity and light intensity. The linear vision module (b) is composed of 64 photoreceptors covering a visual field of 36° (center). The output of the controller generates the motor commands (c) for the robot. Right figure shows the sensory disposition of the Khepera robot.

of behaviors chosen by the robot. In Fitness Space (figure 5) this function is behavioral, internal (the computation is based on variables read through the sensors of the robot), and almost implicit (only one component is used).

The robot has The controller we have used in our experiments is a fully-recurrent discrete-time neural network. It has access to three types of sensory information from the robot (figures 10 and 11):

1. *Infrared light*: the active infrared sensors positioned around the robot (figure 10, a) measure the distance from objects. Their values are pooled into four pairs and the average reading of each pair is passed to a corresponding neuron.
2. *Ambient light*: the same sensors are used to measure ambient light too. These readings are pooled into three groups and the average values are passed to the corresponding three light neurons.
3. *Vision*: the vision module (figure 10, b) consists of an array of 64 photoreceptors covering a visual field of 36° (figure 10, center). The visual field is divided up in three sectors and the average value of the photoreceptors (256 gray levels) within each sector is passed to the corresponding vision neuron.

Two motor neurons are used to set the rotation speed of the wheels (figure 10, c). Neurons are updated every 100 ms.

The fitness results reported in figure 12 show that individuals with adaptive synapses and Node Encoding (graph on the left) are much better than individuals with genetically-determined synapses and Synapse Encoding (graph in the center) in that:

1. Both the fitness of the best individuals and of the population report higher values (0.6 against 0.5).
2. They reach the best value obtained by genetically-determined individuals in less than half generations (40 against more than 100).

Figure 13 shows the behaviors of two best individuals evolved with adaptive synapses and Node Encoding (left) and with genetically-determined weights and Synapse Encoding (right). In both cases individuals aim at the area with the light switch¹⁰ and, once the light is turned on, they move towards the light and remain there. The better

¹⁰Their performance is badly affected if the vision input is disabled, indicating that they do not use random search to locate the switch (data not shown).

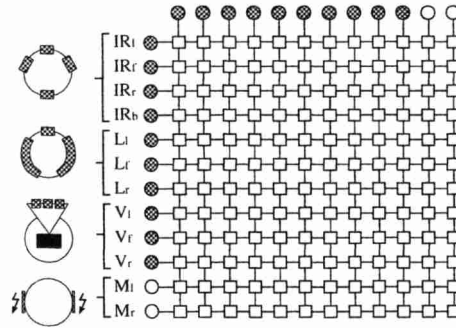


Figure 11: The neural controller is a fully-recurrent discrete-time neural network composed of 12 neurons giving a total of $12 \times 12 = 144$ synapses (here represented as small squares of the unfolded network). 10 sensory neurons receive additional input from one corresponding pool of sensors positioned around the body of the robot shown on the left (l=left; r=right; f=front; b=back). \vec{IR} =Infrared Proximity sensors; \vec{L} =Ambient Light sensors; \vec{V} =vision photoreceptors. Two motor neurons \vec{M} do not receive sensory input; their activation sets the speed of the wheels ($M_i > 0.5$ forward rotation; $M_i < 0.5$ backward rotation)

fitness of the adaptive controllers (given on the top of each box, see figure caption) is given by straight and faster trajectories showing a clear behavioral change between the first phase where they go towards the switching area and the second phase where they become attracted by the light. Instead, genetically-determined individuals display always the same looping trajectories around the environment with some attraction towards the stripe and the light. This minimalist behavior that depends on invariant geometrical relations of the environment gives them a chance to accomplish the task but with a lower performance.

Additional tests have been carried out to assess the role of adaptation in the behavior of the individuals with adaptive synapses. For example, one might argue that what matters is the sign of the synapse and not its strength as long as it is non-zero, or that adaptive synapses may have the same effect of fixed synapses with strengths set to their average values¹¹. The results reported by our control experiments and analyses clearly indicated that evolved adaptive networks modify their parameters in ways that are functionally related to the survival criterion [15].

4.4 Cross-platform Adaptation

Cross-platform transfer is a very useful feature, but we are not aware of any control system that can be transferred across different robots without changes. Cross-platform becomes useful in adaptive and evolutionary systems where initial training experiences can cause the robot to produce harmful actions. One may train (or evolve) control systems for a desktop sturdy robot like the miniature Khepera and then download them to larger and consequently more fragile robots¹². In this case, it would be desirable that the control system self-adapts to the new sensory-motor characteristics

¹¹This latter suggestion was made by Flotzinger [16] who replicated our previous experiments on Adaptive Synapses with Synapse Encoding [9]

¹²Obviously, the two robots must share some characteristics, such as type of sensors and actuators used, that allow a suitable interfacing of the control system.

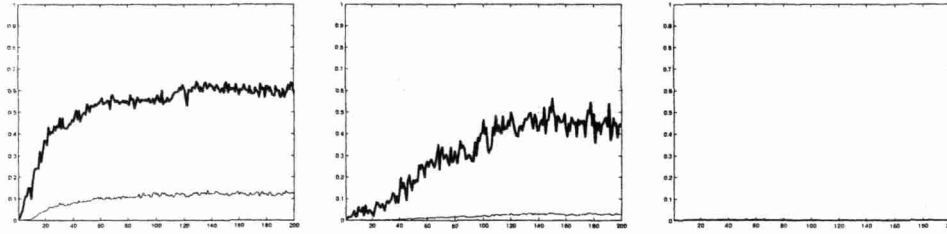


Figure 12: Comparison of adaptive synapses with Node Encoding (*left*) versus genetically-determined synapses with Synapse Encoding (*right*). Thick line=best individual; thin line=population average; dashed line=genetic diversity. Each data point is an average over 10 replications with different random initializations. Population size is 100 and 20 best individuals reproduce by making 5 copies. Crossover probability is 0.2 and mutation probability is 0.05 (per bit).

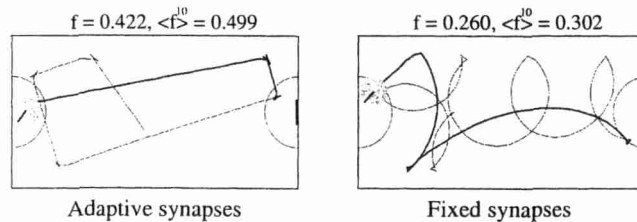


Figure 13: Behaviors of two best individuals (from last generation) with adaptive synapses and Node Encoding (*left*) and with genetically-determined synapses and Synapse Encoding (*right*). When the light is turned on, the trajectory line becomes thick. The corresponding fitness value is printed on the top of each box along with the average fitness of the same individual tested ten times from different positions and orientations.

and morphology. In previous work we have shown that this can be achieved by using incremental evolution of genetically-determined networks [10]. However, even for a simple reactive navigation behavior it took additional 20 generations to re-adapt to the new robot.

Here we test the adaptive properties of the evolutionary adaptive strategy by transferring onto a physical Koala robot (figure 14, left) the best individuals of the last generation evolved on the miniature Khepera robot. The Koala robot has six wheels driven by two motors (one on each side) and 16 infrared sensors (figure 14, right) with a different and stronger detection range.

Taking advantage of the setup offered by *TeleRoboLab*¹³, a mobile robot Koala equipped with a vision module is positioned in the rectangular environment shown in figure 15. As in the previous experiment with the Khepera robot, the Koala robot must find the light-switching area, go there in order to switch the light on, and then move towards the light as soon as possible and stay there in order to score fitness points.

An external positioning system emitting laser beams at predefined angles and fre-

¹³<http://TeleRoboLab.epfl.ch> is a web site created and maintained by P. Saucy and F. Mondada that allows an external user to teleoperate a Koala robot in a physical environment.

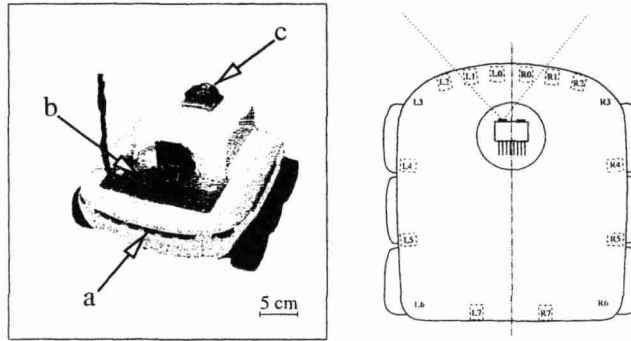


Figure 14: The Koala robot used in the experiments. Infrared sensors (a) measure object proximity and light intensity. The linear vision module (b) is the same as used in the experiments with the Khepera robot. The localization module (c) provides the position of the robot at every time step. Right figure shows the sensory layout of the Koala robot. Only 8 equally-spaced sensors are selected as input to the network.

quencies is positioned on the top of the environment and the Koala robot is equipped with an additional turret capable of detecting laser and computing in real-time the robot displacement. This information is used in order to control the light and to compute the fitness. The performance of adaptive individuals is not affected by the transfer from the Khepera robot (striped bars) to the Koala robot (dotted bars), whereas genetically-determined individuals report a significative fitness loss. Individuals with noisy synapses are not affected by the transfer because their behavior is always random and not effective in both Khepera and Koala robots (for more details see [39]).

Individuals evolved in simulation for the Khepera robot display a satisfactory behavior when tested on the Koala robot. They correctly approach the light-switching area and they are clearly attracted by light (figure 16, left). As in the case of real Khepera robot, once arrived under the light the Koala robot moves around the fitness area while remaining close to it until the testing time is over.

On the other hand, genetically-determined individuals (center) perform spiralling trajectories around the environment and do not display any attraction by the black stripe or the light. They eventually manage to pass through the light-switching area, turn the light on, and occasionally score fitness points passing through the fitness area. In several cases, genetically-determined individuals get stuck on the walls of the environment (behaviors not shown). Individuals with noisy synapses (right) score a low performance because their strategy is based in random navigation.

5 A look ahead

Over the last few years the number of projects in Evolutionary Robotics around the world has been constantly increasing. In this chapter we have described the main approaches to this field and provided a personal interpretation of more and less significant areas of research for the years to come.

We have also suggested Fitness Space as a framework to design, assess, and compare fitness functions with respect to the outcome and evolvability of evolutionary robots. Although the fitness function is not the only factor that characterize an evo-

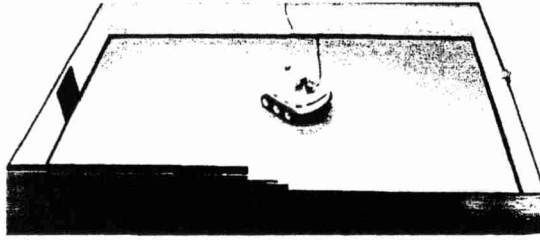


Figure 15: A mobile robot Koala equipped with a vision module gains fitness by staying near the lamp (right side) only when the light is on. The light is normally off, but it can be switched on if the robot passes near the black stripe (left side) positioned on the other side of the arena. Position of the robot is controlled by an external positioning system and passed to the computer in order to control the light and to compute the fitness.

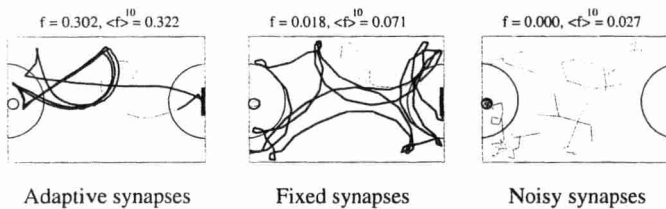


Figure 16: Behaviors of individuals with adaptive synapses (left), genetically-determined synapses (center), and noisy synapses (right) tested on the Koala robot. Individuals belong to the last generation evolved in simulation for the Khepera robot.

lutionary experiment, we think that Fitness Space is a useful tool to guide the setup of an experiment and to clarify where an experiment stands with respect to others in the literature.

The success of Evolutionary Robotics will ultimately depend by its ability to generate new robotic systems that could not be designed with conventional techniques. This means the ability to generate robots that display complex skills and can cope with unpredictable changes.

We think that the evolutionary method presented in the second part of this paper represents a significative step forward towards making Evolutionary Robotics applicable to real-world applications of autonomous robotics. In scenarios like those—for example—of robots probing an asteroid surface or robots interacting with an handicapped person it is impossible to evolve the control system on the spot (not even incrementally). However, one might reproduce the working conditions in the laboratory to some degree of approximation and evolve the adaptive controller in there. The controller would then be transferred on the final robot and let free to adapt to actual working conditions in a few seconds.

We also think that our adaptive strategy will be useful for evolving more complex and powerful control architectures. In current methods there is a trade-off between the complexity the genotype/phenotype mapping and the evolvability of such systems which is partly due to the fact that the phenotype largely depends on genetic instructions. By evolving the adaptive characteristics along with other high-level parameters

(position and type of nodes, e.g.) of the controller, one may obtain simpler genetic encodings and a higher tolerance to mutations. This would make the evolved controllers more viable, add neutrality to the genetic landscape, and ultimately improve evolvability.

6 Acknowledgements

The second part this manuscript (section 4) is based on a more extended report that will appear elsewhere [39]. Joseba Urzelai is supported by grant nr. BF197.136-AK from the Basque government. Thanks to Patrick Saucy for his help on the cross-platform experiment.

References

- [1] D. H. Ackley and M. L. Littman. Interactions between learning and evolution. In C.G. Langton, J.D. Farmer, S. Rasmussen, and C. Taylor, editors, *Artificial Life II: Proceedings Volume of Santa Fe Conference*, volume XI. Addison Wesley: series of the Santa Fe Institute Studies in the Sciences of Complexities, Redwood City, CA, 1992.
- [2] M. A. Bedau, E. Snyder, C. T. Brown, and N. H. Packard. A comparison of evolutionary activity in artificial evolving systems and in the biosphere. In P. Husbands and I. Harvey, editors, *Proceedings of the 4th European Conference on Artificial Life*, Cambridge, MA, 1997. MIT Press.
- [3] R. K. Belew and M. Mitchell, editors. *Adaptive Individuals in Evolving Populations: Models and Algorithms*. Addison-Wesley, Redwood City, CA, 1996.
- [4] V. Braitenberg. *Vehicles. Experiments in Synthetic Psychology*. MIT Press, Cambridge, MA, 1984.
- [5] D. Cliff and G. F. Miller. Tracking the Red Queen: Measurements of adaptive progress in co-evolutionary simulations. In F. Morán, A. Moreno, J. J. Merelo, and P. Chacón, editors, *Advances in Artificial Life: Proceedings of the Third European Conference on Artificial Life*, pages 200–218. Springer Verlag, Berlin, 1995.
- [6] D. Floreano. Emergence of Home-Based Foraging Strategies in Ecosystems of Neural Networks. In J. Meyer, H. L. Roitblat, and S. W. Wilson, editors, *From Animals to Animats II: Proceedings of the Second International Conference on Simulation of Adaptive Behavior*. MIT Press-Bradford Books, Cambridge, MA, 1993.
- [7] D. Floreano. Evolutionary Robotics in Artificial Life and Behavior Engineering. In T. Gomi, editor, *Evolutionary Robotics. From Intelligent Robots to Artificial Life*. AAI Books, Ontario, Canada, 1998.
- [8] D. Floreano and F. Mondada. Evolution of homing navigation in a real mobile robot. *IEEE Transactions on Systems, Man, and Cybernetics-Part B*, 26:396–407, 1996.
- [9] D. Floreano and F. Mondada. Evolution of plastic neurocontrollers for situated agents. In P. Maes, M. Mataric, J-A. Meyer, J. Pollack, H. Roitblat, and S. Wilson, editors, *From Animals to Animats IV: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior*, pages 402–410. MIT Press-Bradford Books, Cambridge, MA, 1996.

- [10] D. Floreano and F. Mondada. Evolutionary Neurocontrollers for Autonomous Mobile Robots. *Neural Networks*, 11:1461-1478, 1998.
- [11] D. Floreano and S. Nolfi. Adaptive behavior in competing co-evolving species. In P. Husbands and I. Harvey, editors, *Proceedings of the 4th European Conference on Artificial Life*, Cambridge, MA, 1997. MIT Press.
- [12] D. Floreano and S. Nolfi. God Save the Red Queen! Competition in Co-evolutionary Robotics. In J. Koza, K. Deb, M. Dorigo, D. Fogel, M. Garzon, H. Iba, and R. L. Riolo, editors, *Proceedings of the 2nd International Conference on Genetic Programming*, San Mateo, CA, 1997. Morgan Kaufmann.
- [13] D. Floreano, S. Nolfi, and F. Mondada. Competitive Co-Evolutionary Robotics: From Theory to Practice. In R. Pfeifer, B. Blumberg, J-A. Meyer, and S. Wilson, editors, *From Animals to Animats V: Proceedings of the Fifth International Conference on Simulation of Adaptive Behavior*. MIT Press-Bradford Books, Cambridge, MA, 1998.
- [14] D. Floreano, S. Nolfi, and F. Mondada. Co-evolution and ontogenetic change in competing robots. *Robotics and Autonomous Systems*, in the press, 1999.
- [15] D. Floreano and J. Urzelai. Evolution of Neural Controllers with Adaptive Synapses and Compact Genetic Encoding. In D. Floreano, J-D. Nicoud, and F. Mondada, editors, *Advances in Artificial Life*. Springer Verlag, Berlin, 1999.
- [16] D. Flotzinger. Evolving plastic neural network controllers for autonomous robots. Msc dissertation 9580131, COGS, University of Sussex at Brighton, 1996.
- [17] C. R. Gallistel, editor. *The Organization of Learning*. MIT Press, Cambridge, MA, 1990.
- [18] J. J. Gibson. *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston, 1979.
- [19] F. Gruau. Automatic definition of modular neural networks. *Adaptive Behavior*, 3:151-183, 1994.
- [20] I. Harvey. Artificial evolution for real problems. In T. Gomi, editor, *Evolutionary Robotics*, pages 187-220. AAI Books, Ontario, Canada, 1997.
- [21] I. Harvey, P. Husbands, and D. Cliff. Seeing The Light: Artificial Evolution, Real Vision. In D. Cliff, P. Husbands, J. Meyer, and S. W. Wilson, editors, *From Animals to Animats III: Proceedings of the Third International Conference on Simulation of Adaptive Behavior*. MIT Press-Bradford Books, Cambridge, MA, 1994.
- [22] N. Jakobi. Half-baked, ad-hoc and noisy: Minimal simulations for evolutionary robotics. In P. Husbands and I. Harvey, editors, *Proceedings of the 4th European Conference on Artificial Life*, Cambridge, MA, 1997. MIT Press.
- [23] M. Komosinski and Sz. Ulatowski. Framsticks: Towards a simulation of a nature-like world, creatures and evolution. In D. Floreano, J-D. Nicoud, and F. Mondada, editors, *Advances in Artificial Life - ECAL99*, Berlin, 1999. Springer Verlag.
- [24] M. A. Lewis, A. H. Fagg, and A. Solidum. Genetic programming approach to the construction of a neural network for a walking robot. In *Proceedings of IEEE International Conference on Robotics and Automation*, pages 2618-2623. IEEE Press, 1996.
- [25] H. H. Lund, J. Hallam, and W-P. Lee. Evolving robot morphology. In *Proceedings of the IEEE 4th International Conference on Evolutionary Computation*. IEEE Press, 1997.

- [26] H. H. Lund and O. Miglino. Evolving and Breeding Robots. In P. Husbands and J.-A. Meyer, editors, *Proceedings of the First European Workshop on Evolutionary Robotics*. Springer Verlag, 1998.
- [27] G. Mayley. Landscapes, Learning Costs and Genetic Assimilation. *Evolutionary Computation*, 4(3):213–234, 1996.
- [28] F. Menczer and R. K. Belew. Latent energy environments. In R. K. Belew and S. Mitchell, editors, *Plastic Individuals in Evolving Populations*. Addison Wesley, Redwood City, CA, 1993.
- [29] O. Miglino, H. H. Lund, and S. Nolfi. Evolving Mobile Robots in Simulated and Real Environments. *Artificial Life*, 2:417–434, 1996.
- [30] F. Mondada and D. Floreano. Evolution of neural control structures: some experiments on mobile robots. *Robotics and Autonomous Systems*, 16:183–195, 1995.
- [31] S. Nolfi and D. Floreano. Learning and evolution. *Autonomous Robots*, 7(1):in the press, 1999.
- [32] S. Nolfi, O. Miglino, and D. Parisi. Phenotypic Plasticity in Evolving Neural Networks. In J.-D. Nicoud and P. Gaussier, editors, *Proceedings of the conference From Perception to Action*. IEEE Computer Press, Los Alamitos, CA, 1994.
- [33] J. B. Pollack, H. Lipson, S. Ficici, P. Funes, G. Hornby, and R. Watson. Evolutionary techniques in physical robotics. In A. Thompson, J. Miller, T. C. Fogarty, and P. Thomson, editors, *Proceedings of the Third International Conference on Evolvable Systems: from Biology to Hardware*, Berlin, 2000. Springer Verlag.
- [34] E. Ronald, M. Sipper, and M. S. Capcarrère. Design, Observation, Surprise! A Test of Emergence. *Artificial Life*, 5(3):225–240, 1999.
- [35] K. Sims. Evolving 3D Morphology and Behavior by Competition. In R. Brooks and P. Maes, editors, *Proceedings of the Fourth Workshop on Artificial Life*, pages 28–39, Boston, MA, 1994. MIT Press.
- [36] A. Thompson. Evolving electronic robot controllers that exploit hardware resources. In F. et al. Moran, editor, *Advances in Artificial Life: Proc. of ECAL95*. Springer Verlag, Barcelona, 1995.
- [37] A. Thompson. On the automatic design of robust electronics through artificial evolution. In D. Mange M. Sipper and A. Prez-Urbe, editors, *Proceedings of the 2nd International Conference on Evolvable Systems: From biology to hardware (ICES98)*, pages 13–24. Springer-Verlag, Berlin, 1998.
- [38] W. Torgerson. *Theory and methods of scaling*. Wiley, New York, 1958.
- [39] J. Urzelai and D. Floreano. Evolutionary Robotics: Coping with Environmental Change. In J. et. al. Koza, editor, *Second International Conference on Genetic and Evolutionary Computation*, San Mateo, CA, 2000. Morgan Kaufmann.
- [40] J. Urzelai and D. Floreano. Evolutionary robots with fast adaptive behavior in new environments. In T. C. Fogarty, J. Miller, A. Thompson, and P. Thomson, editors, *Third International Conference on Evolvable Systems: From Biology to Hardware (ICES2000)*, Berlin, 2000. Springer-Verlag.
- [41] R. Watson, S. Ficici, and J. B. Pollack. Embodied evolution: Embodying an evolutionary algorithm in a population of robots. In P. Angeline, M. Michaeliewicz, G. Schonauer, X. Yao, and Z. Zalzalá, editors, *Proceedings of the 1999 Congress on Evolutionary Computation*. IEEE Press, 1999.
- [42] D. Willshaw and P. Dayan. Optimal plasticity from matrix memories: What goes up must come down. *Neural Computation*, 2:85–93, 1990.

- [43] Y. Yamamoto, T. Sasaki, and M. Tokoro. Adaptability of Darwinian and Lamarckian Populations toward an Unknown New World. In D. Floreano, J-D. Nicoud, and F. Mondada, editors, *Advances in Artificial Life*. Springer Verlag, Berlin, 1999.

First Evolution Experiments on a Physical CAM-Brain Machine

Hugo de Garis et al.
Brain Builder Group
Starlab
Rue Engeland 555
B-1180 Brussels
Belgium
degaris@starlab.net
<http://foobar.starlab.net/> degaris
Tel: +32 2 270 0740

February 2001

Abstract

This paper presents results of some of the first evolution experiments undertaken on an actual CAM-Brain Machine (CBM), using the hardware itself and not software simulations. A CBM is a specialised piece of programmable (evolvable) hardware that uses Xilinx XC6264 programmable FPGA chips to grow and evolve, at electronic speeds, 3D cellular automata (CA) based neural network circuit modules of some 1000 neurons each. A complete run of a genetic algorithm (e.g. with 100 generations and a population size of 100) is executed in a few seconds. 64000 of these modules can be evolved separately according to the fitness definitions of human "EEs" (evolutionary engineers) and downloaded one by one into a gigabyte of RAM. Human "BAs" (brain architects) then interconnect these modules "by hand" according to their artificial brain architectures. The CBM then updates the binary neural signaling of the artificial brain (with 64000 "hand"

interconnected modules, i.e. 75 million neurons) at a rate of 130 billion CA cell updates a second, which is fast enough for real time control of robots. Before such multi-moduled artificial brains can be constructed, it is essential that the quality of the evolution (the "evolvability") of individual modules be adequate. This paper reports on the first evolution results obtained on CBM hardware.

1 Introduction

It has been a decade long dream of the first author that it might be possible to evolve neural network circuits at electronic speeds and in such large numbers that building artificial brains would become practical. In 1993, a research project was conceived [1] which aimed to evolve neural networks embedded in a medium of cellular automata. By storing the state of each CA cell with 1 or 2 bytes, today's technology, with its gigabytes of RAM in workstations, would mean a potential space of billions of CA cells in which to grow and evolve neural nets - more than enough in which to place an artificial

brain. Software simulation studies showed that such a thing was possible. The initial CA based neural net model was too complex to be implemented in electronics, so a simplified model was conceived in mid 1996 [6]. Soon after, a contract was signed between Genobyte, a hardware engineering company in Boulder, Colorado, USA, and the first author's previous lab ATR in Kyoto, Japan, to design and construct a specialised piece of programmable (evolvable) hardware (called a CAM-Brain Machine (CBM)) that uses Xilinx XC6264 chips to grow and evolve 3D CA based neural network circuit modules in about a second or so each (see Figs 1,2,3). With limited man power and budget, the first CBM was delivered in early 1999, and work on debugging and improving the CBM's performance continued throughout that year and 2000. There are now (January 2001) 4 CBMs in the world, 2 in Belgium, 1 in Japan, and 1 in the US. After many technical and commercial problems (e.g. Xilinx took the XC6200 family of chips off the market, so that Genobyte had to work with untested chips, which created all kinds of delays), the CBMs have finally come on line, and are beginning to function as designed, i.e. they implement the algorithms as originally planned, bug free. Of course, since this is an ongoing research project, the fitness definitions used to evolve the modules, and perhaps even the neural net model itself [6] used in the CBM, may need to be changed in the future, depending on the quality of the experimental results. The CBM is a programmable hardware machine, so changing its functionality is not too difficult.

The content of this paper is summarized as follows. Section 2 presents some initial results on the "evolvable capacity (EC)" of a module (i.e. the module's "MEC" = modular evolvable capacity [7]), in other words, for how long (how many ticks of the clock) can an evolved output signal of a module follow closely some randomly specified target waveform. If the CBM cannot evolve such an output with reasonable quality (i.e. have a reasonable "evolvability"), then the whole CBM approach is not worth much. Section 3 shows some early results on how it is possible to "concatenate" the outputs of several "time sliced" waveforms, so that the total waveform is a lot longer than the MEC (measured in number of clock ticks) of any module used to generate any one of the time slices. This concatenation is vital for generating long signals to control robot behaviors. Section 4 presents a successful minimal requirement test using an evolved exclusive OR (XOR) module. Section 5 presents a rather simple 1D (1 dimensional) dynamic input example of a frequency halver which generalizes. This example gives a feel for what can be evolved by the CBM. Section 6 gives an

example of a 2D dynamic input pattern detector (in this case detecting the direction of motion, up or down, of a line moving across a 2D input grid). Section 7 discusses future plans and hopes.

2 Modular Evolvable Capacities (MECs)

This section presents an attempt to see just how well the CBM, in its current state (i.e. with its current global module fitness definition [4,5] and its current neural net model [6], both of which can be hardware reprogrammed if necessary), is able to evolve modules with "acceptable" evolvability levels. By "acceptable" we mean, for the example illustrated in this section, that the evolved output signal of a module follows closely some arbitrary target waveform. There have already been a string of papers published on the basic principles of how the CBM works [2,3,4,5,6,7]. A very brief summary will be provided here. CA based neural nets grow and evolve in a 3D CA space. Once the circuits are grown, the binary neural signaling spreads over the grown network. The output signal(s) are convoluted with a digitised convolution function (which looks a bit like a gaussian or hill curve). The result of this convolution is a digital to analog conversion. The resulting analog output curve is then matched with a user specified target curve. The closer the match, the higher the fitness.

Figs. 4 to 10 show examples of target wave forms (the spikier ones) and their corresponding evolved waveforms (the flatter ones) which are supposed to be as close as possible and for as long as possible, to the former. Common sense says that there is a limit to how long such an evolved waveform can follow closely some target waveform - a finite number of bits used to evolve the module cannot generate a waveform of infinite length that remains accurate to an infinitely long target curve. This limit we call the module's MEC (modular evolvable capacity) [7]. The smaller the number of clock cycles (the duration) of the target waveform, the easier it is for the evolved waveform to follow its target waveform closely, e.g. Figs. 4, 5 and 6 (20, 50 and 70 clocks (i.e. clockticks) respectively). As the number of clocks increases, the evolvable capacity of the module approaches its MEC (defined as the maximum number of clock ticks over which the evolved curve follows the target curve closely). Figs. 7 to 10 (100, 150, 200, 300 clocks) show a form of averaging effect, where the evolved curve takes the middle path along a fluctuating target curve.

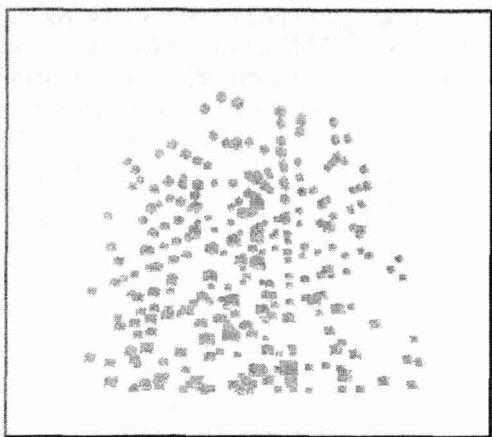


Fig. 1. First step in the growth phase, showing the initial neurons

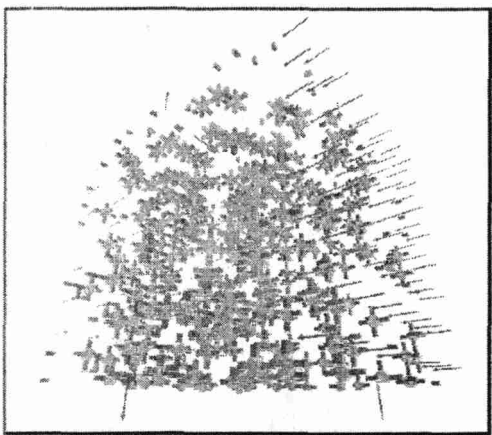


Fig. 2. Half grown module showing neurons with growing dendrites and axons.

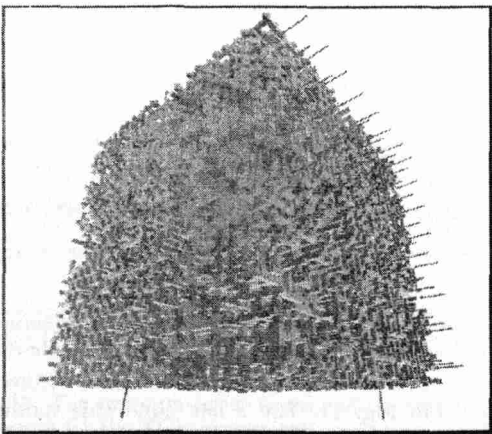


Fig. 3. Fully grown module showing neurons, dendrites and axons.

We are not satisfied with these results. We feel we can improve upon them by changing our global fitness definition (used to measure how closely the evolved and target waveforms match). At present, the earlier and later spikes in the output evolved spiketrain are equally weighted. This may be a mistake, resulting in the above averaging effect. Perhaps with a more appropriate weighting, a better accuracy, and hence a better evolvability may be achieved.

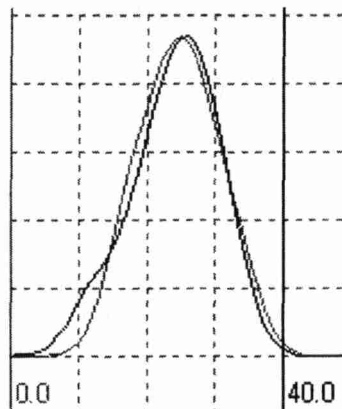


Fig. 4. Target (dark line) and response (grey line) of evolved module over 20 clocks

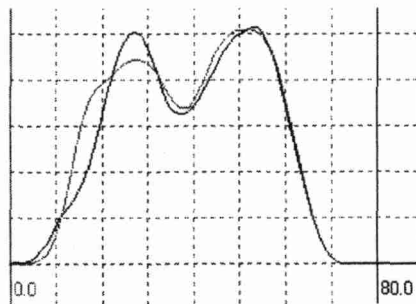


Fig. 5. Target (dark line) and response (grey line) of evolved module over 50 clocks

3 Concatenator

The limited MEC of a CBM module (estimated to be less than a few hundred clockticks) presents a severe conceptual problem to the brain building research field. One of the major application areas of artificial brains will be robotics. For example, modules will be evolved to control a specific behavior (sending control

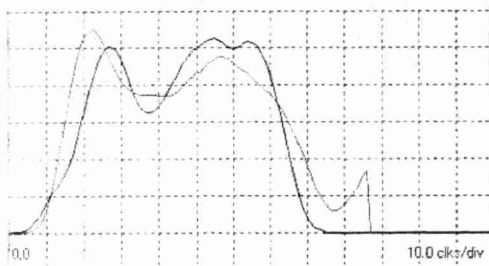


Fig. 6. Target (dark line) and response (grey line) of evolved module over 70 clocks

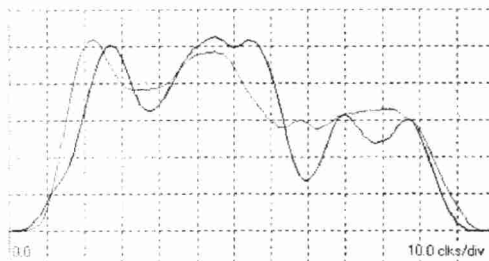


Fig. 7. Target (dark line) and response (grey line) of evolved module over 100 clocks

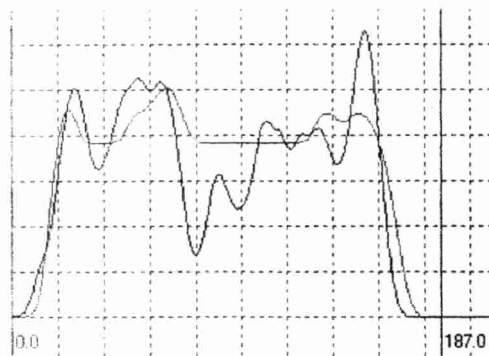


Fig. 8. Target (dark line) and response (grey line) of evolved module over 150 clocks

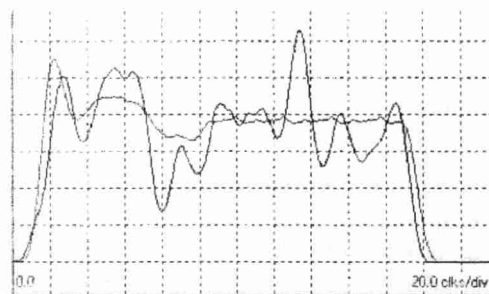


Fig. 9. Target (dark line) and response (grey line) of evolved module over 200 clocks

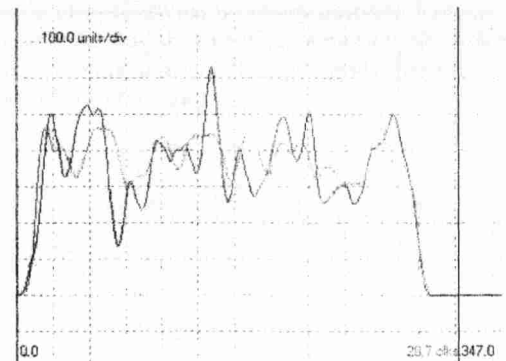


Fig. 10. Target (dark line) and response (grey line) of evolved module over 300 clocks

signals to the robot's motors to make arms and legs move in a particular way, etc). It is highly likely that these control signal wave forms will last far longer than the MEC of a single module, so how can a long duration target waveform be generated by a single module? One idea [7] is to cut up a long target wave form into "time slices", where the duration of each slice is less than a module's MEC and then to concatenate the outputs (appropriately delayed) of the separate modules (as suggested in [7]).

The delay modules can be either evolved or hand crafted cell by cell. There need be no prejudice against hand crafting a module if it is quicker and easier than evolving it, which is sometimes the case. In practice, brain building uses a mixture of the two approaches. As an example of a handcrafted multimodule concatenator, we attempt to concatenate four spiketrains, each of 48 clocks, into one long continuous spiketrain of 192 clocks. This is done as follows using 4 separate inputs :

1. $a_0, a_1, \dots, a_{47}, 0, 0, 0, \dots$
2. $b_0, b_1, \dots, b_{47}, 0, 0, 0, \dots$
3. $c_0, c_1, \dots, c_{47}, 0, 0, 0, \dots$
4. $d_0, d_1, \dots, d_{47}, 0, 0, 0, \dots$

and results in a single concatenated output:

$0, a_0, a_1, \dots, a_{47}, b_0, b_1, \dots, b_{47}, c_0, c_1, \dots, c_{47}, d_0, d_1, \dots, d_{47},$

To implement this multi-module system, 3 handcrafted modules are created. Each module has two input points. The 3 modules are interconnected as shown in Fig. 11. The 2 left hand side modules have each an upper input which is sent straight to the module's output point with minimum signal delay. The lower output is delayed by 48 clocks. A similar story holds

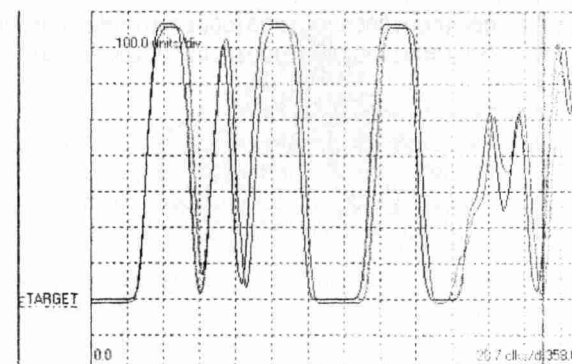


Fig. 13. The target and evolved convoluted output curves for the XOE experiment

5 Frequency Halver

One of the attractions of a CBM module is that it can accept static or dynamic input (in the sense of a waveform (before it is deconvoluted into a spike train [4,5]) *which does not change or does change its value in time*). Modules can be evolved which detect dynamic input patterns (in 1D or 2D). In this subsection we present a case of a 1D dynamic input signal, whose spike train frequency (a spike every N clocks) is halved at the output. Interestingly, the resulting evolved module generalizes, in the sense that if one inputs to this evolved module, a spike train with a spike frequency of 1 spike every M clocks, it will output (approximately) a spike train with a spike frequency of 1 spike every $2M$ clocks. We provide the parameter values used in the evolution and some results below.

The parameters we used were:

```
Population size: 100
Number of Generations: 100
Selection method: Rank
Number of Elite: 1
Convolution filter digitized values:
 1  2  3  5  8 16 26 35 44 52 59 62 63 62 61
57 52 45 37 29 21 13  7  4
Random seed: 123
Crossover rate: 0.1
Mutation rate: 0.01
Branch rate: 0.5
Neuron density: 0.2
Synaptic ratio: 0.5
```

Stimulus on input 60
(a constant frequency of 1 per 4 clocks):
00010001000100010001000100010001000100010001

[illegible]

```
(a constant frequency of 1 per 8 clocks):
0000000100000001000000010000000100000001
0000000100000001000000010000000100000001
0000000100000001000000010000000100000001
```

Figure 14 shows a very good evolution result. The lower line shows the response of the evolved module and the upper line shows the target. The two lines are depicted with a slight vertical offset for comparative purposes. Also, a slight horizontal offset is seen in Fig. 14. This is due to the delay of the binary signals flowing through the module. (The 1 bit signals move one CA cell per clock tick).

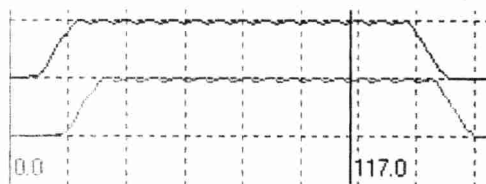


Fig. 14. Evolved and target signals of the best module after evolution of a frequency halver (lower line is the evolved signal, upper line is the target signal).

After this module was evolved we applied signals of various frequencies to its input and the response was indeed half of the input frequency. The following are the spiketrains of the brain run.

(several constant frequencies from 1 spike per 2 clocks to 1 spike per 8 clocks):

[illegible]

Response on output 4:

6 Moving Line Direction Detector

The experiment discussed in this section is an example of how the CBM can evolve a 2D dynamic pattern detector. An 8-by-8 input point array on one face of a CBM module cube is presented with a series of images representing a moving line. The images are binary coded with a '1' presenting a part of the line. The module was evolved to detect whether the line moves up or down across the array. If the line moves up, the module should generate a high output signal frequency (= high convoluted analog output value) and a low value if the line moves down. For the evolution the following parameters were used:

```
Population size: 200
Number of Generations: 600
Selection method: Rank
Number of Elite: 5
Convolution filter digitized values:
 1  2  3  5  8 16 26 35 44 52 59 62 63 62 61
57 52 45 37 29 21 13  7  4
Random seed: 123
Crossover rate: 0.03
Mutation rate: 0.01
Branch rate: 0.5
Neuron density: 0.2
Synaptic ratio: 0.5
```

[illegible]

```
[ 1 1 1 1 1 1 1 1 ]
[ 0 0 0 0 0 0 0 0 ]
[ 0 0 0 0 0 0 0 0 ]
[ 0 0 0 0 0 0 0 0 ]
[ 0 0 0 0 0 0 0 0 ]
[ 0 0 0 0 0 0 0 0 ]
[ 0 0 0 0 0 0 0 0 ]
[ 0 0 0 0 0 0 0 0 ]
```

```
[ bar moving down ] t = ...
[ several times   ]
[ and then up     ]
[ several times   ]
[ and this is     ]
[ repeated twice  ]
```

Target for moving down:

```
00000000000000001000000000000001
00000000000000001000000000000001
```

Target for moving up:

[illegible]

This stimulus was applied to the front face of the module (a space of 24*24*24 CA cells) described in the following array of input point number identifiers, as can be seen in figure 15.

```
[ 56 57 58 59 60 61 62 63 ]
[ 48 49 50 51 52 53 54 55 ]
[ 40 41 42 43 44 45 46 47 ]
[ 32 33 34 35 36 37 38 39 ]
[ 24 25 26 27 28 29 30 31 ]
[ 16 17 18 19 20 21 22 23 ]
[  8  9 10 11 12 13 14 15 ]
[  0  1  2  3  4  5  6  7 ]
```

The CBM evolved a module capable of distinguishing between a line moving up and a line moving down, although the target frequencies were not produced exactly (see figure 16). Nevertheless, the response frequencies are much higher when the bar moves up than when it moves down. So if one were to use a threshold function (or module) we could build a multi-module circuit that distinguishes movement (i.e. a dynamic 2-D input pattern detector). The following spiketrains are the outputs in the two separate test cases of a line moving up and a line moving down.

Response with bar moving down:

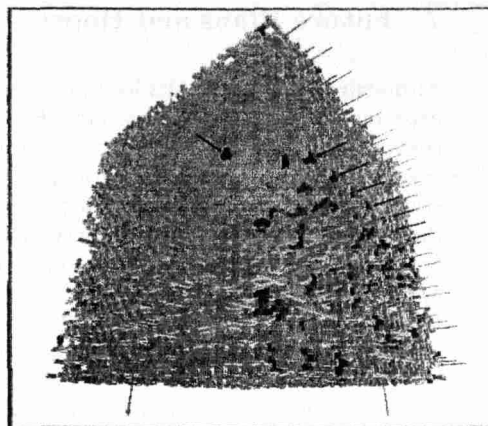
[illegible]

Fig. 15. Stimulus signal of one time step applied to the front face of the moving line direction detector

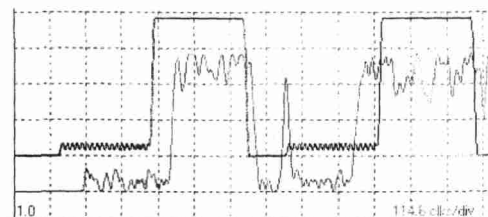


Fig. 16. Evolved and target signal of the best module after evolution of a moving line direction detector. (The lower line is the evolved signal. The upper line is the target signal).

```
1111110111101110100010001000000000000100001101
0001101000001101000111010111101111111111111
1111010011010001000010000000000010000000010000
00000000000000001
```

Response with bar moving up:

[illegible]

It can be seen clearly from these two outputs that the module “fires” more vigorously (more 1s) when the bar is moving up than when the bar is moving down. This is encouraging. Further work should show to what extent the CBM is able to evolve dynamic 2D pattern detectors (and in a few seconds).

7 Future Plans and Hopes

Although the initial results look promising, a lot more experimentation needs to be done. For example, the convolution technique we use in measuring the fitness value of an evolved curve may be inappropriate for generating higher levels of evolvability. Figs 4 to 7 show that the evolved curves are approximately accurate but not impressively so. Future work will be needed to choose better ways to evolve more accurate curves (with higher quality and longer MECs). Perhaps we may have to change the global fitness definition (used by the CBM) so that correctly evolved earlier spikes are given a heavier weighting, thus forcing the initial part of the evolved curve to follow the target curve more accurately, and then use step wise (incremental) evolution (where the evolved population in a GA run "N" becomes the starting population for a GA run "N+1" with a different fitness definition). Since the CBMs have only very recently come on line at the time of writing (Jan 2001), we still don't really know what they are capable of, so a thorough testing of their capacities and a study of their limitations are needed. In parallel with this, software development packages need to be built to facilitate the evolution, running and testing process of modules and artificial brains. The "automation" of brain building will be very important, because it will allow the full exploitation of the considerable capacities of the CBM.

The unrelenting pressure of Moore's law now forces us to think hard about the principles and architecture of the second generation brain building machine, that we call "BM2" (Brain Machine 2nd Generation). We are thinking of using a self-configuring electronics approach, that according to initial estimates, ought to provide a performance level of about 1000 times that of the CBM, by about the year 2004, and for a development cost of perhaps several million dollars. BM2 should be a lot more biological brain based, due to the collaboration of some senior people in the neuro science and neural network fields. In the meantime, we feel we have not yet scraped the surface of what the CBM has to offer.

With 64000 modules to evolve, we plan to give brain builder teams around the world remote access to the Starlab CBM(s), so that the machine(s) can be used while the Starlab team sleeps.

8 References

1. Hugo de Garis, "Evolvable Hardware : The Genetic Programming of Darwin Machines", Int. Conf. on Artificial Neural Nets and Genetic Algorithms, ICANNGA, 1993, Innsbruck, Austria.
2. Michael Korkin, Hugo de Garis, Felix Gers, Hitoshi Hemmi, CAM-Brain Machine (CBM) : A Hardware Tool which Evolves a Neural Net Module in a Fraction of a Second and Runs a Million Neuron Artificial Brain in Real Time, Genetic Programming Conf., July 1997.
3. Koza John R., Deb Kalyanmoy, Dorigo Marco, Fogel David B., Garzon Max, Iba Hitoshi, Riolo Rick L. (eds.), Stanford University, San Francisco, CA, USA.
4. Hugo de Garis, Felix Gers, Michael Korkin, Arvin Agah, Norberto Eiji Nawa, CAM-Brain, ATR's billion neuron artificial brain project : A three year progress report, Artificial Life and Robotics Journal, Vol.2, 1998, pp56-61
5. Hugo de Garis, Michael Korkin, Building an Artificial Brain Using an FPGA Based CAM-Brain Machine, Applied Mathematics and Computation Journal, Special Issue on Artificial Life and Robotics, Artificial Brain, Brain Computing and Brainware, North Holland, to appear 2000.
6. Hugo de Garis, Michael Korkin, The CAM-Brain Machine (CBM) : Real Time Evolution and Update of a 75 Million Neuron FPGA-Based Artificial Brain, Journal of VLSI Signal Processing Systems (JVSPS), Special Issue on Custom Computing Technology, to appear 2000.
7. Felix Gers, Hugo de Garis, Michael Korkin, CoDi-1Bit : A Simplified Cellular Automata Based Neuron Model, Proceedings of AE97, Artificial Evolution Conference, October 1997, Nimes, France.
8. Hugo de Garis, Andrzej Buller, Thierry Dob, Jean Honlet, Padma Guttikonda, Derek Decesare, "Building Multimodule Systems with Unlimited Evolvable Capacities from Modules with Limited Evolvable Capacities (MECs)", 2nd DoD/NASA Workshop on Evolvable Hardware EH2000, Silicon Valley, California, USA, July 2000.

Building Gods or our Potential Exterminators

Prof. Dr. Hugo de Garis
Head of "Starbrain"
Starlab's Artificial Brain Project
Brussels, Belgium

Abstract

Late 21st century computing technologies such as 1 bit per atom memory storage, heatless, reversible, nanoteched, self assembling, asteroid sized, femtosecond switching, quantum computers will allow godlike massively intelligent machines called "artilects" (artificial intellects) to be built with potentially trillions of trillions of trillions of times human intelligence levels, which may or may not decide one day to wipe out the human species for whataever reason. The toughest decision humanity will have to make in the 21st century is whether "to build gods or to build our potential exterminators". Prof. Dr. Hugo de Garis, the father of artificial brain technology, believes that a "gigadeath" war over the species dominance question will erupt in the second half of the 21st century between two human groups, the "Cosmists" who will see building artilects as a religion, the destiny of the human species, and the "Terrans" who will fear that the artilects one day might wipe us out. In the limit the Terrans will destroy the Cosmists to preserve the survival of the human species. The Cosmists will defend themselves. Hence the gigadeath artilect war. Prof. de Garis is currently writing a book on these ideas. Building Gods or our Potential Exterminators

Robot artificial intelligence is evolving a million times faster than human intelligence. This is a consequence of Moore's law which states that the electronic performance of chips is doubling every year or so, whereas it took a million years for our human brains to double their capacities. It is therefore likely that it is only a matter of time before our machines become smarter than we are. It is also likely that this development will occur this century if humanity chooses to allow it to happen.

My name is Prof. Hugo de Garis. My team and I are starting to design and build the world's first artificial brain at my lab Starlab in Brussels, Belgium,

Europe, which should contain nearly 100 million artificial brain cells (neurons). In about 4 years, the next generation artificial brain should contain a billion neurons. Our human brains contain roughly 100 billion neurons, so it is not surprising that someone like me is preoccupied with the prospect of robot intelligence surpassing the human intelligence level. (Admittedly, massive computational speed and size do not automatically equate to massive intelligence, but they are prerequisites. The potential is there. My brain building team still faces the considerable challenge of architecting the artificial brain. We will need to become "BAs" - Brain Architects. Despite this qualification, not only do I believe that artificial brains could become smarter than human beings, I believe that the potential intelligence of these massively intelligent machines, which I call "artilects", could be truly trillions of trillions of trillions of times greater. If these astronomically large numbers sound like science fiction to you, consider the following.

Moore's law is achieved by the size of electronic components such as transistors by a factor of two roughly every year. This halves the distance between components, and hence doubles the speed at which electronic signals can move between them (at the speed of light, a constant of nature). This trend has been valid for 30 years, and is likely to continue until 2020, by which time the scale of electronic circuitry will have reached atomic levels. In other words, within a single human generation it will very probably be possible to store a single bit of information on a single atom. There are a trillion trillion (a 1 with 24 zeros after it) atoms or molecules in objects of human scale, such as an orange. An object as large as an asteroid (to be found in the asteroid belt circling the sun between Mars and Jupiter) can be hundreds of kilometers across and contain a trillion trillion trillion atoms. The bits stored on these atoms could switch (bit flip) from a 0 to a 1 and vice versa in a femto-second (a thousandth of a trillionth of a second). That's an information processing capacity of about ten million trillion trillion trillion trillion (a 1 with 55 zeros) bit flips a second. When one compares the comparable information handling capacity (in bit flips per second equivalent) of the human brain, the estimated answer is about ten thousand trillion bit flips a second (a 1 with 16 zeros), which is a thousand trillion trillion trillion times smaller. These artilects could potentially be truly god like, immortal, have virtually unlimited memory capacities, and vast humanly incomprehensible intelligence levels.

I foresee humanity splitting into two major ideological, bitterly opposed groups over the "species dominance" issue, i.e. should humanity build artilects or not. These two groups I label the "Cosmists" (in favor of building

them) and the "Terrans" (who are opposed). To the Cosmists (based on the word 'cosmos'), building artefacts will be a religion (compatible with and based upon modern science), as the destiny of the human species, as the magnificent goal of creating the next rung up the ladder of dominant species. To the Terrans (based on the word 'terra', the earth), building such artefacts means accepting the risk that one day, in an advanced state, these artefact gods might decide, for whatever reason, that the human species is so inferior and such a pest, that they should exterminate us. With their gargantuan intellects, such a task would not be difficult for them. The Terrans, in the limit, will try to exterminate the Cosmists if the latter insist on building artefacts, for the sake of preserving the survival of the human species. Since the stake is so high (namely whether the human species survives or not) the passion levels will be high. The Cosmists will anticipate the murderous hatred of the Terrans and will defend themselves. We have thus all the makings of a major war. About 200 million people died for political reasons in the 20th century (wars, purges, genocides, etc) using 20th century weapons. Extrapolating up the graph until late 21st century, with 21st century weapons, we arrive at billions of dead - "gigadeath".

So which am I, a Cosmist or a Terran? I'm both. Ultimately, I think it would be a cosmic tragedy if humanity chooses to freeze evolution at the puny human level (with our pathetic little lives of 80 years in a universe billions of year old, that contains a trillion trillion stars - the "big picture"). For me, the tragedy of seeing the human species wiped out is less significant than not seeing the birth of the artefacts. This sounds monstrous, and it is, in human terms, but to deny the creation of the first true artefact, which would be "worth" a trillion trillion trillion human beings, would be a far greater tragedy, a "cosmic" tragedy.

As the planet's pioneering brain builder, I feel a terrible burden of responsibility towards the survival of the human species and the creation of godlike artefacts, because I am part of the problem. I am quite schizophrenic on this point. I would love to be remembered after I'm gone as the "father of the artificial brain", but I certainly don't want to be seen in future historical terms as the "father of gigadeath". Hence I try to raise the alarm now while there is still time before the artefacts come into being. If I were a true Cosmist, I would keep quiet and just get on with my work, but instead I feel that humanity should be given the chance to nip the rise of the 21st century artefact in the bud if it so chooses. So should work on artificial brains be stopped now? I think not. For the next 30 years or so, brain based computers will be far too useful to be suppressed. For example, they will become

smart enough to clean the house, teach the children, provide sex, and help human experts in decision making, etc. They will do most of the work and thus create great wealth for the whole planet. So, in the short to middle term, brain building technology will be seen as a great boon to humanity. It is the longer term that terrifies me and keeps me awake at night. I see no way out of a gigadeath artilect war, so relentless is the logic. The rise of the artilect will probably be inevitable. The economic and military pressures to build them will be enormous - hundreds of trillions of dollars a year world-wide will be spent in the brain based computing market within 10-20 years, I believe. The debate over whether artilects should be built or not is already starting to heat up, at least amongst the researchers concerned with brain building and AI (artificial intelligence). This debate is starting to spill over to other specialties, for example, I'm trying to persuade Prof. Peter Singer (Princeton University), the planet's best known "applied ethics" professor, to write a book about "Artilect Ethics". At the rate at which this issue is hitting the headlines lately, my bet is that within a few years the "artilect debate" will be on everyone's mind.

The decision to build artilects or not will be the toughest decision that humanity will ever have to make. Personally, I'm glad to be alive now. As I said in a recent European Discovery Channel documentary on my work and ideas, "I fear for my grandchildren. They will see the horror, and they will be destroyed by it".

Section II:

ROBO PLAZA

Groningen RoboCup Team
Dutch RoboCup Team
GMD RoboCup Team
Bolesian
IKAT Maastricht
EPFL Lausanne

Robotics is profoundly an experimental science. New Artificial Intelligence is driven by the quest for complex robotic systems. When discussing modern robotics, one has to take all practical aspects into account as well. Therefore, Robo Sapiens provides a Robo Plaza on which several kinds of robots, with different kinds of tasks, on the basis of different kinds of methods and techniques are presented. This should draw attention to the many problems faced by roboticists, and the intimate relation between theory and practice.

The RoboCup foundation aims to construct a team of humanoid robots to beat the leading soccer world champions before the year 2050. A rapidly growing number of universities and research institutes participate in the local and global RoboCup-soccer tournaments. An even greater number of universities is trying to gain interest and experience to enter the competition at a point in the future. At Robo Plaza, three RoboCup teams will present some of their robots.

Groningen has just started their robotic efforts. They have an educational project with the ultimate goal to become the second Dutch RoboCup team. The First Dutch RoboCup Team, Clockwork Orange, is a joint project between three universities. This team has entered in the European Championships in Amsterdam and will enter in the German Open and the world championships in Seattle later this year.

The GMD RoboCup Team is the most experienced of the teams present on Robo Plaza. They have entered in the competition for some year now and are organising the German Open. They will demonstrate their robots and the behavior-based control architecture they developed.

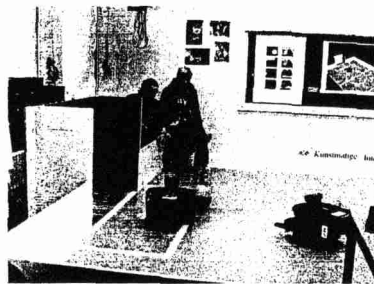
Ofcourse, robots can do much more then playing soccer. Bolesian uses robots to visualize the functionality of knowledge-base information processing. Paul Vogt, from IKAT, uses robots to solve the symbol grounding problem. He lets the robots build their own lexicon through the interaction with their environment and with each other. Dario Floreano and his students use robots to test various kinds of artificial evolution. They demonstrate the new generation of evolutionary robotics, evolving adaptive controller instead of the controllers themselves.

The Groningen RoboCup Team

Erwin Mulder
Artificial Intelligence
Technical Cognitive Science
Grote Kruisstraat 2/1
9712 TS Groningen

March 2001

At the department of Artificial Intelligence in Groningen we recently started an educational robotics project with the aim to ultimately create a second Dutch Robocup team. In this project we take a mainly 'behaviour based' approach and try to implement subtasks that can be identified in a game of Robocup football as (a combination of) robust basic behaviours.



We use 4 Pioneer 2 robots. They are all equipped with 14 sonar sensors, an on-board pan-tilt-zoom camera and a gripping device. The picture shows an attacking robot and a defending one.

In our demonstration we will show several things we have managed to incorporate so far. This includes a field player that manages to visually find the ball and push it into the opponent's goal, and a Keeper that tries to track the ball with a camera and manoeuvre itself between the ball and its own goal.

Clockwork Orange: The Dutch RoboSoccer Team

Marco Wiering
Institute of Information and Computing Sciences
Utrecht University
The Netherlands

March 20, 2001

The Dutch RoboSoccer Team, called "Clockwork Orange" is a Dutch research project in which 3 universities participate: the University of Amsterdam, the Technical University of Delft, and University Utrecht. The team consist of 2 types of robots: three Nomad Scout robots and the Pioneer 2 robot (Dexter). Figure 1 shows our robots.

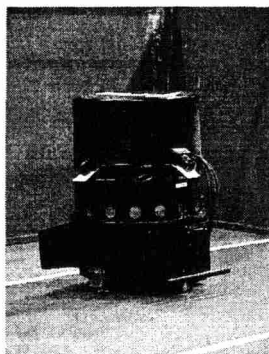


Figure 1: Our robots: Left the Nomad Scout Goalie. Right Dexter, the Pioneer 2 robot.

Our research focuses on different research aspects of RoboCup: vision, sensor data fusion, self-localisation, communication between robots, multi-agent cooperation, and behavior-based reinforcement learning.

The team has already participated in the Mid-sized league of the European Championship RoboSoccer, and was kicked out of the competition in the quarter finals against the final champion team Shariff of Iran. This year we have more than 20 students working on the project, and intend to participate at the World Championships held in Seattle in August, 2001.

Demonstration. At the Robo Sapiens Symposium we will demonstrate the keeping abilities of the Dutch goalie and the player skills of two robotic soccer players. The players will play one-to-one against each other, both trying to get the ball and score against the same defending goalie.

GMD robots Powered by Dual Dynamics

Hans-Ulrich Kobialka Peter Schll
Behaviour Engineering

AiS (Institute for Autonomous intelligent Systems)
GMD - German National Research Center for Information Technology
Schloss Birlinghoven
Germany

4 April, 2001

To build soccer playing robots, an architecture must be designed which integrates hardware and software and balances between system flexibility and performance. The basic elements are: a fast and reliable vision component, an effective and strong motor system and some software strategies for the robots to increase the soccer playing capabilities. For the robots, GMD uses a modular design of hardware and software which can be considered as a robot construction system. This approach facilitates the implementation of controlled experiments with different designs. The hardware of a robot consists of an aluminium chassis with differential drive. It uses low level sensors (odometry, distance sensors etc.) and a NewtonLab vision system.

The Dual Dynamics (DD) scheme is a mathematical model of a behavior control system for autonomous mobile robots. It has grown from three roots: the behavior-based approach to robotics, the dynamical systems approach to cognition, and the mathematical theory of self-organizing dynamical systems.

Dual Dynamics is a methodology for designing the behavior of autonomous robots. The demonstration will show how the robot is controlled by different behavior modules according to the current situation.

Kennistechnologie in Uitvoering

Bolesian B.V.
Europalaan 24b
5203 DH s Hertogenbosch
Tel: +31 (0) 73 648 3311

4 April, 2001

De voornaamste doelstelling waarom deze robot is ontwikkeld, is het inzichtelijk maken van kennistechnologie. Daarom is er een expliciete scheiding aangebracht tussen de robotkennis: hoe ziet mijn wereld er uit en welke beslissingen moet ik nemen, en robotprogrammatuur: het daadwerkelijk aansturen van bijvoorbeeld zijn wielen. Door middel van een demonstratie wordt aangetoond dat op grond van veranderende kennis, de robot ander gedrag zal vertonen.

De programmatuur van de robot is in staat een aantal primitieven (zoals voorwaarts bewegen en draaien) aan te sturen, de input van zijn sensoren naar de kennisbank te sturen en commandos te ontvangen van deze kennisbank. Op grond van hetgeen de robot waarneemt met zijn sensoren, neemt de kennisbank beslissingen, die de robot weer kan uitvoeren. Daarnaast wordt deze informatie verwerkt in de kennisbank, zodat er geleerd wordt van de vergaarde informatie.

Voor het ontwikkelen van deze robot is gebruik gemaakt van technieken en methoden zoals UML en SDF II, zoals die standaard door Bolesian in hun trajecten wordt gebruikt. De fysieke robot is gebouwd met LEGO en de kennisbank is gecomplementeerd met AION v.8. De communicatie tussen de robot en de kennisbank gebeurt door middel van infrarood.

Adaptive Grounding of Lexicons on Mobile Robots

Paul Vogt

IKAT Universiteit Maastricht

paul@cs.unimaas.nl

Abstract

This paper reports on experiments done on mobile robots. In the experiments the robots develop a lexicon of which the meaning is grounded in the real world. The experiments are based on an approach that is in line with the embodied cognition paradigm. Robots construct a lexicon by engaging in a series of language games. In a language game, two robots try to name an object they can perceive in their environment. When they fail to communicate successfully, they adapt their memories in order to improve performance in future games. This way a shared lexicon is formed. The experiments showed that the robots can do this fairly well within a simple experimental setup.

1 Introduction

Symbol grounding is one of the hardest problems in robotics research. How can robots develop and use symbols that have their meaning grounded in the real world? Or in other words: how do seemingly meaningless symbols acquire their meaning in relation to the real world? This is the main question of what is known as the *symbol grounding problem* [Harnad, 1990].

Several attempts have been made in the past to solve this problem. Some of these investigated the problem in relation to language acquisition, e.g. [Billard and Hayes, 1997; Sugita and Tani, 2000; Yanco and Stein, 1993]. In these experiments, however, (parts) of the language has been predefined. In the work of Billard and Hayes, for instance, a teacher robot has been preprogrammed with the language and a student robot learns it by imitation. Jun Tani provides the language by human-robot interaction. And robots in Yanco and Stein's work have also been preprogrammed with the signals that exist in their language and meaning. In their work, the robots learn to associate the proper signals with the proper meanings by means of reinforcement learning where feedback about the robots' performance is provided by a human instructor. So, all these investigations require some input from a human.

In the past few years, research at the AI Lab of the Vrije Universiteit Brussel has been focussed on the origins of lexicons on physical robots, see e.g. [Belpaeme et al., 1998; Steels, 1997; Steels and Vogt, 1997]. In these experiments real robots have been developed that interact with their environment, including each other. The interactions are modeled by so-called language games [Steels, 1996a]. In a language game, two robots try to communicate the name of some object they observed in their environment. They first sense what is in their environment. The sensing is then preprocessed and categorized. Then the speaker of the game (one of the two robots) tries to name the categorization of one of the objects. In turn, the hearer (the other robot) tries to interpret this name in relation to one of the categorized objects. Afterwards, feedback is provided whether or not the two robots communicated the same object. Depending on the outcome, the lexicon is adapted to improve their future performance.

The development of the lexicon is based on three mechanisms that define an *adaptive complex dynamical system*. These mechanisms are: (cultural) interaction, individual adaptation and self-organization [Steels, 1996a]. As a result of these mechanisms a shared lexicon emerges in a similar way as ant-paths are formed. The lexicon is an attractor of the adaptive complex dynamical system.

This paper is organized as follows: A brief discussion of the symbol grounding problem, its relation to embodied cognition and some adopted definitions are presented in the following section. Section 3 very briefly introduces the language game model. Some experimental results are provided in section 4. And finally, section 5 concludes.

2 The Symbol Grounding Problem

As mentioned, the symbol grounding problem [Harnad, 1990] questions how symbols acquire their meaning in relation to the real world. In Harnad's work, symbols are defined in the classical way. I.e. they are defined in terms of *physical symbol systems* as proposed by [Newell, 1980]. In Newell's definition, symbols are patterns that provide access to some distal structure. They are completely represented inside an agent's brain and have in some

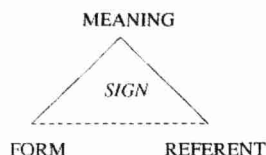


Figure 1: The semiotic triangle illustrates the relations between referent, form and meaning that constitute a symbol.

way a relation to the real world. The latter notion is subject of the symbol grounding problem.

Also around 1990 Rodney Brooks, proposed his *physical grounding hypothesis* [Brooks, 1990]. This hypothesis states that agents should have their intelligence grounded in the real world, but that they do not need symbolic representations. The intelligence could be represented by sensorimotor couplings and part of it is *situated* in reality. Hence it is an agent's interaction with reality from which intelligence emerges. Brooks' hypothesis is the foundation on which the embodied cognition paradigm is built.

So, one may argue that agents need not solve the symbol grounding problem as there are no symbols, see e.g. [Pfeifer and Scheier, 1999]. However, humans seem to use symbols. But if one comes up with an alternative definition of a symbol, symbols might still fit within the embodied cognition paradigm [Clancey, 1997]. Such an alternative is provided by the theory of semiotics. In semiotics the notion of a *sign* is central. A sign is defined by a relation between a *form*, a *meaning* and a *referent* [Chandler, 1994]:

Referent A referent is the thing "to which the sign refers".

Form A form is "the form which the sign takes (not necessarily material)".

Meaning The meaning is "the sense made of the sign".

If the form in relation to the meaning is either arbitrary or conventionalized so that the relationship must be learned, the sign is called a symbol [Peirce, 1931]. The relation between the three elements are usually illustrated by a semiotic triangle, see figure 1.

An advantage of this definition of the symbol is that it is *per definition* grounded, since the referent (a real world object) is an intrinsic part of the symbol. Yet the problem remains that the semiotic triangle still has to be constructed. This is, like the symbol grounding problem a very hard problem. This is so, because when a robot detects a referent under different condition (e.g. from different positions), its sensory stimulation differs extremely. To solve the symbol grounding problem, the robot must identify these different sensings invariantly.

The fact that the semiotic definition provides a structural coupling between referent, meaning and form makes the symbol fit within the embodied cognition paradigm. It is situated and embodied. The latter

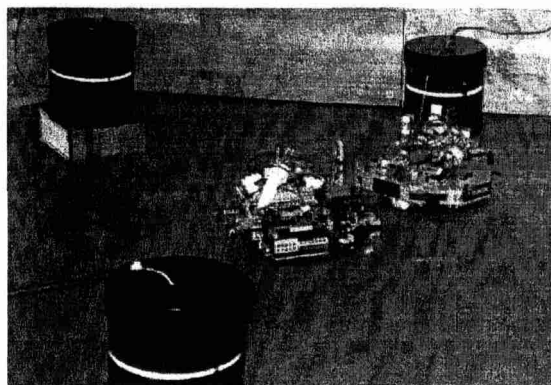


Figure 2: The robots situated in their environment as used in the experiments.

means that it depends on the an agents physical interaction with the real world. The language games that have been implemented on the robots implement the interaction and the construction of the semiotic symbols.

3 The Language Game Model

The experiments are done using two LEGO robots that are situated in a small environment consisting of four light sources, see figure 2. Each light source is placed at a different height and the robots are equipped with four light sensors, each mounted at a corresponding height. The aim of the experiment is that the robots develop a shared lexicon of which the meaning is grounded in their world. The lexicon is a set of form-meaning associations that relate, ideally, to some light source. Prior to the experiments, the robots have no categories and their lexicon is empty.

The lexicon development is guided by a series of *guessing games*. A guessing game is a variant of a language game in which the hearer tries to guess what light source the speaker is naming. Below follows a very brief step-by-step description of the guessing game. The interested reader is referred to [Vogt, 2000] for the details.

1. The guessing game starts when both robots are close to each other with their backs oriented towards each other.
2. One by one, the robots rotate 720° in order to detect their environment. The sensing results in a spatial view of the middle 360° . The sensing is described by a set of raw sensory data which is sent to a PC for further processing.
3. The sensing is preprocessed by segmentation and feature extraction processes. This results in a description of the sensing, now called a *context* of feature vectors. A feature vector is a 4 dimensional vector describing some properties of the sensing. Each feature vector is supposed to relate to the detection

of one light source. It is important to note that not always all four light sources are detected.

4. After the context is constructed, the speaker chooses one arbitrary feature vector as the topic. The hearer considers all feature vectors as a potential topic; it has to guess which feature vector is the real topic.
5. Both robots individually try to find a distinctive categorization for the (potential) topic(s). This is modeled by so-called *discrimination games* [Steels, 1996b]. The aim of a discrimination game is to find a categorization of a topic that distinguishes the topic from all other feature vectors in the context. A category is represented by a prototype (i.e. a point) in the feature space that is spanned from all possible feature vectors. When no distinctive category can be found by the discrimination game, new categories can be created for which the feature vector of the topic is used as an exemplar. If the discrimination game succeeds, the distinctive categories are forwarded to the naming part of the guessing game. If the guessing game succeeds, the prototypical category is moved towards the feature vector of the topic. This way the category becomes a representative example of the feature vectors it categorizes.
6. After both robots thus acquired distinctive categories that relate to the (potential) topic(s), the speaker tries to produce an utterance that names the distinctive category of its topic. It does so by searching its lexicon for elements of which the meaning matches the distinctive category. If there are more than one, it selects the one that has the highest association score and the associated form is uttered. The association score indicates the element's past effectiveness in the communication. If the speaker does not find such an element, it creates a new form, associates this with the distinctive category (which now becomes a meaning) and adds the new element to its private lexicon.
7. When the hearer receives an utterance, it tries to interpret it. It searches its own lexicon for elements of which the form match the utterance, and of which the meaning matches a distinctive category of one of the potential topics. If there are more than one such elements, the hearer selects the one that has the highest association score. The feature vector to which the matching distinctive category relates is then chosen as the hearer's topic of the game. I.e. this feature vector is what the hearer guessed the speaker has named.
8. Feedback is provided on whether the hearer found a lexical element and if so, whether both robots communicated the same topic.
9. Depending on the outcome (provided by the feedback) the lexicon is adapted. Three possible outcomes / adaptations remain:
 - (a) The hearer has not found an element in its lexicon that matches the received utterance and of

which the meaning is consistent in the game's context. In this case, the hearer adopts the uttered form and associates it with the distinctive categorization(s) of one arbitrarily selected feature vector. The speaker decreases the association score of the used form-meaning association. The guessing game fails.

- (b) The hearer has found a matching element, but the selected topic is not consistent with the speaker's topic. In this case, the hearer again adopts the uttered form and associates it with the distinctive categorization of an arbitrarily selected feature vector. Both robots decrease the association score of the used form-meaning association. The game fails.
- (c) Both robots have selected a lexical element in relation to a consistent topic. The hearer thus guessed right and the guessing game succeeds. Both robots increase association score of the used element and competing elements are *laterally inhibited*. An element is competing when the form matches the communicated form, but its meaning is inconsistent. It is also competing when the meaning matches the meaning of the used element, but its form is inconsistent.

The guessing game thus models a communication act. The mechanisms on which the game is based guide the formation of a shared global lexicon. The lexicon is constructed locally inside each robot and is spread through the population through cultural interactions. Score adaptations and selection drive the self-organization of the global lexicon.

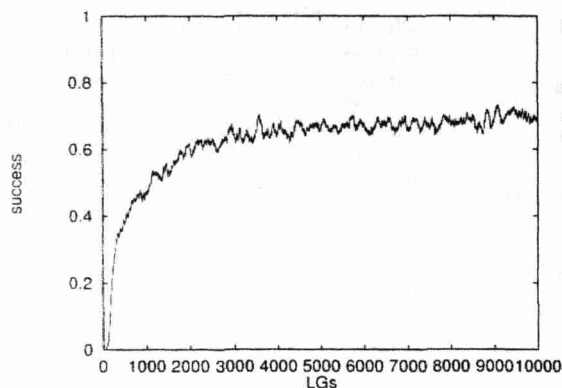
4 Experimental Results

Given the above description of the guessing games a large series of experiments have been done. The results of these experiments are reported in [Vogt, 2000]. In this section results of one such experiment is presented.

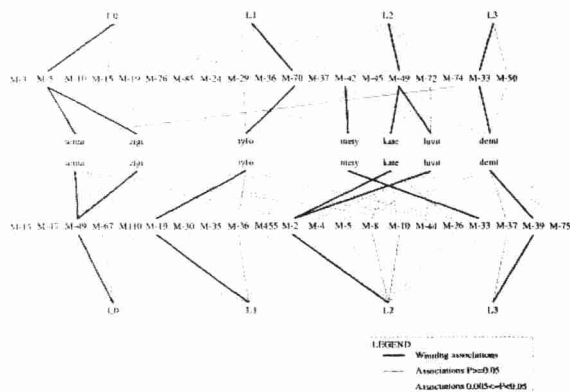
Prior to the experiments a data set of raw sensory data from the sensing of 1,000 situations is recorded. One situation consists of the sensing as if both robots play a guessing game. With this data set, a series of 10 runs have been done each experiment. One run consists of 10,000 guessing games. Each game one of both robots is arbitrarily selected as the speaker, which arbitrarily selects one feature vector as the topic.

From the sensory data set, some expectations can be extracted. For instance the a priori chance that the robots choose the same topic when each selects a topic arbitrarily. This a priori chance has been calculated to be 23 %. Furthermore, since the robots not always detect all light sources, they may not have acquired a coherent context. This yield a maximum in their potential understandability. I.e. there is a maximum in the expected communicative success. This potential understandability has been calculated to be 80 %.

Figure 3 (a) shows the evolution of the communicative success. The communicative success is the success of the



(a) CS



(b) Lexicon

Figure 3: The results of the experiment. Figure (a) shows the communicative success and figure (b) shows an instance of an emerged lexicon.

guessing games averaged over the past 100 games and averaged over the 10 runs. As can be seen, the success increases towards a value around 75 %, which is much larger than the a priori chance for success and slightly lower than the potential understandability. Hence the robots seem to be pretty well capable of developing a shared lexicon.

An instance of such a lexicon is shown in figure 3 (b). This figure shows a semiotic landscape that displays the connections between referent (or light source) L, the meanings M and forms like **tyfo**. This is for both robots r0 (upper half) and r1 (lower half). The connections are weighted by their relative co-occurrence frequencies during one entire run. The connections between referent and meaning are relative to the occurrence of the referent and the connections between form and meaning are relative to the occurrence of the form.

Ideally, the graph is orthogonal with respect to the referents. This means that, ideally, the connections between referent, meaning and form for both robots should not mix. A referent may be related with several meanings and several forms, as long as there is no relation to another referent. Furthermore, each referent is ideally associated with only one form. Figure 3 (b) shows that in most cases the graph almost meets the orthogonality criterion. Connections that mix up are connections that have low co-occurrence frequencies. The figure also shows that, although there are some synonymous relations between referent and form, these are far less than the one-to-many relations between referent and meaning. Moreover, the polysemous relations between form and referent are even lower. The latter means that when a robot uses some form, it almost uniquely determines the relation to some referent. The former means that when a robot tries to name some referent, it only uses a few forms to do so.

5 Conclusions

This paper showed an experiment in which two mobile robots solved the symbol grounding problem. The robots do so by engaging in a series of guessing games in which they try to communicate a referent. While they do so, they construct a lexicon which is shared through their communication and of which the meaning is grounded in the real world. The lexicon consists of arbitrary, though conventionalized forms that are related with one or more meanings. The meanings are categorizations that the robots constructed from their sensing of the real world. This is what constitutes a situated (or semiotic) symbol, i.e. the relation between referent, meaning and form.

Given local mechanisms like categorization, naming and adaptation, the robots develop a shared lexicon. This is an emergent phenomenon. The goal of constructing a shared lexicon has not been preprogrammed, neither is the structure of it. It emerges through the self-organizing effect implemented by the local mechanisms.

As the semiotic landscape showed (figure 3 (b)), the lexicon is formed of symbols that have one-to-many relations between referent and meaning and one-to-many relations between form and meaning. The result is that the one-to-many relations between referent and meaning are canceled out by the one-to-many relations between form and meaning, yielding one-to-one or one-to-few relations between referent and form. So, the symbols are identified more or less invariantly only at the naming level. This provides a strong argument in favor of a co-evolution between language and meaning, rather than two separate evolutions as is for instance hypothesized by [Deacon, 1997].

Acknowledgments

This work has been done at the Vrije Universiteit Brussel leading to a Ph. D. thesis under the supervision of Luc Steels. The author wishes to thank Luc Steels for providing an excellent research environment and many

fruitful discussions. In addition all colleagues from the VUB AI Lab are thanked for their contributions in one way or the other.

References

- Belpaeme, T., Steels, L., and van Looveren, J. (1998). The construction and acquisition of visual categories. In Birk, A. and Demiris, J., editors, *Learning Robots, Proceedings of the EWLR-6, Lecture Notes on Artificial Intelligence 1545*. Springer.
- Billard, A. and Hayes, G. (1997). Robot's first steps, robot's first words ... In Sorace and Heycock, editors, *Proceedings of the GALA '97 Conference on Language Acquisition - Edinburgh*, Human Communication Research Centre. University of Edinburgh.
- Brooks, R. (1990). Elephants don't play chess. *Robotics and Autonomous Systems*, 6:3-15.
- Chandler, D. (1994). Semiotics for beginners. [WWW document] [URL http://www.aber.ac.uk/media/Documents/S4B/semiotic.html](http://www.aber.ac.uk/media/Documents/S4B/semiotic.html).
- Clancey, W. J. (1997). *Situated Cognition*. Cambridge Univsity Press.
- Deacon, T. (1997). *The Symbolic Species*. W. Norton and Co., New York, NY.
- Harnad, S. (1990). The symbol grounding problem. *Physica D*, 42:335-346.
- Newell, A. (1980). Physical symbol systems. *Cognitive Science*, 4:135-183.
- Peirce, C. (1931). *Collected Papers*, volume I-VIII. Harvard University Press, Cambridge Ma.
- Pfeifer, R. and Scheier, C. (1999). *Understanding Intelligence*. MIT Press.
- Steels, L. (1996a). Emergent adaptive lexicons. In Maes, P., editor, *From Animals to Animats 4: Proceedings of the Fourth International Conference On Simulating Adaptive Behavior*, Cambridge Ma. The MIT Press.
- Steels, L. (1996b). Perceptually grounded meaning creation. In Tokoro, M., editor, *Proceedings of the International Conference on Multi-Agent Systems*, Menlo Park Ca. AAAI Press.
- Steels, L. (1997). The synthetic modeling of language origins. *Evolution of Communication*, 1(1):1-34.
- Steels, L. and Vogt, P. (1997). Grounding adaptive language games in robotic agents. In Husbands, C. and Harvey, I., editors, *Proceedings of the Fourth European Conference on Artificial Life*, Cambridge Ma. and London. MIT Press.
- Sugita, Y. and Tani, J. (2000). A connectionist model which unifies the behavioral and the linguistic processes: Results from robot learning experiments. Technical Report SCSL-TR-00-001, Sony CSL.
- Vogt, P. (2000). *Lexicon Grounding on Mobile Robots*. PhD thesis, Vrije Universiteit Brussel.
- Yanco, H. and Stein, L. (1993). An adaptive communication protocol for cooperating mobile robots. In Meyer, J.-A., Roitblat, H., and Wilson, S., editors, *From Animals to Animats 2. Proceedings of the Second International Conference on Simulation of Adaptive Behavior*, pages 478-485, Cambridge Ma. The MIT Press.

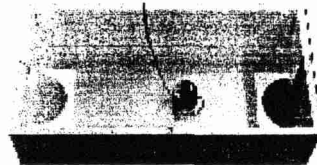
The "Light-Switch" Problem: Evolution of Adaptive Synapses

Joseba Urzelai Dario Floreano Jesper Blynel

Evolutionary and Adaptive Systems Team (EAST)
Laboratory of Microprocessing and Interfaces (LAMI)
Swiss Federal Institute of Technology (EPFL)
CH-1015 Lausanne, Switzerland

Demonstrators:

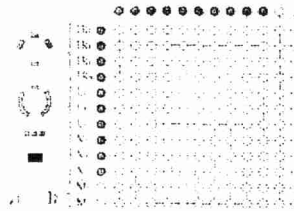
Jesper Blynel, Claudio Matussi, Jean-Christophe Zufferey



A mobile robot Khepera equipped with proximity sensors, light sensors and a vision module is placed in a rectangular arena. A light bulb is attached to a wall in one end of the arena. The light is initially off, but can be switched on by the robot when it passes over a black-painted area in the opposite end of the arena. A black stripe is painted on the wall over this "light-switch" area.



An evolutionary algorithm is used to evolve robot controllers to solve the problem. Each controller is a fully connected recurrent neural network. The genotype encodes the adaptive properties of the synapses in the neural network and not the synaptic weights. Each individual starts its life with randomly initialized synaptic weights. Each synapse is then set free to adapt using one of four variations of the standard hebbian learning rule. Each individual is tested in the arena and given a fitness-value proportional to the time it spends in the gray area under the light bulb when the light is on.



The demo shows the behavior of one of the individuals of the final generation of an evolutionary run. The visitors can press a red button to start the demo and then watch the behavior of the robot in the arena and follow the activation of the neural network on a monitor.

