Towards a Scientific Foundation for Engineering Cognitive Systems - an interim report -

Hans-Georg Stork
(h-gATcikon.de, AT = @)

Contents

1 Engineering, science, and cognition 2
2 Cognitive Systems 5
3 Artificial cognitive systems - whether, why and what 12
4 A scientific foundation - what it should support and how . . 23
5 A scientific foundation - bricks and mortar 27
5.1 Analysis 28
5.2 Modelling 29
5.3 Synthesis 36
6 A scientific foundation - building sites and builders 42
6.1 Platforms and environments 45
6.1.1 Cognitive machines 45
6.1.2 Artificial cognition in natural, artificial and hybrid worlds . 54
6.2 Core competencies 62
6.2.1 Maintaining the artificial system's body in its environment 63
6.2.2 Making the artificial system's mind explicit 69
6.3 Closing the gaps 80
6.3.1 Adaptation and learning in artificial systems 83
6.3.2 Beyond discrete patterns and computation 90
7 Postscript 102
Acknowledgements 102
References 102

Abstract (1) What is a cognitive system? (2) Why do we want to engineer such systems? (3) Why do we need a scientific foundation for engineering cognitive systems and what would be required of it? (4) What are its building blocks and what is the glue that binds them? (5) How can we achieve it? We propose answers to each of these questions. Answers to (5) also refer to projects currently being funded under the Cognitive Systems "Strategic Objective" of the European IST programme.[1]

1 Engineering, science, and cognition

> *"Ich behaupte aber, daß in jeder besonderen Naturlehre nur so viel eigentliche Wissenschaft angetroffen werden könne, als darin Mathematik anzutreffen ist."*
> **Immanuel Kant (1724-1804), in: Metaphysische Anfangsgründe der Naturwissenschaft (Vorrede), Riga, 1786**[2]

Engineering requires design. Design should be grounded in explicit knowledge. Explicit knowledge is gained through targeted research. Principled and methodical research is the hallmark of science.

For a long time in human history engineering has been guided by intuitive knowledge, quite separate from contemporaneous scientific endeavours - if there were any. This has gradually changed since the dawning of the modern era some 400-500 years ago. Many if not most of today's engineering products and the industries thriving on making them, would not exist without the fundamental insights gained through scientific research, for instance into the nature of matter at its smallest scales. The designs underlying our energy supply, our transport and communication devices and networks, the tools used by surgeons and other medical practitioners, many of the tools, devices, machines and appliances available to everyman: they are all based on a solid foundation in natural science. Physics in particular, has always been a key contributor of building blocks. But none would have fitted without the binding glue of mathematics, providing the intellectual underpinning and the formalisms needed to find and express models, theories, methods and solutions in an actionable way.

Many of today's most outstanding engineering feats are in the domains of electronic computing and its applications - for instance in the above mentioned areas. Harnessing these achievements keeps calling for further advances. We are, however, coming close to certain limits in these domains: not those yet that will eventually slow down progress according to Moore's law but limits to our own capacity of mastering the complexity of our own inventions.

Indeed, electronic computing as practised since the middle of the 20th century may no longer be a panacea in dealing with problems of managing networks of all sorts, industrial production and distribution systems, or simply using the

[1] December 2007; the views expressed in this note are those of the author and do not engage his employer.

[2] "But I maintain that in every special natural doctrine only so much science proper is to be met with as mathematics." Immanuel Kant (1724-1804), in: The Metaphysical Foundations of Natural Science (preface), Riga, 1786

A system is ...	
robust	if in the presence of previously unspecified perturbations (that may be caused externally or internally) it keeps functioning as specified; or, in potentially quantifiable terms: if small perturbations lead to small changes in performance;
versatile	if it can serve several different purposes and/or allows for different approaches to carrying out a given task;
adaptive	if it can change its structure and/or operation in response to changes in its environment while performing the same basic functions;
autonomous	if it is independent of external control;

A system behaves **naturally** if (1) its modes and modalities of communication and interaction with people (where necessary) comply with established interpersonal communication and interaction *modes* and *modalities*, and (2) communication and interaction contents are consistent with human interpretations of "the world".

Table 1: A brief glossary of general systems features

high-tech machines and devices we live with. Requirements such as flexibility, robustness, adaptability, reliability or interactivity are certainly not new. But they are becoming increasingly important and at the same time increasingly difficult to satisfy within more and more demanding deployment scenarios in largely non-deterministic and loosely structured natural or artificial environments. A new engineering approach is needed towards creating artificial systems that can meet these requirements more readily and with less human intervention than is customary today.

This note is about the science and - if only implicitly - the mathematics that ought to inform this new approach towards building systems and artefacts that are more robust, versatile, reliable, adaptable, convivial[3] and less demanding of human intervention than is possible given the current state-of-the-art. Some of these terms are explained in Table 1. Section 3 will be more explicit about the kind of artificial systems we have in mind.

We start from the assumption that in order to meet the above requirements our engineering products must be endowed with certain capabilities that nature has bestowed on her creatures in the course of billions of years of evolution. In fact, individual animals and animal populations, including humans and human societies, do enjoy to a greater or lesser extent the above characteristics. They are

[3] Presumably it was Ivan Illich ([Illich]) who introduced this term into the technology debate. Here it simply means that we want machines that are adapted to human needs rather than the other way round. To prevent humans from becoming (parts or extensions of) machines we need machines that behave more "human like".

able to survive under potentially adverse circumstances, owing to their particular physical make-up and ways of interacting with and within their own worlds.

We use the term "cognitive" to describe the particular capabilities that enable living organisms to control their presence in their worlds and thereby to guarantee, at least for a limited period of time, their continued existence. The term "Cognitive System" refers to any system with such capacities. We give a more detailed account of these and related concepts in Section 2 of this note.

Inasmuch as a science is defined by the research questions it is supposed to answer, the science we are concerned with here will likely encompass more than, say, Cognitive Science, Cybernetics or any of their constituent disciplines. It will be less of the artificial (to recall the title of Herbert A. Simon's classical monograph [Simon]) than for the artificial. Above all, it will have to clarify beyond intuitively appealing and more or less scholastic "definitions", the meaning of cognition in operational and possibly quantifiable terms - similar perhaps to 19th century Physics which clarified the concept of energy in terms of measurable quantities (well, at least to the extent of making it more readily usable, to the benefit of many). In Sections 4 and 5 of this note we expand on the general requirements of a scientific foundation for engineering Cognitive Systems and on its potential building blocks, respectively.

Finally, in Section 6, we look into ways of implementing research leading to establishing and strengthening that foundation, addressing (if only implicitly) questions such as "What theoretical and practical challenges spawn fruitful research?" (*What do we want to understand and what do we want to do?*) and "Who co-operates?".

Quite generally, what kinds of technical systems can be - and are being - built, depends on

- the available hardware and communication technologies;
- the available knowledge (in terms of theories, models, abstract designs of architectures, viability of (partially competing) paradigms etc.) of systems specifications, and of how hardware and communication components can be configured and operated to meet these specifications.

Therefore, research on engineering artificial cognitive systems has to focus on increasing our knowledge as to ...

- whether, to what extent, and how, desired, required, or to be determined cognitive capabilities can be implemented in systems (a robot, a communication support system, a domestic device, a sensor-network, a manufacturing plant, a power plant, etc.), given the available hardware and communication technologies and, failing the availability of the technologies needed,
- what (new) technologies - in terms of to be specified characteristics - would be needed in order to implement these capabilities in systems;

- how high-level capabilities can be "bootstrapped" from elementary ("low-level") ones;
- what is possible, what are the limitations;
- what new requirements regarding cognitive competencies can be or should be specified in view of emerging technological, societal or economic needs and opportunities.

We illustrate our answers primarily with examples from current activities in relevant sectors of the European IST Programme but, where appropriate, shall also refer to other initiatives worldwide.

2 Cognitive Systems

What is a cognitive system? There appears to be no universally agreed answer. But we are quite certain: cognitive systems abound in living nature, ranging from unicellular organisms to human beings.

This somewhat sweeping statement is meant to establish at the outset our position on cognition. The term shall not be restricted to tasks and processes peculiar to the human mind, such as conscious reasoning and planning, (symbolic) language based communication, abstract thinking, the creation and use of tools, or inventing games, mathematical problems and entire theories. Of course, these so called higher cognitive functions are of paramount interest and ultimately we want to understand how they come about. Yet they constitute but the uppermost layer of an immensely richer fabric of faculties based on mechanisms we are only just beginning to explore[4].

In general we may attribute the qualifier *cognitive* to an open system[5] that *actively* maintains[6], at least over a limited period of time and often against all odds, its integrity and stability within its environment (its encompassing system, that is) or, to put it differently, its internal workings and particular ways of interacting with other open systems, cognitive or not. To do so it must be equipped with certain structures and functions, depending on its environment and its own requirements regarding integrity and stability. These requirements are largely dictated by its physical make-up which includes the very structures

[4] One of the principal goals of a research programme in Cognitive Systems (and yet another formulation of the elusive mind-body problem) may in fact be exactly this: to find out *how on earth* (and why!) something like the human mind (including the phenomenon we call *consciousness*) could eventually emerge from the bubbling mud that covered our planet eons ago. (It most certainly did not come out of the blue!)

[5] An open - as opposed to a closed - system is a clearly distinguishable subsystem A of some larger system S the rest of which affects A and vice versa. A pebble in a brook is an open system but it has no way of defending its integrity against erosion; it is not active.

[6] ... from the perspective of an outside observer.

needed to maintain itself, and which in turn include the structures by which the system is coupled to its environment.

This "definition" seems to come very close to one proposed by Humberto Maturana and Francisco Varela ([Maturana]) who contend that being cognitive is a key feature of being alive - as expressed in our opening statement, which ascribes cognitive abilities to a human being and also to a single cell, and hence postulates a *continuum of cognition*[7]. But clearly: different biological species employ quite different means to maintain their integrity and stability; single cells for instance can not and need not develop symbolic communication to sustain their life in the same way as humans do. But they certainly do interact with their environment in ways not necessarily controlled by it ([di Primio], [BenJacob]).

Indeed, cognition is first and foremost a biological phenomenon; as such it is entirely subordinate to an organism's need to constantly reconstruct its unique self. Its scope and its degree of sophistication are contingent on the organism's evolved structures and functions.

Today's life forms have acquired their cognitive capabilities over a period of 3,5 to 4 billion years (the alleged age of life on earth) through a bundle of processes collectively known as evolution (leading up to phylogenesis and a multitude of co-evolving species); individual living entities develop these capabilities through growth (or ontogenesis) and learning (including learning from each other).

Learning can perhaps be best described as a process through which cognitive functions (some will be identified further below) and the corresponding behaviour can become more effective, efficient, and more precise. It happens under the constraints imposed by the environment and the physical equipment of the organism in question, and at different levels (species, individuals, groups, swarms, ...). Many different mechanisms are driving it. It is probably safe to say that the very ability to learn is itself subject to evolutionary change and that learning and memory are essential features of natural cognitive systems.

Whatever their degree of sophistication, the cognitive capabilities of a natural organism hinge on the incorporation of its environment through sensory channels that convey, in traditional parlance, *signals* in terms of molecules, photons, sound, direct mechanical impact, etc. (exteroception). They may also involve some sort of proprioception which may lead to changes of the internal state or

[7] There is little reason to believe that the principle "NATURA NON FACIT SALTUS" would not hold here as it does in other departments of natural history and science. It may of course be open to debate whether or not certain natural (i. e., non-man-made) objects that we do not consider living entities but which nevertheless process energy (for instance, stars) should be considered cognitive. From the point of view of observable *interaction* at least there is little to support such arguments. To ask for the "starting point" (a *Zero* in said continuum) of cognition is, however, a legitimate question.

the overall configuration of an organism, either independent of or in conjunction with what exists and happens outside.

But what exactly are these cognitive capabilities? How can they be characterised? What is their particular contribution to keeping an organism alive? And how do they relate to other capabilities that can not be deemed cognitive?

A brief reminder of non-equilibrium thermodynamics may provide a first and tentative answer: natural organisms are dissipative structures; their energy is being transformed into heat (*down-graded*) and, according to the second law of thermodynamics, irretrievably lost. They are prone to physical decay. (For *Biological Thermodynamics* see e.g., [Haynie].)

Their most basic need therefore is to restore the energy and matter required for keeping in motion the processes that (re-)produce their physical organisation, including of course - in a recursive twist - those parts that enable them to replenish energy and matter[8].

This is presumably where *cognition* sets in. The maintenance of an organism's structural and functional integrity (*survival*) requires ingesting matter and *low entropy* energy. A system that can do nothing but consume the resources at its immediate disposal is bound to stop functioning (consumption being its only occupation) and to disappear rather sooner than later - unless these resources are practically unlimited. (On a cosmic time scale this usually happens to stars.) By contrast, a system that can actively seek resources can stay intact over an extended period of time. Hence the primary (and most primitive?) cognitive capabilities of a natural organism ought to ensure that it can search for and detect sources of usable energy and matter, and thus counteract decay[9]. Needed in the first place to sustain the organism's self-construction these capabilities should

[8] Erwin Schrödinger, in his remarkable seminal 1943-1944 lectures on "*What is life?*" ([Schrödinger]), puts it as follows: "*What is the characteristic feature of life? When is a piece of matter said to be alive? When it goes on 'doing something', moving, exchanging material with its environment, and so forth, and that for a much longer period than we would expect of an inanimate piece of matter to 'keep going' under similar circumstances. When a system that is not alive is isolated or placed in a uniform environment, all motion usually comes to a standstill very soon as a result of various kinds of friction; differences of electric or chemical potential are equalized, substances which tend to form a chemical compound do so, temperature becomes uniform by heat conduction. After that the whole system fades away into a dead, inert lump of matter.*" He continues to argue that organisation is "*maintained by extracting 'order' from the environment*": "*How would we express in terms of the statistical theory the marvellous faculty of a living organism, by which it delays the decay into thermodynamical equilibrium (death)? We said before: 'It feeds upon negative entropy', attracting, as it were, a stream of negative entropy upon itself, to compensate the entropy increase it produces by living and thus to maintain itself on a stationary and fairly low entropy level.*"

[9] On our planet nature has found at least two fundamentally different solutions to the problem of resource scarcity and depletion: plants which rely on largely non-exhaustible resources (e.g., solar energy), and animals which are equipped with more or less sophisticated facilities allowing them to roam within their environment and search for nutrients.

also enable it to avoid destruction by recognising and preempting potentially harmful situations in its environment.

We do not specifically ask why a given open system or group of open systems would develop a propensity to last and - at the same time - the constituents, mechanisms and operations needed for that purpose. It is a question with no definitive answer yet (probably equivalent to the quest for the origin of life on earth, cf. [de Duve]).

We shall not primarily be interested either in the Why and How of the evolution of such systems but rather in the way they develop and work. Hence we content ourselves, at least for the time being, with asserting that they obviously exist and that it is possible to study cognitive capabilities within their specific physical context.

That context is given by the system's body and its environment, while the system itself may be understood in terms of a collection of processes[10], supported by the physical substrate of the system's body. In a first approximation they may be classified as either controlled or controlling processes. Specifically cognitive processes belong to the latter class. In biological terms they are bound to the sensory equipment (organs or organelles and, in a final analysis, suitable molecular structures) of an organism whereas controlled processes are associated with executive devices such as cilia, flagella or muscles (also, in a final analysis, *implemented* in cellular and molecular structures).

In a second approximation, however, the distinction may not be as clear-cut, due to intricate feedback and feedforward relationships between individual processes. Predominantly controlling processes may also be controlled and vice versa. In natural organisms for example this may happen through the presence and level of concentration of substances such as hormones or neurotransmitters (leading to global states usually referred to as *emotions*). This makes it difficult indeed to draw a line between genuinely cognitive processes of a living organism and other processes (e.g., digestion) supporting and preserving its life.

There are, nonetheless, a number of cognitive processes whose *right to exist*, we believe, can be derived more or less directly from the above postulated primary capabilities; for instance processes related to:

- attention, expectation, and active exploration of the environment (in response to internal or external stimuli);
- the identification, categorization or classification of objects and (internal or external) phenomena.[11]

[10] It may be convenient to take *processes* to be the abstract constituents of systems; a process manifests itself in the particular ways a system's physical substrate changes its state, driven by forces inherent in its material components (where the energy behind these forces may even be supplied by a battery).

[11] (See also [Roth] who propose a sort of hierarchy of cognitive processes.)

A common trait of many of these processes appears to be their strong reliance on operations on patterns, configurations that is, of physical phenomena in various dimensions (space, time, mass; derived dimensions like temperature, colour, etc.). These patterns may be internal or external to a given system; the set of operations on them includes their formation, detection, retrieval, (partial) matching and association, in conjunction with (integrative multi-modal) sensation and perception. Pattern formation in particular, driven and constrained by forces inherent in an ensemble of physical components (cf. footnote 10, page 8), is a general occurence in dynamical systems, living or not. It is likely to be a crucial factor not only in the emergence of life but also for some of the most prominent cognitive capabilities such as adaptation and learning or, more generally, memory (cf. for example, [Kelso]).

Indeed, many of the physical and/or behavioural adaptations to changes in their environment biological species have undergone in the course of evolution can be described in terms of changes of patterns. Most visible among those are the features and the assemblage of parts of their bodies (their *bauplan*). They represent in their totality what an organism has *learned* in its phylogenetic history.

Note that so far we have avoided the term *representation* whose connotations may be somewhat biased by its use within the context of discrete symbol manipulation. Yet, one may justifiably argue that patterns in the above mentioned sense can be understood as some kind of representation that occurs[12] in living organisms as they go along, interacting with and within their environment. Through evolution bodies themselves become representations of certain aspects of the world around them whereas the specific connectivity patterns for instance, that arise in an animal's neural tissue, can be viewed as representations resulting from the animal's individual (ontogenetic and epigenetic) history and from its activity in its world (see for example, [Jirsa]).

These representations encode the organism that brings them forth (including its ways of relating to its environment) through the ways it relates to its environment[13]. Being grounded in evolution, growth, action and learning they have meaning *sui generis*. In millions of years, concomitant with the development of cortical structures in brains, they have become more varied in some phylogenetic lineages, influenced by and causing increasingly *complex* behaviours, for example

[12] We deliberately refrain from putting this in *active mode*: organisms do not make or create internal representations of their external world and their own situation in that world; rather, these representations arise quasi as a byproduct of the organisms' acting in their world.

[13] This may be a somewhat convoluted way of repeating the fact that living entities are self-constructing (or *autopoietic*, to use the term coined by Maturana and Varela [Maturana]).

(human) language (in the sense of vocal or gestural utterances)[14]. Presumably, language, in this most general sense, is a prerequisite for an organism to create symbolic, structured, communicable and preservable external representations of its internal non- (or sub-) symbolic representations (or internal worlds!) and their transformations.

Internal representations and their transformations, if present at a sufficiently *high level*, are usually referred to as thoughts, imagination and thinking, which underly, as experience seems to confirm, the familiar particular mental faculties Homo Sapiens appears to be endowed with.

There is little doubt that the degree of his competence for (internal and external) language distinguishes Man from the rest of the animal kingdom. It unlocks cognitive dimensions way beyond the very basic capabilities necessary for an organism's survival: dimensions that have, in a final analysis, led to the remarkable insights into nature's ways, ranging from the very large (the cosmos) to the very small (e.g., the elementary constituents of matter), and which make up our current scientific knowledge.

These insights enable humans to control their environment and to change it, to an extent that no other living species has ever attained[15], while the changes, in turn, impinge on their individual and social development[16] in unprecedented ways[17]. The changes they bring about include the creation of physical artefacts of all sorts, and of the organisational structures underlying and determining human societies.

Man is also likely to be the animal that is best at learning and certainly the only animal that is capable of inventing and using symbol systems (or *formal worlds*) such as games, music, mathematics, money, religion (including God and gods) and laws - more or less abstract, more or less detached from firm physical, biological or social grounds.

Hence there is equally little doubt that to date Homo Sapiens is indeed the most highly developed natural cognitive system. The specifically human faculties of understanding the world, a given situation or an individual's situation within a given environment, and of making that understanding explicit through external representations, decisions or goal-oriented actions, are often collectively referred to as *intelligence*[18]. However, as pointed out at the beginning of this

[14] We understand the concept of *behaviour* to imply observability; what gives rise to behaviour may or may not be observable.

[15] Clearly, all living entities do affect their environment to a greater or lesser extent. Man's unique faculties, however, make him an actor in and - at the same time - a scriptwriter of Nature's play. A scriptwriter though, with rather limited knowledge of what he is writing about and limited understanding of the consequences of the actions his script entails.

[16] And - one may add - if worst comes to worst then also on their genetic development.

[17] Or, more succinctly: We shape the world that shapes us.

[18] ... with marked, yet difficult to measure differences in capacity and capabilities between individuals.

section, we must not identify this particular notion of human intelligence with the general concept of cognition. Intelligent behaviour in a real world environment, as exhibited by humans, requires cognitive capabilities whereas the latter do not necessarily entail the former. A cognitive system as described in this section, is not necessarily also an intelligent system, a system that is, which is capable of performing the intellectual feats suitably talented humans (but apparently no other animals) can perform (e.g., as listed in the second paragraph of this section).

Notwithstanding the fact that human intelligence has come up with intricate rule based symbol manipulation systems of various kinds, it must not be confused either with the ability to do rule based symbol manipulation fast - abiding by and following the rules and nothing but the rules. This requires no cognitive capabilities whatsoever (let alone intelligence)[19]. Yet human intelligence may be (and indeed, is) very useful in solving formally specified problems - for the very reason that it is rooted in *low-level* cognition which *operates* on patterns we are not normally aware of, and which - by all accounts - does not meticulously construct and peruse complex webs of logical deductions expressed in terms of discrete symbols.

Games such as Go or Hex, are prime examples of formal systems where subconscious *reasoning* on patterns forming in human brains can still outsmart any algorithm (sequential or parallel) running on current computer hardware; for Chess it took several decades after all, before the algorithmic (more or less) *brute force* approach gained the upper hand on human intelligence. Mathematics, of course, is another inexhaustible source of examples.

Hence, natural cognitive systems are, in many respects, much more powerful than discrete symbol manipulation systems implemented on von Neumann type computers; in other respects they are - as pointed out - much less powerful, in that they cannot compete for instance, against the speed at which modern hardware executes a programme that yields a hundred thousand digits of π. A system of the latter sort, but also a chess playing machine that beats the world champion, must therefore not be mistaken either for the kind of *human-like intelligent system* that cognitive systems (at least the natural ones) obviously can become. Electronic number crunchers, game-tree builders and pruners are mere *intellectual cranes, cars or planes*, tools that is, whose sole purpose is to overcome the limitations of Man's physical capabilities - here of course we refer to those of his brain which, no doubt, is able to compute (crunch numbers, build, traverse and prune game-trees) but - we repeat - is very slow at performing this sort of operations. To put it in a nutshell: brains can emulate computers; whether

[19] In electronic computers the execution of a programme is typically driven by something like an oscillating quartz, certainly not an object that can be ascribed cognitive capabilities.

computers (as we know them) can emulate brains is an open question likely to be answered in the negative[20].

It may be a sort of narcissistic hybris, an obsession or perhaps fascination with his own mental faculties, that has biased Man's understanding of his own mind[21], and led him to focus on his own cleverness in dealing with formal systems (e.g., through symbolic reasoning) and, for some time at least also obscured the fact that *human likeness* is not at all a general feature of cognitive systems. This focus has perhaps most clearly been adopted within research activities known since around the mid-50's of the 20th century as *Artificial Intelligence (AI)*. Achieving artificial systems with *"human-level intelligence"* has been on the agenda of AI basically from its very beginning. This endeavour has certainly been flawed - inter alia - by the fact that there may be no agreement on what "human-level intelligence" actually means. Arguably, the capacity (as low as it may be) for symbolic reasoning in isolation is not at all the only distinguishing mark of "human-level intelligence" (although, of course, to the best of our knowledge, no other species has attained it). And perhaps humans are not so very intelligent after all: a conjecture difficult to refute given the past and current failure of humankind to organise its life on this planet for the benefit of all[22].

We shall see, however, that under certain circumstances some sort of *human likeness* - to be interpreted very carefully - may indeed be de rigueur when we enter the realm of artificial cognitive systems.

3 Artificial cognitive systems - whether, why and what[23]

In view of the preceding discussion one may suspect the very notion of *artificial cognitive systems* to be a contradiction in terms. However, to accept that living entities (as amply present in nature) are cognitive systems does not imply that we have to agree to the converse of that proposition. Our systemic "definition" of *cognitive system* does not refer to biology; and being *alive*, while possibly sufficient, may not be a necessary condition for a system to be deemed to have cognitive capabilities[24].

---

[20] In subsection 5.3 we will say more about *emulation*, and in subsection 6.3.2 more about unconventional computing and computers.

[21] The need for bringing that understanding down to earth has been pointed out in footnote 4, 5

[22] Early critiques of some of the approaches taken under the AI label have been put forward for example by Dreyfus ([Dreyfus]), Weizenbaum ([Weizenbaum2]), and Searle ([Searle]) - and by each for different reasons.

[23] This and the following section in part draw on early discussions of possible directions of the Cognitive Systems initiative under the European Commission's research programmes (cf., [Stork2]); they also benefitted from personal communications from Aaron Sloman (University of Birmingham, `http://www.cs.bham.ac.uk/~axs/`); they reflect, however, only this author's interpretation.

[24] In [Maturana] Maturana and Varela boldly declare *"living is cognition"*. However, Evan Thompson, referring to earlier writings by Maturana, points out that the "is"

We may therefore rightfully ask (1) whether processes such as evolution, growth and interaction with the environment are the only means of endowing matter with such capabilities; (2) whether cognition is possible in entities that have not evolved (like a biological species), that do not grow (like an animal or a plant), and that do not learn while interacting with their environment? Or (3) whether an artefact that can learn to operate autonomously and purposefully in a given environment must in some sense be considered alive?

At first glance the answers to these questions appear to be "*no, yes, no*": at least some of the cognitive functions and processes that help maintaining a living organism's integrity and stability might well be implementable in entirely artificial technical settings, without being involved in keeping up other - non-cognitive - processes (such as energy supply), without having reached a certain degree of perfection through some kind of evolutionary process, and without having to learn anything that it can not already do. We might want to make use of these cognitive functions and nothing else, somewhat similar to the border guard who wants his dog to bark if and only if it sniffs illegal drugs in a traveller's bag. All that person is interested in are the dog's particular sensors and its reaction to specific molecules its nose is sensitive to. In this particular context, whether or not the sense of smell is of any importance (while searching for food, mates, etc.) beyond that basic service does not matter at all. (In other words: we do not care whether or not the smell of things has any meaning for the dog.)

The feasibility of this approach, wholly based on deliberate design, has long since been demonstrated. There are *artificial "dogs"*, for instance, and there are machines with artificial eyes attached to them that can distinguish between objects of different shape, colour, texture, and other features. Likewise, the well established discipline of *Machine Learning* (of which more will be said in section 6.3.1) in particular has, over many years, generated numerous computational methods for improving the precision of presumably rather basic cognitive functions like object (pattern, event, ...) identification, classification and categorisation, implemented on (sequential or parallel) von Neumann type computer platforms, and put to work within a wide range of application scenarios. Lastly, machines have been around for a fairly long time that can exert some control over themselves and over processes in their environment, based on evaluating input from sensing or measuring devices.

These examples seem to prove that cognitive functions can indeed be detached from their natural biological context and that artefacts can be endowed with cognitive capabilities. In fact, there seems to be no a priori (let alone a posteriori) argument against this being the case.

---

in this assertion must not be interpreted as an identity of concepts but as an inclusion: *"living systems are cognitive systems"* ([ThompsonE]).

As explained in the previous section, natural cognition in its most basic forms is strictly subordinate to the supreme goal of the survival of some organism[25]. In a final analysis, this may also be true for *less basic* cognitive functions such as Man's well known *high level* mental faculties (cf. the previous section, 2nd paragraph). This leads to questions like the following: Why would we want artificial systems - within narrower or wider limits - to be cognitive? What tasks should their cognitive functions be applied to - if not to ensuring the survival of some living organism? Or else: does it make sense to talk about the *survival* of an artificial system and if so in what way? What is so special about natural cognitive systems that we want artefacts to be - in certain respects at least - like them? And of course we might ask why we need machines that are even *cleverer* than the ones we already have?

Depending on the enquirer's stance, the latter question is probably either the most difficult or the most straightforward to answer. Taking to the middle ground we may argue that the artificial systems we already have - including those which we have just alluded to - are not yet very clever at all!

These systems cover a wide range of infrastructures and infrastructural components, of products and services, that Man has been creating during his long history as a species. They have become increasingly complex, and are now integral parts of his world(s). To a large extent Man's existence in its present form hinges on them (at least in some parts of this planet).

There are for example those that are supposed to meet the growing demands of human societies for uninterrupted energy supply, safe transportation, efficient production, and reliable communication. Infrastructural systems - e.g., energy, transport, communication or content networks, but also power plants and industrial facilities of all sorts - constitute entirely new environments and *bodies*, artificial, yet more or less strongly connected to processes with indeterminate and hard-to-predict dynamics. Many of these processes typically involve people, one of the major sources of uncertainty.

Other systems, often of a more compact nature, are employed in science, healthcare, education, information or entertainment, and in environments which are either inaccessible or hazardous to humans. Many of these systems ought to be designed for ease of use by people from all walks of life.

Most, if not all of our artificial systems, rely solely on human intelligence (as outlined in the previous section), both in their design and their operation. Where some sort of interpretation of data is required, be they obtained through sensors, through interaction with people, or through other input channels, that interpretation has to be supplied by a human designer or operator. So far, data processing machines, be they embedded, networked or stand-alone, have no *understanding*

[25] ... which in turn may be subject to the even higher goal of perpetuating the species the organism belongs to.

of their own; they work purely syntactically, like the number crunchers and game-tree manipulators mentioned in the previous section. Their intelligence may in fact be referred to as *syntactic intelligence* (as of Searle's *Chinese Room*, cf. [Searle][26]), a mere amplification and speed-up of Man's own symbolic reasoning and algorithmic capabilities[27].

By contrast, natural cognitive systems, while - for the most part - lacking the *syntactic intelligence* man-made machines can provide, are individually or jointly more robust, versatile and flexible, more responsive and also more autonomous (e.g., pro-active, self-sustaining, ...) than any of our artificial systems. Harnessing the potential of artificial systems would require precisely these characteristics. They are expected to make such systems ...

- fitter for use in contexts where they are normally used, or
- fit for use in contexts where they could otherwise not be used.

In the remainder of this section we discuss in very general terms relevant contexts of artificial cognition ("What do environment, body and cognition mean in artificial contexts?"), corresponding tasks ("What should artificial bodies do in their environments?"), and desirable or required features and traits ("What qualities should bodies and their behaviour have?").

Generally speaking, all contexts are relevant where artificial systems could or should render some service to processes running in their environment (i.e., outside the system); this may in particular include or simply boil down to ...

(1) providing information about (parts of) their environment and/or themselves (through classification, categorisation, recognition and description of objects, events and processes); and/or

(2) exerting some control over processes in their environment (directly through immediate action or indirectly through suggesting human decisions) and/or

(3) within their own "bodies".

We shall argue here and in subsequent sections that (3) may actually be necessary for (1) and (2). Further below (especially in sections 5.2 and 5.3) we shall also endorse the conjecture that anything worthy of being called *"cognitive"* in artificial systems must come about through processes similar (but of course not

[26] Searle's Chinese Room argument, meant to criticise bold claims popular with the Artificial Intelligence community, spawned a long lasting and still not fully subsided discussion on whether or not an artificial system can have some - human-like - understanding of language. See also [Stork] for a brief explanation of the significance of the argument.

[27] To the best of this author's knowledge none could ever be induced to come up of its own accord, with an ingenious new algorithm, for example. And up until now machines can be "intelligent" if and only if they are built / programmed by intelligent people.

identical) to those that natural systems are undergoing (i.e., evolution, growth and post-natal learning).

The concept of *environment* is very broad in scope. It ranges from (parts or aspects of) natural *real-worlds* to (man-made) *artificial worlds*. The former may be uncharted territory (outer space, distant planets), volcano slopes, the deep sea, a disaster area or one or several human beings; the latter include networks of all sorts and their contents, but also power plants, factories, laboratories, operating theatres, buildings, offices, many other *socio-technical* constructs; and there is a range of hybrid or *mixed*, environments.

The concepts of *environment* and *body* may also not be as clear-cut in *artificial contexts* as they seem to be when applied to natural systems. (And even there the distinction is not as sharp as one might believe: bodies - e.g., cells - unite with what used to be items in their environment, and form larger bodies, with different environments, etc.) A power or production plant for instance, may from one (internal) perspective be perceived as an environment; from a different (external) perspective it may be seen as a body within the larger environment of a society and its economy. A building is an environment for people but people could well be understood to be (part of) the environment of a building ([Rutishauser]). Hence the notions of body and environment are indeed relative. (We discuss these "environmental issues" in more detail in section 6.1.2.)

In keeping with footnote 10 (page 8), perhaps the best way of resolving this *relativity dilemma* is to define body and environment in terms of two classes of processes: those that belong to *the body* and those that belong to *the environment*, depending on context and perspective. Usually, but not necessarily, the processes in these classes are supported by physical substrates that are separated by some clearly discernible boundary (as is the case with most natural systems). What matters is that the separation in two classes reflects major structural differences between two different sets of processes, the ones *inside* and the ones *outside*, and that these sets can be viewed as coupled via (material and/or informational) links between all or some of their respective members.

An artificial system typically obtains its cognitive capabilities - if any - through one or several specific, physically manifest subsystems which may or may not be physically integrated (*in the same body*) with other parts of the system; in either event these subsystems must take into account the physical and architectural[28] characteristics of the larger system (e.g., a robot's moving parts, or the above mentioned production plant when seen as a *body*), including the particular ways the larger system is connected to its environment. Consequently, there are several *layers of control* within an artificial cognitive system and between the entire system and its environment.

[28] Here, as in further instances the term architecture refers to the abstract functions of components and subsystems and the relations existing between them.

This is similar to architectures of *higher* forms of life - e.g., mammals - where various body-internal subsystems at different *levels* (e.g., neural, endocrinal) cooperate in the *implementation* of cognitive functions. These subsystems (e.g., brains, glands), however, are always embedded whereas one may well conceive artificial cognitive systems whose *cognitive parts* are separate from the *non-cognitive parts*. The *physical instantiations* of architectures for artificial cognitive systems may be *distributed* in their environments (and even form *swarms*), with components (e.g., sensors, cameras, actuators and corresponding processes) communicating via some (wireless) network. This relativizes further the concept of *body* in artificial contexts; and where it not for the fact that the more general term *organism* appears to be reserved for living entities, it would indeed be more appropriate to use that very term instead, meaning a *functional structured assemblage of functional physical components*.

(Living) organisms succeed in their environments if they manage to achieve their supreme goal of staying alive for as long as possible (or for at least as long as they need to help perpetuating the existence of their genetic substance through some form of reproduction; cf. footnote 25, page 14). Clearly, this success criterion does not apply to artificial systems, or at least not literally. After all, natural organisms are there only for the sake of being there whereas artificial systems are supposed to render some service to people. They are "successful" if they can provide that service at a quality level that meets their users' expectations. A consistently high (and possibly increasing) "quality of service" could hence be postulated as a (the?) general (supreme?) goal to be pursued by an artificial system. This of course can mean many things, depending on the kind of service (within a given environment) and the expectations regarding that service. It can in particular mean that a system be able to continue functioning (to "survive"!) under critical conditions, maintaining its stability and, more or less, its performance levels.

It follows that regardless of the nature of its service an artificial system would profit from cognitive functions whose role is analogous to the life-maintaining role of cognition in living organisms: figuratively speaking, they should contribute to keeping the system *alive*. They should enable the system to recognise situations, i.e. patterns in the space and time of its environment and itself, that might jeopardise its proper operation, and to counteract accordingly. The precision and robustness of its perception, the efficiency with which it uses its resources, the ability to acquire and organise relevant knowledge (representations) about its world and itself, to cope with frequent and hard-to-predict change, and to act in time: these are among the key basic features that can make a system fitter for whatever service it is supposed to provide. Here, one has to keep in mind that rendering a service is certainly not a one-off occurrence. In many environments

it requires constant attention and readiness for action and interaction (often in *real-time*).

Endowing artificial systems with cognitive capabilities would - under the above interpretation - entail further desirable general quality traits, such as reliability and availability, and even more so if artificial cognition could be extended to the level of the elementary physical components of such systems. Cognitive functions working at that (*low*) level may for instance detect faulty components and trigger appropriate repair processes without greatly disturbing the overall operation of the system. This would lead to graceful degradation at worst[29].

We repeat: all artificial systems provide - at least in a final analysis - their services to people and in more or less direct contact with people. For those systems that people interact with closely, the ability to communicate on human terms can be an essential and, in fact, indispensable quality aspect. This (but not only this) will, in the best case, require *high-level* cognitive capabilities such as language recognition and production, rational thinking and common sense, human style. It also has to take into account the emotional (pre-)disposition of human interactors. These cognitive capabilities must be compatible with socially agreed human *ontologies* (see Table 2) and interpretations. In that sense certain types of artificial cognitive systems ought to be *human-like*. Human-likeness does not mean that the system has to look, move or in some other way behave like a human being (although this may help tickle some people's fancy). It is neither a Golem nor a Frankenstein.

There is a variety of tasks that pertain to different types of environments and relevant services, and whose execution by, within or in interaction with artificial systems would greatly benefit from *cognitive approaches* taken by these systems - quite apart from the *life sustaining* role of cognition. Such tasks would typically have to do with augmenting and extending human cognitive and/or physical capabilities, and with lessening human *cognitive load* while coping with all sorts of (natural and man-made) complex systems and situations.

This is certainly not surprising, given that most (if not all) of our artificial systems (perhaps even our societies) have been created for the very purpose (albeit not always on purpose) of overcoming our natural limitations[30] (see also our comments towards the end of the previous section, on certain forms of electronic computing). Indeed: our (like any other living organism's) cognitive capabilities

[29] Artificial systems with this kind of "low-level" cognitive capability have been in existence for a long time and they keep growing: computer communication networks for instance, where *quality of service* is a well-defined (and quantifiable) notion. While they have built-in mechanisms to detect faulty components (such as nodes or links) and even to find out what is wrong, they cannot as yet repair themselves. Computer networks can also be subsumed under the general heading of *fault-tolerant computing*.

[30] This being said, one also has to admit that our artificial systems (including societies) have long since been imposing their limitations on us; these must be dealt with as well.

What is (/an) **(O/o)ntology**?

With an upper case **"O" (Ontology)** the term denotes the philosophical discipline that deals with the general idea of "*Being*" (*"Why is there something rather than nothing?"*). There is only one Ontology.

The more mundane lower case **"ontology"** refers to an understanding of (a part of) the world in terms of concepts that relate to (or *"model"*) phenomena in (that part of) the world. In principle there are as many ontologies as there are (autonomous) understanders of something. A system cannot be deemed to behave *naturally* (see Table 1) if its design and behaviour are not compatible with socially agreed human ontologies. Meaningful interpersonal communication (verbal or not) is not possible without shared ontologies.

More narrowly, an **"ontology"** is a *"formal, explicit specification of a shared conceptualisation"* ([Gruber]). Axiom systems (e.g., for Euclidean geometry) are classical examples. Explicitness and formality are required in order to make a shared conceptualisation actionable for and within an artificial system. As such it usually comes down to some data structure representing the terms, relationships and rules that are pertinent to - and hence describe - "the world" at issue. It provides the *semantic ground* for all sorts of inferencing algorithms.

Table 2: Ontology and ontologies

are constrained by their biological underpinnings which determine the kind of energy flow and conduits which enable a human body to interact with its environment, and the manner and speed of processing that flow. We can, in plain and simple terms, see the light, but we have no organ to pick up radar or radio waves; sounds beyond a certain frequency we cannot hear; and our direct impact on the world hinges on what we can feel through our skin and do by using our muscles and limbs, manipulating or just touching the things around us. The speed at which we perform algorithmic symbol manipulation (e.g., arithmetic, scheduling, etc.) mentally (and even with aids such as pen and paper) is piteously slow, our own consciously accessible memory for facts and figures is not very reliable and its capacity is rather low.

Technology has come a long way to dealing with many of these impediments. We now have at our disposal a large number of tools, appliances, devices, machines, and so on, that help us cope with our shortcomings and greatly surpass our manual and - at least in some respect - mental capabilities and capacities. All sorts of appliances assist us in our daily chores, at home and at work; others

provide information and entertainment; vehicles, roads and other transport systems take us from A to B across large distances; numerically controlled ("NC") machines equipped with grippers and sensors (*robots*) move parts along assembly lines and put them together to form cars, TV sets or other machines; we have devices that can react in one way or another to signals that we have no natural receptor for; we have devices that can draw our attention to certain changes in artificial or natural environments; that operate under conditions that would be prohibitive for us - in outer space, on distant planets, in disaster areas and other places that are too dangerous for us to be at; semi-automatic real-time control of large plants takes place as a matter of course.

Much of what these devices, machines and systems do is owed to digital computing technology which has made it possible to process quantized signals and symbolic data at great speed and precision. To us these data usually represent - among other things - numbers, text or images; hence, suitably programmed computers can help us to understand what these data mean and to discover relationships between them (e.g., through data analysis, data mining, etc.). They help us to plan and make informed decisions wherever some formal evaluation of possible scenarios or options is required.

As implicitly mentioned further above, many of these machines already have functionalities that can be considered *cognitive*. However, as we also pointed out, these functionalities are still - to coin a phrase[31] - fraught with a sort of *syntactic rigidity* (and ensuing *brittleness*) that in the long run we should not have to put up with. Hence, all our current technical systems are candidates for becoming even *smarter* by acquiring largely *of their own accord* and through suitable *training* some *understanding* of what they are doing, what is happening to them now and what might be happening to them in the future. Machines people use in their daily lives should "understand" their users rather than require their users to understand them. So far, the only indication that this should be possible is given by nature, and her unique CHNOP++ -based[32] technology (horses and dogs are prime examples). Nature has indeed set a "standard" in solving seemingly simple problems[33] that none of our past and current technologies has been able to meet.

The following bullet points summarise what could be key capabilities artificial cognitive systems should possess in order to be (where and as appropriate) more flexible, robust, adaptive, convivial, autonomous and, at least, appear smarter than traditional computer-based information processing systems. Each of these capabilities is certainly related to some of the others, and all should be present,

[31] ... at least in this context; Google returns some 20 hits after all, referring to linguistics or data interchange formalisms.

[32] C-Carbon, H-Hydrogen, N-Nitrogen, O-Oxygen, P-Phosphorus

[33] "simple" because our (and other animals') bodies can solve these problems without even pausing for reflection.

to greater or lesser extent, depending on the specific environment and tasks at issue. Artificial cognitive systems should ...

- interpret (*make sense of*, *give meaning to*) whatever they are poised to sense in whatever environment they are operating (this includes the recognition of affordances[34], the interpretation of environmental data in terms of abstract roles or goals, etc.);
- develop, in terms of suitable representations, some *understanding* (or *awareness*) of their own role and situation in whatever environment they are operating;
- predict (and anticipate) future events in their environment (including, where relevant, the behaviour of other agents - human or not - operating in the same environment);
- learn (supervised or unsupervised, through interactions with their environment) in order to modify if necessary, the way they operate and/or to improve their performance according to given criteria (including criteria related to the system's use of its resources);
- pursue goals (set by humans but also *by themselves*, possibly inherent in their architecture/design, e.g., to bring about changes in a given environment);
- behave *sensibly* and *robustly* under conditions of uncertainty (e.g., in uncharted, non-deterministic environments, and everyday situations), e.g., deal with novel situations in terms of previously learned regularities, through generalisation and analogical reasoning;
- communicate and interact with people on human terms (e.g., in natural language).

Clearly, many of the concepts in the above list are as yet ill-defined. It shall indeed be within the remit of the sought after scientific foundation to make their meaning more precise, as will be expounded in the following section.

Further below (in section 6) we shall also see that various forms of biomimetics are currently among the more popular *philosophies* guiding research on specifying and implementing artificial cognitive systems. The basic assumption is that it should be possible to *learn from nature how to engineer artificial cognitive systems.*

This addresses the technical problem: to produce *better* artefacts. But there is the more fundamental epistemic problem: to find out, as a matter of human curiosity, how our mind works (cf. also footnote 4, page 5). Once - and certainly

[34] In the sense of Gibson: the actions an environment offers its inhabitants (cf. [Gibson]).

for many, still - one of the hottest topics of speculative philosophical debate, it is now also considered one of the most prominent problems at the cutting edge of science. Building *artificial minds* based on models derived from studying cognition as it occurs in nature, and *playing* with the resulting artefacts, is seen as a promising way of approaching a solution. This of course, leads to quite different answers to the *whether, why and what* questions posed in the heading of this section; answers that can perhaps be summed up most succinctly and somewhat metaphorically as in the title of [Harvey]: artificial cognitive systems might enable us *to pursue the philosophy of mind by using a screwdriver* (one may of course add soldering irons, networks of sensors, FPGA chips, and many more tools and components).

However, knowing in ever growing detail *how the mind works* may have more serious consequences - beyond engineering and beyond epistemics - than just satisfying our curiosity. It may open up ways of manipulating minds, that have hitherto undreamed of effects. It leads to ethical and moral questions Man has never been confronted with before. That would also be true should we ever succeed to build systems with capabilities like the ones listed above.

Autonomy (as for example in the *autonomous pursuit of goals*) is the key term. Time and again it will recur in these notes. In each instance though, one has to be very careful in its interpretation and weighing its implications. Under no circumstances can we wish machines to have autonomy in the same sense as we attribute this feature to a person. For arguments that will become apparent (especially in connection with the concept of the *embodiment*[35] of cognitive processes) it is very likely impossible for man-made artificial systems to ever be able to experience the world in the same depth and breadth (and innumerable other dimensions) as a human being or, for that matter, any other living entity. Hence, under no circumstances must we allow machines to take over tasks that require moral judgement, empathy, or choices only humans can and, indeed, must make. Unfortunately, the temptation seems to be great in our modern (western or western-style) societies to give up our own autonomy vis-à-vis our technical creations. We have got used to relying on them in ever greater degree. Whether this is happening because of a certain lethargy or indifference in facing our responsibilities or because empathy and moral judgement are becoming ever scarcer commodities among people, is an open question. (For an excellent discussion of ethical and moral issues related to modern computer technology - and in particular to systems of the sort at issue in this note - see for instance [Weizenbaum2].)

[35] For a discussion of the scope of this concept see for example [Ziemke2].

4 A scientific foundation - what it should support and how

There are generic and specific answers. The generic ones should be obvious, given the historical record of science and its relationship to technology and engineering. Almost none of our modern technologies would exist without the basic knowledge generated through scientific research: knowledge expressed in terms of models, principles and laws of nature, which set the boundaries within which engineering can happen and which inform the practice of engineering itself: the interwined activities usually referred to as specification, design, implementation, test and deployment of artefacts. Given the social dimensions of many of these artefacts, these activities should also be informed by insights gained in various branches of the humanities, such as the social sciences and psychology. In this note our primary focus is on the natural sciences.

The specific answers depend on the specific issues that arise as a particular technology evolves. These issues become apparent through growing sets of research problems. A scientific foundation provides the ground for tackling such problems methodically and with some likelihood of success.

In this section we highlight a number of potential research questions pertaining to our domain of interest, as outlined above. The most general (*big*) question, we believe, is the following:

> *What (if anything) do we need to understand about cognition as a biological phenomenon in order to specify, design and build artificial cognitive systems?*

Indeed, to start from the biological origins of cognition in order to gain an understanding of its *whys, whats and hows* now seems to be commonplace. Understanding natural cognition should contribute to a general *Theory of Cognition* (or *Theories* ...), which in turn should feed into a *Theory of Engineering Artificial Cognitive Systems*.

The *big question* requires a methodical and conceptual framework within which answers can be found and formulated; it should give operational (rather than denotational) meaning to the very terminology that ought to underly a scientific discourse on cognition, permitting a shared understanding of key concepts (e.g., *information*, *knowledge*, *organisation*, *organism*, *body, machine*, *environment*, and many others, not to mention *cognition* itself). A sort of *physics of cognition* would be needed, with (ideally) a set of precisely defined basic and derived notions, and which allows for falsifiable hypotheses and replicable experiments and observation.

Of those elements, aspects and characteristics of natural cognition that we do not yet fully understand, and these are many if not all, we mention but a few:

- The emergence, evolution and development of cognitive capabilities in living organisms; the way they relate to bodily needs and activity;
- the physical structures and functions underlying cognitive capabilities and processes (e.g., "What is the role of the physical substrate?", "What are minimal requirements on the physical substrate?", "What is the role of embodiment?");
- the *mechanisms* of recognising objects, actions and situations, and of adapting behaviour within non-deterministic environments;
- the types and levels of internal representations (of external phenomena in the time and space of the relevant environment or *umwelt*) and the ways they come about, change and interrelate through cognitive processes;
- the role and instantiations of memory and learning in cognitive systems (e.g., learning through system-environment interaction / communication, or based on schemas originating from evolution, culture or previous individual experience);
- goal-setting mechanisms and the development of strategies for achieving goals (sub-goal discovery and purpose-driven learning, e.g., learning predation / predator-avoidance; how can goals (desires, preferences, values, moods, intentions, etc.) be identified by (or to) a natural cognitive system?);
- the nature and role of emotion and affect;
- self-awareness, consciousness (for example in the sense of *anticipation / simulation of bodily activity*), intentionality and Theory of Mind, and how these relate to *higher-level* human cognition;
- the role of language in cognition and of cognition in language.

Apart from these points which call for an explanation of largely *inner* contexts of cognition a theoretical framework should also address the impact of communication, cooperation, and competition among cognitive agents. It should account for the socio-cultural (*outer*) contexts of human cognition, including the role (physical and social) artefacts play in controlling individual behavior and mediating social interaction.

One of the most challenging research problems may indeed be the How and Why of the emergence of human intelligence and creativity. The *definition* of Cognitive Systems, given in Section 2, deliberately does not mention explicitly Homo Sapiens's admirable achievements on the cognitive front and in fact, does not seem to even imply anything of the sort. Yet, the potential for these achievements must be implicit unless we posit the invisible hand of some divine entity.

On a very general level an engineer of artificial cognitive systems might wish to address questions like the following:

- Which sorts of artificial cognitive systems need what form of embodiment and why? (See also the discussion in the preceding section, regarding *environment* and *body*.)
- Which sorts of memory (mechanisms) are required in an artificial cognitive system?
- What are the modes and mechanisms of learning needed in an artificial cognitive system?
- What form and degree of autonomy is desirable and achievable? To what extent does the impossibility of identifying all problems in advance require designs to allow for the possibility of self-debugging and self-modification at run time?
- To what extent can natural cognitive traits such as affect, consciousness or *Theory of Mind*, be modeled and used in artificial systems? (And vice versa: to what extent can research on artificial versions shed light on the natural counterparts?; cf. the comments at the end of the previous section.)
- For an artificial system to be (or to become) cognitive, does it necessarily have to be *self-X* (X∈{monitoring, maintaining, healing, configuring, controlling, adapting, understanding, aware, generating, ...})?

We believe that none of these questions can be answered satisfactorily without a deep understanding of natural cognition.

A *Theory of Engineering Artificial Cognitive Systems* (or TEACS) should also address basic questions concerning the very specification and implementation of such systems, for instance:

- How can cognitive capabilities be specified and integrated in artificial systems?

This is the quest for formal, testable specifications of systems requirements, for example in terms of structural and behavioural constraints, learning abilities and adaptive performance, or temporal demands (*what can/must/should be done when?*).

It is the quest ... for ways of embedding cognitive agents in large-scale, long-running systems; for knowledge representations that can support action in non-deterministic environments; and for ways of integrating existing *Artificial Intelligence (AI)* components.

A TEACS should cater for effective design and construction principles and methods that meet general demands on robustness in the presence of noisy and

ambiguous perturbations (of the system's state, by its environment), on flexibility, versatility, stability and autonomy. Typical questions to ponder would probably be the following:

- What distinguishes a blueprint for an artificial cognitive system from *ordinary* blueprints?
- To what extent can, must or should cognitive systems engineering take a biomimetic approach?
- What balance has to be struck between evolution, growth and learning on the one hand and complete premeditated design on the other hand?
- What are basic abstract architectural components for complex cognitive agents (e.g., to become reflective, introspective and affective)?
- What is needed - in abstract architectural terms - for a system to be *self-X* (if it has to be; X as above)?

A TEACS should give guidance on what needs to be built into systems to bootstrap action, perception (e.g., in terms of basic *reflexes* and *pattern recognition* capabilities), and (unsupervised or supervised) learning, as determined by the problem space at issue.

It should provide the (empirically corroborated) models and methods needed to exploit fully the potential of components such as sensors, processing elements and communication devices, in their presently available and in their likely to become available forms. This would imply shedding light on the fundamental limitations imposed by whatever physical components are at hand, on the realisation of cognitive functions in artefacts.

It should offer further clues as to how to bridge a perceived *software-hardware gap*, for instance by exploring requirements on hardware that can support *cognitive architectures* and processes directly, without any mediating layer. This may also yield an answer to:

- What is needed - in physical terms - for a system to be *self-X* (X as above)?

There is also an issue of scaling. Artefacts range from the very small to the very large, their activity cycles can be of varying length and overlapping. Ideally, a TEACS would suggest approaches that span several relevant dimensions, such as space (e.g., complex wide-area plants), time (multiple time-scale loops), rationality (levels of reflection), and size (down to embedded).

Last but not least, and in keeping with Immanuel Kant's insight in the importance of the role of mathematics in any scientific endeavour (cf. footnote 2), we must insist on the availability of sound mathematical methods, not only

for modelling and analysing natural cognition but also for the creation and validation of man-made designs. Hence, our quest for a scientific foundation for engineering artificial cognitive systems extends to asking

- what various branches of mathematics can offer towards providing a solid theoretical underpinning of that foundation, and
- where "new mathematics" (i.e., mathematical research that is motivated by problems arising in Artificial Cognitive Systems (ACS) research) would be required.

We shall revisit some of these questions in subsequent sections (especially section 6) of this note.

5 A scientific foundation - bricks and mortar

We repeat: a thorough understanding of natural cognition is needed if we want to build artificial cognitive systems or make artificial systems cognitive. In that regard our engineering discipline differs from any previous endeavours to systematically design and build artefacts of all sorts. *Biomimetics* has certainly made its way into many traditional branches of engineering. But: we could build aeroplanes without a complete grasp of how birds fly and we could build computers and write programmes without knowing anything about the way brains calculate. It is highly unlikely, however, that we are able to create cognitive machines (in the sense apparent from these notes) unless we comprehend at least the basics of the *mechanisms* of natural cognition as these mechanisms define the very concept of *cognition* in the first place.

In order to become operational that understanding must be expressed in terms of models that will necessarily be abstractions from relevant phenomena as they occur in Nature. A scientific foundation for the engineering of artificial cognitive systems may therefore be conceived as consisting of at least three *layers* which pertain to

1. the analysis of natural cognition (*"collecting the data"*),
2. the abstract modelling of cognitive processes and their interactions (*"interpreting the data"*), and
3. the implementation (synthesis) of cognitive machines or of cognitive processes in machines, based on abstract models (*"testing the interpretation"*).

It goes without saying that each of these layers (analysis, modelling and synthesis) each in its own way, has a long history. In this section we delineate them in a few rather broad strokes. It must be clear though that the *layer* metaphor

reflects only part (and a small one at that) of the story. In actual fact the activities usually attributed to these layers are intricately intertwined, not unlike the processes that make up the very subject of the present notes: cognition itself. Science is its pinnacle after all.

5.1 Analysis

Given our interpretation of the term *cognition* and insofar as that term refers to a class of processes occurring in living Nature, the scope of *analysis* ranges from the most primitive to the most *complex* forms of life. It should address cognitive phenomena at all levels, molecular, cellular, neural, mental (perhaps *epiphenomenal*?) and social.

Analysing these phenomena presupposes recognising them: what interactions of organised matter can be considered cognitive and what can not be considered cognitive? Is there a test to determine whether a system is cognitive? What is the scope of the term "*cognition*"? Are there degrees of cognition? Or hierarchies of cognitive faculties? (See also our discussion in Section 2.)

What kind of structure admits what kind of interaction? What causes the interactions? What happens inside their physical substrate? In what way does it change and why? What are necessary and what are sufficient conditions on an organisation of matter for cognition to happen? What brings about an organisation of matter that is capable of supporting processes deemed cognitive? What are its characteristics? How did it and does it evolve? What makes it *remember* and *learn* and how does learning become manifest?

Analysis extends to these and many of the questions raised in the preceding sections but is of course limited to what is observable, measurable and describable. It is indeed instrumental in understanding the processes involved in creating and maintaining the physical and functional structures of a living organism; processes that cater to the evaluation of states both external and internal to a given structure, and to triggering actions with internal or external effect. (In short, it is instrumental in *understanding how cognition in living entities work*s.)

One of the most challenging quests is, no doubt, for the physical (*neural*) correlates (or rather: basis) of human and animal *consciousness* ([Koch]), and for the features that distinguish human from animal consciousness. What happens in a human brain when it becomes aware of its own existence? What enables Man to create and operate on sophisticated external symbolic and shareable representations of his mental states and processes?

In tackling this host of problems one has to have recourse to methods and techniques that are being developed in fields such as chemistry, biochemistry, molecular and cell biology, neuroscience, ethology, psychology, linguistics and even sociology. It seems to be a daunting task and yet, many threads already exist within these disciplines that may ultimately make up a coherent fabric.

Scientific publications abound that pertain to cognition related issues within these disciplines.

The neurosciences seem to be of key importance. They address the central nervous system (CNS) of higher life forms and hence a central layer in the overall architecture of life. Brain imaging methods for instance (fMRI, MEG, etc.), do contribute in no small measure to our understanding of the correlation between neural processes and cognitive behaviour (i.e., behaviour an external observer would interpret as based on the recognition and evaluation of what exists and happens, and on the prediction of changes in an animal's environment and body).

Studying the layers underneath should yield no less important insights into the ways neural processes come about in the first place (e.g., through processes of *self-organisation* as opposed to organisation by design), and into the evolution of CNS's and their capabilities. Lastly, attention must be paid to the layers above, the traditional playing field of psychologists: the various forms of cognitive competencies and behaviour (such as spatial and temporal orientation, language and proto-language, reasoning, communication, and so on), their development (in animals, infants and children) through growth and learning, their social and cultural dimensions where these exist, their effects on the environment and their being affected by the environment.

5.2 Modelling

The term *"model"* has different meanings in different contexts. Here we use it as shorthand for any explicit and communicable representation of the knowledge gained through analysis. Analysis goes hand in hand with modelling. It is in fact *guided by models*, by some pre-existing and then incrementally updated knowledge. The questions we posed in the previous subsection (and further above) could not even be formulated without some idea of what it is that we want to find out. As in every scientific discipline progress in studying cognitive phenomena comes about through a potentially infinite cycle of theories, discoveries, experiments and observations. The models emerging from this cycle must of course be implementable, in order to be of any value for creating artificial cognitive systems, as will be further discussed in the next subsection.

Models - not only of natural cognition - may be characterised for instance as plain descriptive, flat, hierarchical, discrete (symbolic, digital), continuous (dynamic, analogue), deterministic or probabilistic. Their formal apparatus draws on many branches of mathematics and theoretical computer science (which may in fact be considered a branch of mathematics). It includes differential equations, probability and statistics, recursive functions and complexity theory, automata theory and formal languages, and formal semantics - to name but a few. Approaches to modelling can differ in many ways: in their degree of abstraction, their generality and granularity; in the way they reflect the discrete and the

continuous aspects of what is to be modelled; in their use of formal frameworks; and in the explanatory and predictive power of their underlying theories (their ability to produce falsifiable hypotheses, for example). As in the case of analysis these approaches pertain to as many disciplines as are concerned (individually or jointly) with the study of cognitive phenomena.

Apart from these rather general characteristics, models for cognition may be classified according to the modeller's stance. The set of concepts underlying a model, or the interpretation of a particular term, may indeed vary greatly, depending on the *"school of thought"* the modeller adheres to. *Computationalism* (sometimes dubbed *cognitivism*), *connectionism* and *enactivism* are popular labels attached to these schools (occasionally, the latter two are subsumed under *emergentism*). Whitaker ([Whitaker]) succinctly (but with a bias towards *enactivism*) describes some of their distinguishing features.

In a nutshell, a traditional *computationalist* tends to understand cognition in terms of symbol manipulation based on data structures which serve as *explicit* representations (perhaps better: descriptions) of the aspects of the world a system has to deal with (... very much like an array of random access memory cells can represent a chess board). Hence, from his perspective, a cognitive system is "simply" a symbol manipulation system whose operation can be specified by formulating programmes, executable on an abstract machine for which there may be many possible physical instantiations (including pen, paper, the hand and the brain of a suitably drilled clerk who need not have the slightest understanding of what he is doing[36]). The meaning of the symbols manipulated by such a machine is not intrinsic to the machine but subject to external interpretation, in line with the designer-programmer's view of the world at issue. They are not linked to processes which impinge on the quality of - or perhaps threaten - the machine's functioning, and to which the machine can of its own accord respond in a self-sustaining manner. In other words, these symbols are not of any "value" to the machine itself. (For an illustration of the difference between intrinsic and extrinsic semantics see Table 3, on page 31). Pure *computationalism* tends to ignore the biological origins and groundings of our cognitive faculties and turns cognition into a kind of abstract game which can be played entirely detached from its "original board". A *computationalist* model of cognition captures hardly any of the essential qualities of natural cognition. Rather, it largely draws on mathematical logic, algorithmic theorem proving techniques, methods originating in Operations Research (e.g., game theory) and the like, dressed in a large variety of formal frameworks (see for example [AndersonJR] for an architecture called ACT-R, and [Newell2] for SOAR - State, Operator And Result). We have discussed this approach (which seems to be frequently but, we believe, wrongly

[36] In fact, the famous formalism Turing proposed models the suitably drilled but unwitting clerk.

The following ideated scenarios may further illustrate the distinction between *intrinsic* (i.e., internally generated) and *extrinsic* (i.e., externally imposed) semantics of a systems's inner and outer world:
When I look (e.g., through an electron microscope) at a **living cell** all I can possibly see (if I am lucky) is some matter (all sorts of radicals) going back and forth through the cell's membranes and swirled around in the cell's interior.
When I look at a **computer** I can see characters and pixels flickering up on a screen following key strokes, mouse clicks or the reading of a punched card. (In the old days - say, the fifties to seventies, of the previous century - one could even catch a glimpse of some representation of what happened inside - again through little lights turned on and off in seemingly random patterns.)
Computer: the characters / pixels / lights don't mean anything to the computer - they don't do anything for its wellbeing; but they may mean a lot to me (numbers, text, account statements and what have you) or whoever programmed the computer to produce its output. So, indeed, the computer does process what is information to me.
Living cell: the molecules diffusing through the cell's membrane and reacting in the cell's interior mean a lot to the cell - they keep it *alive*; they don't mean anything to me - apart from telling me that there are certain processes going on which do not seem to be completely random. I can analyse these processes, give names to the ingredients involved, draw neat diagrams representing the results of my analysis (molecular pathways for instance), and so on; I can create information based on what I see and analyse, and pass that information on to a colleague, etc.; however, that information is completely irrelevant to the cell at issue. The cell simply does what it does.
Computer: the situation may be somewhat different if the computer also receives input from, say, a temperature sensor. In that case it could turn itself off if it gets too hot, or reduce its power consumption through some change of state. So, whatever it gets from the sensor means something to it - it is (literally:) *vital* information (in that vague sense). But as in the case of the living cell, it does not mean anything to me - unless I have been the programmer of that particular behaviour. Depending on its programme it could even *learn* how to behave in like situations. However, this would have to be arranged quite carefully lest this process will be too costly: the computer could be rendered useless through overheating before it can apply the correct behaviour. (Not being the programmer I could again analyse what I observe what the computer is doing and if I am lucky I might even surmise the underlying programme or learning mechanism (just as I could surmise the structure of the catalytic cycles in the living cell) and then produce one myself (play *God*?) - for the benefit of my computer.)

Table 3: Intrinsic versus extrinsic semantics

equated to *Artificial Intelligence*, *AI*) at some length in section 3. While inspired by and supposed to address the so called higher cognitive functions (reasoning, planning, etc.) it meets with certain limits in precisely these domains[37] and seems even more difficult to extend fully to lower level cognitive tasks, such as spatial orientation and navigation, or motor control.

Connectionism has become popular partly in reaction to these shortcomings. *Artificial Neural Networks* (ANN) are among its chief assets. The very term suggests conceptual proximity to natural neural networks occuring in animals and in particular, the human brain. An ANN is composed of arrays of nodes, arranged in layers, of connections between these nodes, connection weights, and rules governing the propagation of signals (with excitatory or inhibitory effect on nodes) along connections. Proponents of *connectionism* contend that ANNs - while still a far cry from "real life" - lend themselves to modelling presumably basic cognitive processes (e.g., pattern recognition, object classification and categorisation) more faithfully and, in a final analysis, more effectively than any abstract symbol manipulation device a computationalist could dream of. Representations in ANN models are often referred to as *sub-symbolic*. As classifiers for instance, ANNs do not recognise objects in terms of particular explicit shapes and explicit association tables. An ANN keeps its "knowledge of its world" *implicit* in the weights (or "strengths") of the connections between its nodes, as brains presumably do. It is *trained* rather than programmed. Through training it can change the weights of its connections according to predefined schemes and thereby gain new or revise existing knowledge. It is, in concise terms, a structure that learns, an active memory (i.e., a memory and a processor at the same time).

Its nodes are rather crude simplifications of real neurons. They work like primitive processors with simple programmes and small storage space built in which implement the propagation and learning rules. ANNs can also deal with classical symbolic computation: logic gates for instance, can be conceived in ANN terms. However, ANNs, while *Turing-complete* when seen as formal computational "devices"[38], differ greatly from the classical *von Neumann* model. And they are usually not studied from a computationalist perspective. Many variants exist, depending on connectivity structures, propagation (including timing) and learning rules, and also on whether the nodes and the connections between them

[37] This is not only due to the computational complexity of formal symbol manipulation problems related to these functions. As noted towards the end of section 2: The fact that the human brain is capable of emulating discrete symbol manipulation does not allow to conclude that discrete symbol manipulating machines can emulate the human brain.

[38] A computational formalism is *Turing-complete* if it allows to specify the computation of any Turing-computable function, that is any function that is computable on a Turing machine ([Turing1]). Certain ANN models have been shown to have super-Turing capabilities ([Siegelmann]).

operate in digital (discrete) or analogue (continuous) mode. They are indeed special *parallel distributed processing* (*PDP*) architectures.

PDP architectures and connectionist models in general are not limited to abstracting from natural neural networks but also lend themselves to studying many kinds of natural or human-induced phenomena that can be described in terms of collections of elementary components whose interactions are subject to physical or man-made laws. These collections are often viewed as *dynamical systems* and numerous studies focus on explaining how local relations and interactions within such systems can give rise to highly ordered observable *global states* and behaviour (processes commonly referred to as *self-organisation* and *emergence*).

Like their computationalist counterparts, connectionist models and related studies may remain at an entirely abstract, game-like level. They may disregard biological reality and the grounding of cognitive functions in a body that needs them, and in an environment where that body is supposed to act successfully. It is nonetheless in the vicinity of *dynamical systems theory* (DST) where *connectionism* meets the *enactivist's* approach. *Enactivism* adds the physical environmental and bodily contexts ("*situatedness*") as crucial ingredients to understanding cognition and - ultimately - intelligence and consciousness, in animals and humans ([Clark1, Clark2]). It also adds a historical, evolutionary perspective which is largely absent from purely computational or connectionist models (and not covered by DST either). Its mantra, stemming from its biological (and, partly, philosophical) origins, is *embodiment*[39]. In spite of attempts at formalizing key concepts[40] underlying this approach no satisfactory physical theory (in the sense of [Pierce][41]) appears to exist that provides a testable explanation of the emergence of (especially *higher forms* of) cognition within an enactivist framework[42]. Much still is under more or less philosophical debate, occasionally prone to almost esoteric digression.

From an enactivist point of view cognition develops and is maintained through constant interaction of an entity with its environment. In a manner of speaking

[39] For an extensive discussion of *Embodied Cognition* see [AndersonML]. See also the discussion in section 3, regarding *environment* and *body*.

[40] ... such as *autopoiesis* and *structural coupling*, cf. footnote 13 (page 9) and the discussion in section 2. *Embodiment* is probably one of the hardest to formalise.

[41] Pierce distinguishes *mathematical* and *physical* theories. A mathematical theory may have either no or several largely distinct interpretations outside mathematics (examples abound: number theory, graph theory and branches of algebra, geometry and topology are only a few of many sources), whereas a physical theory (e.g., mechanics, electrodynamics, quantum theory etc.) is directly related to experience, experiment, and observation and allows for very narrow interpretations in "real worlds" only. This is not to say of course, that physical theories are devoid of mathematics. On the contrary, they can give rise to highly abstract mathematics (as exemplified by modern theoretical physics), corroborating Kant's dictum (cf. footnote 2).

[42] An interesting though perhaps not too widely accepted attempt at grounding mathematics in "the body" is [Lakoff].

the entity enacts (or *constructs*) its knowledge of its world, based on its phylogenetic and morphogenetic memories and within the constraints of its bodily structures and functions: its (changing) body in its entirety[43] *is* the *(dynamic) representation* of its world. It is, however, so far only at the lowest level of biological systems where this idea[44] becomes most concrete. Disciplines such as *Systems Biology, Theoretical Biology* and their "engineering companions" *Control Theory* and (more recently) *Artificial Life* (ALife) contribute probably most to this effect. ALife deals with a variety of "low level" models of which *cellular automata* (CA) were among the earliest ([von Neumann], [Burks]); they still are (e.g., [Rocha2]) popular objects of investigation, lending themselves to rigorous mathematical treatment. A rich mathematical framework has also been created around the notion of *adaptation* through, for instance, *genetic* and - more generally - *evolutionary* processes ([Holland]). Likewise, there have been attempts at formalising the concept of autopoiesis ([Beer1], [McMullin], [Nomura], [Wiedermann] and in particular, [Rosen1] and [Rosen2]), introduced by Maturana and Varela ([Maturana]). [Rocha1] presents an interesting view on symbolic codes as a prerequisite of *syntactic autonomy*, self-organisation and memory based evolution. However, from here it is a long way towards an adequate understanding of cognitive processes in higher forms of life, let alone in humans, according to enactivist or *constructivist* paradigms.

The three "schools of thought" whose boundaries are not as clear cut as it may seem, have different historical roots: most notably in the foundations of mathematics (computationalism), in neurophysiology (connectionism), in psychology and phenomenology (as a philosophical current, cf. for instance [Moran]) (enactivism). Each has its own traditions, none should be taken for granted, but none should be outright discarded as irrelevant or inappropriate, either. All have received varying attention at different times.[45] All must address in one way or another at least a subset of the requirements actionable models of cognition and cognitive systems are expected to meet.

Information, interpretation, *representation* and *indetermination* are probably among the most fundamental concepts to be covered by these models, regardless of any particular "philosophical stance". A system's cognitive processes,

[43] ...that is: the body and everything that is happening in and through it.

[44] ... or (obvious, trivial?) insight which had only been obscured by a thousands of years old *dualist* tradition in western thought and philosophy?

[45] Cybernetics, a discipline created in the fourties and fifties of the previous century by Wiener ([Wiener]), von Neumann, Ashby, and others, has been an early attempt to provide a unifying view on a number of issues that concern us here, among which are: autonomous and intelligent control, self-organisation ([Ashby1]) and neural architectures([Ashby2]). For multiple reasons (left to historians of science to explain) and perhaps lamentably, Cybernetics lost its impetus as a "unifying force". However, although it did not bring about a more or less coherent scientific community it did influence greatly subsequent developments in constituent fields and spawned many new developments at the interface of science, engineering and even the humanities.

by "definition" (cf., section 2), and no matter how they are physically instantiated, do interact with processes in their environment. They are *formed* by their environment and that which forms them may be called *information*. Conversely, what is and what is not information relative to a process or system of processes depends on the very structure of that process or system in question, on its interpretive capabilities[46]. Information (that which is *outside* the system) and representation (that which is *inside* or *part of* the system / that which the system gets to "know" or make of its world) are inseparably intertwined through *interpretation*: the act of turning information into representation and behaviour, which is itself contingent on representation and behaviour. Models must account for the forms and functions of representations (in the most general sense of that term) in cognitive systems but they must not stop there (at the *product* of cognition so to speak). They must also account for some way of establishing *meaningful* (or *useful*?) *relations* between "internal" structures, functions and processes on the one hand, and the "outside" world on the other hand (the *production* of cognition: *intrinsic semantics*). (Ultimately, this may lead to an explanation of the *aboutness* - or *intentionality* - of conscious experience.) Last but not least a model must address ways the processes of cognition cope with inherent and environmental *indetermination*, the uncertainty about *"what is going to happen next"* in the environment and "the body". Provisions for *self-modification* (which includes evolution, growth, adaptation and all (other ... ?) forms of learning) must be at centre stage in any such model.

Closely related to this is the equally fundamental *Action Selection problem*: *"What should an agent do next?"*. [BromBryson] contains a brief discussion of current approaches to solving this problem under the premises of different modelling "philosophies".

[46] Note that in keeping with the scenarios of table 3, this usage of "*information*" has little if anything to do with its usage in the context of statistical information and communication theory (as expounded in [Pierce]). We denote as "information" that which a cognitive system's own structures *resonate* to. [Freeman] illustrates this understanding in terms of brain functions, linking information to *meaning*. Here is an example in terms of human knowledge: *What is and what is not information to someone is largely determined by the knowledge that person already has; to teach someone the Pythagorean theorem who has not previously studied certain properties of triangles is a rather futile exercise; she cannot make sense of it.* In other words: *information* is a relative concept. Gregory Bateson ([Bateson]) characterises it as "a difference that makes a difference". Indeed, it is the difference (*"change"*) in the environment that gives rise to a change (*"difference"*) of the system's state and to the extent the system *can* actively or reactively change its state. (Whatever happens around a rock it has no way to respond - actively or reactively - of its own accord.)

This is not to say that the *statistical concept* of information is not relevant in the context of cognitive systems. On the contrary: it quantifies - at least in principle - the 'order' such systems can "extract from the environment" (cf., our Schrödinger quote in footnote 8, page 7).

For a more general discussion of the concept of *information* see for instance [Floridi] and, in the context of control systems, [Sloman1].

One of the key questions appears to be: "Given that real bodies have evolved and developed cognitive functionalities through real brains, to what extent can these bodies and brains be virtualised while still implementing the same functionalities or a subset thereof?" Or, to put it differently: "What are the structural and functional properties of matter that make these functionalities possible?", thus closing the analysis - modelling loop.

It is questionable (and indeed questioned by a growing number of researchers) whether the traditional information processing paradigms that underly the workings of most of our current electronic computing devices suffice to deal with this demanding agenda and to capture the ways natural organisms incorporate their environment and act in it. Information processing as invented by humans imposes artificial rules on matter. Although exploiting natural properties it (literally) imprints - in a coercive manner - non-natural structures (such as transistors and switching circuits) on physical substrates. By contrast, information processing in natural organisms (including animals and humans) is driven by physical laws, at greater or lesser degrees of freedom, but not arbitrarily. Hence the need for a *physical theory of cognition* (or "cognitive information processing"), which - as pointed out in footnote 41 (page 33) - will most certainly involve highly sophisticated mathematics. It may perhaps become part of an evolving *Science of Organisation*, that picks up on the early attempts of Cybernetics (cf., footnote 45, page 34) to provide a unified framework covering not only the general principles underlying the structures and functions of living organisms but those of technical and socio-technical artefacts as well - although this may be a futile dream (yet apparently dreamed by many[47]). However, such a theory could also inform us on whatever theoretical or practical impediments and limitations there may be, to the realisation of artificial cognitive systems.

In section 6.3.2 we shall come back on some of the above discussed issues.

5.3 Synthesis

Scientific theories give us clues - in terms of abstract models - as to how some process or system of processes in nature might work. But they are of little value if we cannot test them. Trying them out can take many different forms. In physics, arguably the most basic of the natural sciences, we design experiments that have to meet high standards, concerning for instance their replicability and their yield of data. The latter may either support (but not confirm) a given theory (and provide key parameter values), or prove it inadequate or plain wrong. The same holds - *cum grano salis* - for chemistry (where more "constructive" approaches are common). Similarly stringent requirements are either much more difficult

[47] Much of the work undertaken at the Santa Fe Institute (`http://www.santafe.edu/`) (for instance under the heading of *Complex Adaptive Systems*) points in that direction.

or impossible to satisfy in the sciences of the living. The difficulty increases as we ascend through the ranks of the animal kingdom. A Theory of Cognition as postulated above, would in this regard certainly be at the top end of the scale. How can we try it out? And - of equal import - how can we put it to good use (as discussed in section 3), in line with our prime objective of engineering artificial cognitive systems?

In physics, experiments can be more generally understood as implementations of abstract theoretical models of real-world phenomena, based on concrete devices whose behaviour can be controlled, watched and measured. These implementations require specific knowledge, techniques and technologies, ancillary to the theorist's intent proper, of explaining a natural phenomenon.

This observation carries over to any scientific discipline that aims to provide theoretically well-founded specifications for engineering artificial systems. Hence it applies to the subject matter of this note. The question to pose here is: *what do we need*

- *to know (beyond any actual Theory of Cognition),*
- *to be capable of doing (in terms of skills and techniques) and*
- *to have at our disposal (in terms of technologies, materials, tools),*

in order to turn specifications derived from abstract theoretical models of cognition into working physical models that demonstrate the adequacy and usefulness of our theories?[48] It may be addressed for instance from the perspectives of the "schools of thought" introduced in the previous subsection.

As a consequence of his *cognition=computation* worldview, all the pure computationalist presumably needs in order to transform his algorithms into operational systems, are a sufficiently powerful digital computer with appropriate input-output interfaces, and a well equipped software development workbench. The additional knowledge needed to implement his algorithms and make his machines "cognitive" would fall squarely within the remit of Software Engineering.

However, putting it that bluntly may be somewhat unfair. Nothing prevents the computationalist from employing for example, advanced machine learning techniques, self-modifying code, or highly interactive (and sensitive) interfaces. Yet, his implementations are ultimately constrained by the formalisms he imposes on his target hardware[49]. They are subject to the well-known limitations of formal systems, evidenced by the classical meta-mathematical incompleteness and undecidability theorems ([Gödel], [Turing1]), by results on computational complexity, and last but not least and perhaps more importantly, by the problem

[48] ... and we may add: eventually also into useful and commercially viable products.
[49] ... such as fixed length *registers* and *memory cells*, and operations on their contents, expressible in some some *machine language*.

of imbuing such systems with "robust" real-world semantics. In other words: the computationalist's physical models, not being based on a *physical*[50] theory of cognition, contribute little if anything to our understanding of natural cognition. This need not particularly worry us though, at least not at first glance, given our prime objective.

Moreover, there is merit to the argument that no matter what *physical* theory we adopt, its implementation must, in a final analysis, rely on an *artificial* manufacturing process. Or can organic behaviour be achieved through non-organic methods? Can artificially produced physical components be made to interact, based on their physical properties, in such a way as to exhibit behaviour that we would consider intelligent or the result of some cognitive process?

Of course, *simulating* physical phenomena, as specified by some theory explaining them, is always possible providing that theory allows for discrete (i.e., digitally representable and manipulable) approximations: we have computer simulations of physics experiments (e.g., nuclear explosions), stars, tomorrow's weather, transportation systems, ecologies and economies, protein foldings and the flow of ions through synapses. In fact, computer simulation has become a popular (and reasonably powerful) way of "trying out" theories, not only in the natural but also the social sciences, and in engineering. In many disciplines it helps to get around the difficulty of setting up replicable experiments - be it because they are too expensive, unethical, or simply impossible. However, nobody would deny that, whatever we simulate, the simulation is not the same thing as what is being simulated: simulating a nuclear test produces no harmful fallout, simulating a car crash does not kill anybody and simulating a star does not make it emit any x-rays into the real universe.

So what about simulating in a(n array or some other collection of) digital computer(s) the *physical processes* underlying natural cognition, assuming these processes (1) are well understood in terms of a mathematical modelling framework and (2) can be suitably discretised? Continuing the above line of reasoning we should conclude that while our simulation may well contribute to our understanding of natural and perhaps human cognition, it cannot have any effect whatsoever on the world outside the computer(s) on which it is running[51].

[50] In the sense for instance *mechanics*, *thermodynamics* and *electrodynamics* are *physical* - not *mathematical* - theories (cf. footnote 41, page 33).

[51] Stanislaw Lem's amusing and intriguing satire "*Non Serviam*" [Lem] is worth reading. It is about simulating a world inside a computer. Its inhabitants start philosophising about Good and Evil in their world. Their lofty discussion comes to an end only when the computer operator pulls the plug. Although some people seem to maintain that "*God is a computer programmer*" (e.g., [Chaitin]) we may - if only for practical purposes - assume that we are not living in such a world, trying to come to grips with the intricacies of cognition. (The same theme has been picked up on time and again in popular science fiction writing - e.g., Galouye's *Counterfeit World* - and filming - e.g., *The Matrix*.)

However, we already know that this is not quite so: the connectionist's ANNs are (as noted in the previous subsection) rather crude yet at least partial models of some of the natural structures and functions that determine and support cognitive processes in brains. A classical von Neumann computer can be programmed in such a way as to be functionally equivalent to a given ANN that satisfies suitable conditions[52]. From the brain researcher's modelling perspective this implementation could count as a simulation were it not for the concomitant behaviour which *does* have an effect on the world outside the computer. This type of implementation is in fact known as "*emulation* of some computational formalism (or: information processing system) by some other computational formalism (or: information processing system)", a technique that is frequently used in computer engineering to make - usually at the expense of efficiency - a given operating system or hardware behave like some other operating system or hardware.

What then would, in a nutshell, be the connectionist's agenda? First of all (and this should almost go without saying), his abstract models (ANNs) have to match their biological counterparts (brains) as precisely as possible. They should for instance reflect not only neural but also hormonal activity. He discards of course the option of taking the real thing for its own model. If he did he would be done. Hence, secondly, he implements this model on some hardware other than the "hardware" the brain is made of. He thus emulates the structures and functions of the brain at least to the extent he knows and understands them. One of his main problems consists in finding the most suitable machines to support the emulation, for instance in terms of efficiency. Classical computers may not be the best choice and there are indeed a number of other choices for implementing ANNs, based for instance on parallel hardware that is reconfigurable at run-time (e.g., Field Programmable Gate Arrays - FPGAs, Field Programmable analogue Arrays - FPAAs, or Cellular Neural/Nonlinear Networks - CNN which can also be emulated on FPGAs). Further advances (e.g., in molecular electronics [Heath], nanocomputing [Beckett]), merging processing and memory, are already on the horizon. Research and development along these directions are well under way.

Philosophical subtleties notwithstanding, it is an open question whether or not the differences between the approaches of the computationalist and the connectionist respectively, are that dramatic when it comes to the actual implementation of their designs. There is no doubt that the connectionist has to compute although evidently, his computations are organised differently. And there seems to be a drift of connectionist implementations towards exploiting directly - and not just emulating - the capabilities of certain materials to self-organise in line with their physical properties. Depending on the refinement of their underlying

[52] Both, idealised *von Neumann* computers and ANNs, are Turing-complete; however, we already noted that one can define classes of ANNs that are provably "super-Turing" (cf. footnote 38, page 32, and Section 6.3 below).

abstract models connectionist implementations may be significant contributions towards our understanding of biological neural networks. They seem to be a long way, however, from realising a broad spectrum of brain functions. This may simply be due to the fact that brains are unique and there appears to be no way to fully implement a brain through means other than those employed by Nature herself.

But - *mutatis mutandis* - we repeat: this need not particularly worry us, given our prime objective. It is not our primary objective to be able to build brains. Rather, we want to induce cognitive processes in artificial systems. To achieve this, ANN implementations may indeed be promising candidates, and they may be more promising the more brainlike they become (for instance in terms of plasticity or endocrinal controllability, see [Boahen]). Hence, insights into the way natural brains work may indeed be very relevant for designing and implementing *cognitive architectures* (thus closing the analysis - modelling - implementation loop)[53].

The enactivist's approach is arguably the most encompassing theoretically and - consequently - the most challenging as far as implementation is concerned. Enactivist models of cognition have become quite popular in guiding mathematics education ([Reid], see also footnote 42, page 33). Yet, like their connectionist counterparts and as noted in the previous subsection, they are still far from explaining the emergence of "higher" cognitive capabilities in such a way that it can be reproduced and made operational in an artificial setting. The following (interrelated) problems are presumably among the hardest the enactivist faces:

- to clarify the interplay between body, environment and the controlling agents (*brains*) in the context of technical systems ("What is a body, what is its environment?", see our discussion in section 3 and e.g., [Pfeifer1], [Beer2], [Clark2])[54], and

- to find ways of transforming evolution, growth and learning through interaction (i.e., autonomous structural and functional development, *self-or-*

[53] There is a large body of research exploring the brain's role as a controller of bodily function. [Grush] gives an account of some of the salient aspects. The idea that designing architectures for artificial cognitive systems can greatly benefit from studying the brain also underlies a British research initiative presently known as *Grand Challenge 5, "Architecture of Brain and Mind - Integrating high level cognitive processes with brain mechanisms and functions in a working robot"* (`http://www.cs.bham.ac.uk/research/cogaff/gc/`).

[54] Much of this type of work also appears to happen in areas (and respective communities) such as *real-time systems control* (e.g., [Passino], [Axelsson]), *complex systems engineering* (e.g., [Portillo], [Bar-Yam], [Norman]) or *computer networking* (e.g., [Akyildiz], [Sifalakis]), that have so far been less strongly connected to Artificial Cognitive Systems research.

ganisation), into technically viable procedures (cf. for instance [Sipper] and [ThompsonA])[55].

The questions of "*where and how to start*" and "*how to control*" the emergence of cognition in such systems - providing it can happen (!) - are all-important. The former is often highlighted as the *minimal architecture* question. The latter is crucial because we must not admit of artificial systems developping capacities beyond our control.

The above mentioned disciplines *Artificial Life* and *Control Theory*, new hardware technologies[56] allowing for direct support of (artificial yet physical versions of) evolution, growth, memory and learning, and of course the overarching field of Machine Learning (see section 6.3.1) in all its facets, could be or become providers of robust and resilient technical solutions, exhibiting some of the desirable qualities of natural cognitive systems.[57]

If successful the enactivist approach may indeed be the best way of meeting the requirements on artificial systems discussed in section 3. To describe it as "most encompassing" is not only justified by the fact that it endeavours to address most comprehensively the physical situatedness of cognitive entities. It is also the most *synthetic* approach: in its implementations at least it can build on insights gained and solutions achieved under the connectionist and computationalist research agendas[58]. ANNs for example, and classical symbolic computation may well be embedded in "enactivist implementations". The computationalist's cause may not be as lost as one might think it is. One should bear in mind that, after all, it has emerged from a natural substrate, the human brain[59].

Some of the above discussed issues will be taken up again in section 6.3.2.

---

[55] Evolutionary principles and strategies have long since been applied in various engineering disciplines, see for instance [Rechenberg].

[56] ... including those mentioned in subsection 5.2 as well as others for *molecular (organic) information processing*, informed by Systems Biology and Systems Chemistry (e.g., [Zauner]), but also new materials and sensor technologies. We shall come back on this in section 6.3.2.

[57] Note that we are leaving out entirely from this discussion the possibility - now becoming more and more obvious - of creating artificial systems through the targeted and controlled manipulation of natural ones, including manipulation at the lowest (nanoscale) level. Present day *Genetic Engineering* may only be a feeble beginning and ethically or politically motivated scruples, still felt by many, may not persist.

[58] A diagram reproduced in [Varela] (page 7) expresses this quite aptly in terms of concentric circles: cognitivism / computationalism - emergentism / connectionism - enactivism, from inside out, with numerous names (of researchers) put to these layers.

[59] A recently launched wiki-initiative entitled "Encyclopedia of Computational Intelligence" (`http://scholarpedia.org/article/Encyclopedia_of_Computational_Intelligence`) addresses many of the topics (and of course more) that have been discussed in this section.

6 A scientific foundation - building sites and builders

> *Aber der Erwachte, der Wissende sagt: Leib bin ich ganz und gar, und nichts außerdem; und Seele ist nur ein Wort für etwas am Leibe. Der Leib ist eine große Vernunft, eine Vielheit mit einem Sinne, ein Krieg und ein Frieden, eine Herde und ein Hirt. Werkzeug deines Leibes ist auch deine kleine Vernunft, mein Bruder, die du "Geist" nennst, ein kleines Werk und Spielzeug deiner großen Vernunft.*
> **Friedrich Nietzsche (1844-1900), in: Also sprach Zarathustra, 4**[60]

One may with some justification argue that laying the scientific foundation for creating artificial cognitive systems is an activity that has been going on for centuries. It has been spurred (and sometimes blurred) by more or less profound philosophical speculations and ruminations that have led to the notions of *mind* and *soul* a long time before artificial cognition became a subject of serious study. In the second half of the last century the qualifier "cognitive" has been apposed to a number of scientific, scholarly and engineering disciplines, to wit: psychology, neuroscience, neuropsychology, linguistics, anthropology, archeology, ergonomics and, more recently, robotics. The new field of *Cognitive Science* has evolved in the last sixty odd years and now draws extensively on several of these disciplines. Although no generally agreed upon canon seems to exist it does have a strong academic presence. Cognitive Science departments are proliferating, occasionally as an adjunct to or part of a larger Computer Science outfit, but more often under the roofs of Psychology and Philosophy. The field addresses many of the issues that form the subject matter of these notes. Its early development was greatly influenced by promises that arose from the optimism around ambitions pursued under the previously mentioned heading of *Artificial Intelligence (AI)* (see also footnote 26, page 15). It still seems to focus more on a "detached" understanding of what one could call the "high-level" cognitive capabilities enjoyed by humans, rather than on the grounding of such capabilities in more basic life (and body!) sustaining processes. In fact, a sizeable part of the Cognitive Science community had explicitly or tacitly, in one form or another, subscribed to the (computationalist's / cognitivist's) *cognition=computation* equation[61]. It must be acknowledged though, that especially in the last two decades the *corporate memory* of Cognitive Science has become increasingly aware of its roots in human psychology. The take-up of work on developmental psychology is but one

[60] "But the awakened one, the knowing one, saith: 'Body am I entirely, and nothing more; and soul is only the name of something in the body'. The body is a big sagacity, a plurality with one sense, a war and a peace, a flock and a shepherd. An instrument of thy body is also thy little sagacity, my brother, which thou callest 'spirit' - a little instrument and plaything of thy big sagacity." Friedrich Nietzsche (1844-1900), in: Thus spake Zarathustra

[61] Given that *Natural Intelligence* has been the primary focus of Cognitive Science this was probably a blatant misunderstanding of the AI agenda. To most people working on early AI (not necessarily Cognitive Scientists!) it must (or should) have been obvious (or at least highly plausible) that natural (human) intelligence could not be approximated by, let alone be the same as the operation of intelligently written computer programs.

example. For another example see footnote 42 (page 33). Likewise, the study of animal cognition, a descendant from ethology, now seems to fall within the remit of Cognitive Science. (See for instance [Gärdenfors], who tells the story of Cognitive Science, and [Wynne] for animal cognition.)

There is no doubt that the Cognitive Science community in all (or despite?) its diversity has to play a prominent role in advancing the scientific foundation for engineering artificial cognitive systems, in accumulating and structuring the knowledge needed to implement the practical items on the agenda outlined in this note. However, following the discussion in section 5 it is equally clear that this community cannot be the only one to partake in this endeavour.

In a nutshell, and given that we do not intend to manipulate *real life* (cf., footnote 57, page 41), the Artificial Cognitive Systems (ACS) agenda boils down to *building and furnishing a new Chinese Room* ([Searle], cf. footnote 26, page 15). In a way, this is a metaphorical characterisation of all the *"construction sites"* we are going to visit where scientists and engineers from different backgrounds (could) co-operate in strengthening the desired scientific foundation.

In Searle's original version of his argument against *"traditional AI"* the Chinese Room (perhaps a *Chinese Office*?) is a *syntactic system*, lined with formal logics, brittle and rigid, with structures *we* have to design down to the minutest detail and define what they mean. The general problem faced by the "builders" of the new room is to create a *semantic system* (perhaps a *Chinese Restaurant*?): a system which autonomously grounds its structures and functions in evolution, growth (or some artificial version thereof) and learning (for instance through action and interaction). It would be a system that understands *us* and hence our world(s), that is prone to statistical indeterminacy. That understanding would become manifest through the system's ability to "do the right things" of its own accord (i.e., act autonomously, yet in compliance with human ontologies, see Table 2, page 19). This may include deliberation and communication on human terms, and graceful adaptation to novel situations in its environment.

Searle's argument was instrumental in starting the *New AI* movement in the 1980s. The *New AI* agenda may well be considered a subset of the set of issues that are relevant for ACS research and development (see for instance [Pfeifer2]). Its potential applications cover a large part of the ACS territory demarcated in this note. It subsumes connectionist as well as enactivist approaches to modelling; it addresses embodiment as a key concept, and embraces Machine Learning (see section 6.3.1), artificial neural networks, non-monotonic logic and reasoning (i.e., reasoning that allows for revising conclusions in the light of new evidence, [Morgenstern]), Bayesian methods ([Griffiths]), Fuzzy Logic ([Nguyen]) and other ways of dealing with uncertainty and imprecision in the world at large. Among its earliest and still very popular objects of choice for experimentation and implementation are machines commonly known as *robots*.

The term robot originated in early 20th century Science Fiction ([Capek]). It usually evokes the mental image of a metallic creature which looks like a caricature of a human being. We see it simply as a machine that can render specific services by handling all sorts of physical objects, based on information gained through sensors and/or various modes of interaction within its working environment. Typically (but not necessarily), such machines have limb-like movable parts (*actuators*) or are made to move as a whole (e.g., on wheels or by means of leg-like structures). They usually operate in real world (3D) space and time environments, natural ones or those created through human activity. Locomotion may indeed be essential for a fair number of services. Such machines are not only prime candidates for becoming cognitive in the sense of this note (and thus more robust and versatile, as discussed in section 3) but they are also challenging technical platforms for testing theoretical models of cognition. In section 6.1.1 we shall describe briefly some of the pertinent research directions.

There are of course other environments, where no direct handling of physical objects would be required but where cognitive agents of some sort would nevertheless be very welcome. These environments may be natural, entirely artifical or hybrid. They include the (virtual, multidimensional) *digital content and service spaces* (for instance the World Wide Web or parts thereof) that keep exuberantly growing as secondary structures on top of the ubiquitous communication networks. They also include the networks themselves as well as all kinds of *physical and overlaid infrastructures* that human societies depend on and where these agents would exert some control over what is happening. The physical support or instantiations of such agents may be distributed over wide areas. We shall look into these matters more closely in section 6.1.2.

In section 6.2 we shall present yet another view of ACS related research, from the perspective of *cognitive competencies*, classified roughly as *low-level* (section 6.2.1) and *high-level* (section 6.2.2). The former are deemed to be "in charge" of "keeping the system operational" by controlling the internal workings of the system's body, whereas the latter are supposed to enable the system to do its particular job in its world: to set and pursue goals, to communicate with and to control processes in whatever environment it is supposed to operate in. To do so it has to analyse, evaluate and interpret - to the best of its abilities - what exists and is happening in its world. It has to make its world explicit to itself and - in general - to other cognitive entities as well. (See also our remarks in section 2, page 6 and in section 5.1, page 29.) A crude analogy here is to reactive and proactive patterns of behaviour respectively, of natural organisms.

In section 6.3 we shall pick up on some of the questions and issues raised in sections 5.2 and 5.3 and attend to more generic work spanning (1) the domains outlined in section 6.1 and (2) the various cognitive competencies discussed in section 6.2. Our foci will be on formal theories of learning, on non-standard mod-

els of information processing and the possibility of implementing such models in non-standard physical substrates. Indeed, if ACS's are to meet our expectations, no matter on what platform or in what environment, they must be capable of learning in the most general sense of that term. Likewise, bridging the gap between low-level and high-level cognition may require new theoretical models of *emergence* as well as the means of implementing them. The exploration of both has only just begun.

In the following sections and subsections we shall attempt to identify and delineate in rather broad strokes, not only current research and its drivers but also some of the salient and realistic longer term research challenges. It is obviously not possible to establish a one-to-one correspondence between projects (or activities) and issues. A given project for instance may focus on a particular application domain or put a particular emphasis on certain scientific problems. To be successful though, it needs to take account of the wider context of its research questions.

It should also go without saying that our headings indicate only one of many possible ways to structure the contexts and directions of ACS research. Much of that research is carried out as part of more or less targeted, industrial, academic or mixed projects of which many receive public funding through national and international programmes. The European IST Programmes[62] and their predecessors (in particular ESPRIT[63]) have an impressive track record in that regard. As of this writing the *Cognitive Systems* "Strategic Objective" for instance[64], under the current European Framework Programme 6 (FP6), hosts more than 30 relevant projects. We shall draw on them to illustrate the ambitions and modes of co-operation between the different disciplines that are involved at different stages in the analysis-modelling-synthesis cycle sketched out in section 6. However, where appropriate we shall also refer to initiatives and activities not directly linked to these projects. There are indeed many, taken and undertaken by research institutions, research groups or individual researchers, and supported by industry and/or private and public funding agencies, worldwide.

6.1 Platforms and environments

6.1.1 Cognitive machines

In this section we use the term *machines* in a narrow sense: it denotes devices with a clearly discernible physical "body", designed and built to assist people in carrying out tasks that are physically strenuous, potentially dangerous, repetitive

[62] IST = *Information Society Technologies*, `http://cordis.europa.eu/ist/`

[63] `http://cordis.europa.eu/esprit`

[64] `http://cordis.europa.eu/ist/cognition`; from here on we refer to this initiative as *FP6-IST-CogSys*.

and tiring, or simply impossible to do without. These devices can be stationary or mobile, large or small, depending on the particular kind of service they are supposed to deliver. They usually handle and/or transport physical objects external to their bodies and, as indicated further above, are frequently referred to as *robots*. (This includes unmanned *autonomous* vehicles but appliances such as mobile phones, TV sets, laptop computers or video cameras do not necessarily belong to the category at issue here; they can be "cognitive" too, however, as we shall see further below.) The history of industrialisation could partly be written in terms of the progress achieved in building such time-and-motion saving machines (perhaps better known as more or less complex *tools*). Refining and improving their mechanics (grippers, artificial hands, exoskeletons and locomotion gear, etc.) and sensorial capacities (including non-standard ones that living organisms do not possess) has always been a major concern. To reduce the degree of human intervention (or interference?) in the operation of these machines, has been another persistent trend (as already pointed out implicitly in the introduction (section 1) to this note).

Ultimately, this means more than merely automating the completion of a task according to some preset rules (as for instance so called *numerically controlled* (NC) machine-tools do). It means that machines are able to take decisions *autonomously* (cf., Table 1) on how to proceed with a given task, should conditions arise that had not been foreseen when that task was defined. A (conceptually) simple example would be a roving robot that is supposed to retrieve some object from a distant place but on its way encounters obstacles which had not been anticipated by its human commander (or programmer!). Another scenario would involve a manipulator on the shop floor that need not be laboriously reprogrammed upon changing the objects it has to deal with, but that can learn what it can and should do with the new objects.

The industrially and economically motivated desire to make machines cognitive (in that sense) so that they can take over for instance manual labour in a reliable way, turns them into potentially rewarding platforms for ACS research. Endowed with rich acting and sensing capabilities they become suitable for testing theories of cognition that propose the interplay between perceiving and acting as the very basis for developing the cognitive faculties in living organisms. Much of the work under the above mentioned *New AI* label (also known as *embodied AI*) is in fact being done on such platforms. It has become a research area in its own right, known as Cognitive Robotics (a robotic creature named "Cog" being its - perhaps - most famous early representative, see [Brooks3]). As of this writing two large European projects[65] (under the FP6 IST programme

[65] The two projects are: COGNIRON (`http://www.cogniron.org`), supported under the *"Future and Emerging Technologies"* umbrella of the IST programme, and COSY (`http://www.cognitivesystems.org`) which is one of the 20 projects under FP6-IST-CogSys (`http://cordis.europa.eu/ist/cognition/projects.htm`). Ac-

2003-2006) are addressing a broad range of general aspects pertaining to this new area, whereas others are focussing on particular issues, methodologies or methods, and on creating specific robotic systems to demonstrate the viability of their approach. Further below we shall take a closer look at some of the latter projects.

The search for feasible architectures of cognitive machines has only just begun: for architectures that cater, preferably in an integrative manner, for the various modes of sensing (visual, auditory, haptic, olfactory, temperature, microwave, chemical, etc.) and acting (pointing, grasping, manipulating, roving, gesticulating, talking, etc.). Acting should be based on mechanisms and forms of reasoning that do not succumb to *combinatorial explosions* of state spaces, the *dimensionality curse* (in multi-dimensional representation spaces), and the *frame problem*[66] (see for instance [Hölldobler], [Pylyshyn], [Shanahan], [Ziemke]). Clearly, its scope is limited by what the machine's body *affords*. It therefore seems equally obvious that in order for a machine to be deemed cognitive it should (in some sense) not only be *aware* of the affordances of the objects it is supposed to handle but also of its own body's capabilities; it should (in some sense) *know* what it can do and why, and what it can not do and why not. This concerns not only the tools with which it may be equipped and the way the operation of these tools is organised but also its mechanisms for using externally available resources (for instance energy and - possibly - construction materials).

How can a machine (as described above) be made cognitive (in the described sense)? In light of the preceding discussions it is not surprising that there are a number of - partly competing, partly complementary - approaches towards that goal. In technical terms it amounts to providing the machine with some sort of control mechanism that makes it behave in the desired way (we will have to say more about this in sections 6.2.1 and 6.2.2).

---

cording to the brief of COGNIRON its scope is broad because *"... a cognitive robot companion, beyond the necessary functions for sensing, moving and acting will exhibit the cognitive capacities enabling it to focus its attention, to understand the spatial and dynamic structure of its environment, to interact with it, to exhibit a social behaviour and to communicate with other agents and with humans at the appropriate level of abstraction, according to a given context."* Consequently, the project will *"adress the issues of representations, understanding, reasoning, and learning mechanisms and interaction with humans and other agents."*

COSY expects *"... to produce well-documented implementations of a succession of increasingly sophisticated working systems demonstrating applications of parts of the theory, e.g. in a robot capable of performing a diverse collection of tasks in a variety of challenging scenarios, including various combinations of visual and other forms of perception, learning, reasoning, communication and goal formation."* (All quotes are from the above indicated websites.)

[66] Loosely speaking, this is the problem of focussing on what needs to be known (and represented) about a given environment and ignoring what can be safely ignored. For most real-world environments it is hardly possible to make this distinction at design time (cf., [Stork]).

At least partially based on evolutionary and developmental principles and procedures to construct such mechanisms, some of these approaches pursue an agenda that may very well follow from section 5. *Evolutionary Robotics*[67] ([Nolfi]) and *Developmental Robotics* ([Asada], [Lungarella], [Weng]) for instance, are currently two well identifiable and increasingly popular subdisciplines of Cognitive Robotics. The former is mainly inspired by the widely accepted *synthetic theory of (biological) evolution* (*phylogeny*) whereas the latter draws on *(developmental) psychology* (as influenced for example by Jean Piaget; see [Piaget]) and *developmental neuroscience* (*ontogeny* and *morphogenesis*).

It is not surprising either that suitable classes of artificial neural networks are among the preferred tools for implementing representations - for instance as self-organising maps (SOMs) - and robot controllers, under both paradigms. Evolutionary Robotics applies abstractions of Darwinian evolution to populations of ANNs, usually in some simulation environment, whereas "developmentalists" employ more direct methods (e.g. various kinds of learning, of which we will say more in 6.3.1), gleaned from theories of *child development*, to adapt a given ANN to its task.

As indicated elsewhere in this note some roboticists prefer to add arms, legs and a veritable head to their creatures to make their shape and movements resemble that of a human being or an animal. There are several possible motivations: (1) it is exciting; (2) it may be fun; (3) it may appeal to basic human instincts[68]; (4) it is an intriguing engineering feat; (5) it may be useful in studying specific issues related to human embodiment and human-robot interaction (HRI, see for instance [Breazeal]) . Whether this is a passing or lasting trend remains to be seen. For the time being and as far as *physical models of human bodies* are concerned, it characterises yet another subfield: *Humanoid Robotics*[69] (see for instance [Swinson]) which deals with the engineering challenge of producing artificial skin, limbs, joints, tendons and whatever gear might be needed to imitate a moving human body. In scientific terms, however, it is ancillary to Developmental Robotics by furnishing the platforms on which some development can take place.

Robot-Cub[70], a European project, funded under FP6-IST-CogSys, combines both, humanoid and developmental robotics, by providing such a platform and

[67] `http://www.evolutionaryrobotics.org/`

[68] Many Science Fiction stories and films confirm (1)-(3). Given (1)-(3), people- or pet-shaped machines have indeed caught the attention of the general public (especially for branding and marketing purposes). Honda's Asimo (`http://world.honda.com/ASIMO/`) and Sony's Aibo (`http://www.sony.net/Products/aibo/`) and Qrio (`http://www.sony.net/SonyInfo/QRIO/`) are well known examples. More recently PARO (`http://paro.jp/english/`) has made front-page stories, a robot that looks like a furry seal and which is claimed to have a therapeutic effect on mental stress patients.

[69] `http://www.humanoidrobotics.org/`

[70] Robotic Open-architecture Technology for Cognition, Understanding and Behaviours, `http://www.robotcub.org/`

(1) Discovering the manipulation abilities of its own body:

- Learning to control one's upper and lower body (crawling, bending the torso) to reach for targets.
- Learning to reach static targets.
- Learning to reach moving targets.
- Learning to balance in order to perform stable object manipulations when crawling or sitting.

(2) Discovering and representing the shape of objects:

- Learning to recognize and track visually static and moving targets.
- Discovering and representing object affordances (e.g. the use of tools).

(3) Recognizing manipulation abilities of others and relating those to one's own manipulation abilities:

- Learning to interpret and predict the gestures of others.
- Learning new motor skills and new object affordances by imitating manipulation tasks performed by others.
- Learning what to imitate and when to imitate others' gestures.

(4) Learning regulating interaction dynamics:

- Approach, avoidance, turn-taking, and social spaces.
- Learning to use gesture as a means of communication.

(5) Developing robot "personalities" via autobiographic memory based on interaction histories:

- Learning about meaningful events in the lifetime of the robot.
- Sharing memory (events) during interaction.

Table 4: Robot-Cub scenarios

experimenting with it, in line with the developmental paradigm. It is highly ambitious in that it comprises the construction of new hardware and software: the hardware will be a robot of the shape and approximate size of a two year old toddler (the "*iCub*") although it will - at least to begin with - only be able to sit and crawl but not walk. Within the project it will be used in a number of experimental scenarios (see Table 4, quoted from [Sandini]). They correspond roughly to cognitive competences infants acquire in the first months and years of their life, for instance: eye-head-hand co-ordination, bimanual co-operation, object affordances, interaction by imitation, elements of communication. Both hardware and software will be freely available to relevant research communities, to be employed in different settings.

Cognitive robotics projects are largely bio-*inspired* (in fact, if they were not then they would probably not deserve the attribute "cognitive"). They do, however, differ in their degree of bio*mimetics*. Macro-scale engineered robots simply do not enjoy the structural richness (*"complexity"*) and malleability which characterises the most basic living entities. Yet, macro-scale biomimetics is certainly possible to some extent: on the hardware side this concerns not only the above mentioned artificial limbs but also in particular the specific sensor technologies for vision and touch; on the software side we have architectures (i.e., modules, components and processes) derived from our knowledge of the anatomy and physiology of mammalian brains (brain areas and their activation, neural pathways, etc.).

Several projects, funded under the same initiative as Robot-Cub, develop and integrate these biomimetic features. ICEA[71] and Senso-pac[72] are prime examples. Senso-pac focusses on sensorimotor systems related to haptics. The robotic device used for demonstration purposes is equipped with a dexterous hand and touch sensors patterned after natural skin[73]. It learns to distinguish between different types of liquid substances contained in a vessel, by evaluating both touch and force data when holding and shaking that vessel. Its software architecture is based on the analysis and modelling of key brain areas involved in haptic cognition (sensory and motor cortices as well as the basal ganglia and the cerebellum). The ICEA project emphasises the role of emotions - understood as bioregulatory parameters - in governing behaviour, and draws extensively on models of processes active in pertinent brain areas (e.g., amygdala, hypothalamus, brain stem). For implementing these models it uses several robot platforms, of which one has rat-like sensing capabilities through artificial whiskers. These implementations also allow to investigate the connections between cognition, bio-regulation, self-preservation and autonomous energy management (see also section 6.2.1).

In view of this section's introductory remarks it is no surprise to see at the core of many cognitive robotics projects the idea of bootstrapping cognitive capabilities in a machine/robot, by linking in a cyclic manner, the acquisition of sensorial data (*"perception"*) and the execution of physical action[74]. Senso-Pac is one example. The key to achieving this is, of course, learning! In section 6.3.1 we shall pay due respect to the immense importance of artificial learning. Suffice it here to mention four other projects (of the same series as the above) that employ

[71] Integrating Cognition, Emotion and Autonomy, `http://www.iceaproject.eu/`

[72] Sensorimotor structuring of perception and action for emerging cognition, `ftp://ftp.cordis.europa.eu/pub/ist/docs/cognition/sensopac.pdf`

[73] *Electronic or artificial skin* (see for example [Wagner]) with sensing capacities that even go beyond touch, is currently a highly active research area. Several labs in Europe (e.g., [Butterfass], [Crowder]), Japan (e.g., [Someya]) and the US (e.g., `http://www.nuengr.unl.edu/Lab-Mesoscale-Engineering`) are working on it.

[74] ... an idea that presumably originates from developmental psychology.

different - more or less biomimetic - learning strategies and structures to make a robot execute certain actions on its own (i.e., *"autonomously"*).

Like Senso-Pac, Paco-Plus[75] integrates haptics and vision. Like Robot-Cub, it uses a robotic platform with anthropomorphic traits. The robot is supposed to learn and to learn to imitate what it can do with a given object. One of the planned test environments is a kitchen where it is supposed to fetch, carry, and handle items of interest. Through interaction with objects it produces implicit representations for which project participants have coined the term *Object Action Complex (OAC)*. The underlying assumption is that things become meaningful objects through the actions that can be performed on them, an assumption clearly reminiscent of object orientation paradigms in Software Engineering. The strategy applied to forming these OACs mixes reinforcement learning (maximising reward) and correlation based learning (minimising contingencies). OACs also serve as the basis for grounding elementary linguistic faculties - to be demonstrated by having the machine describe what it is doing and why. In section 6.2.2 we shall say more about human-robot and robot-robot communication in general.

COSPAL[76] also pursues the goal of grounding a machine's symbolic reasoning capabilities in an association of action and perception. The machine is equipped with a gripper. The test case is a simple, 3-dimensional shape sorter type of puzzle that small children can solve. This is a typical assembly task that could be tackled in a fairly straightforward manner through conventional programming based on hard-coded rules and representations. In contrast, COSPAL undertakes to develop a learning algorithm for a specific type of artificial neural network, dubbed *channel associative network* ([Forssén]). It is flexible in the sense of enabling the machine to adapt to any randomly generated instance of the puzzle, regardless of differences in appearance of its various elements. (This sort of capability would clearly befit the manipulator on the shop-floor, putting together changing parts to form a product.) The project's motto has been aptly phrased as "*action precedes perception*" ([Granlund]) or "*seeing by doing*".

All projects mentioned so far touch in one way or another on the concept of affordances as introduced by Gibson ([Gibson], see also footnote 34, page 21), albeit limited to specific (classes of) objects in a robot's world. Project MACS[77] revolves around this notion in somewhat greater generality. The MACS robot explores its entire environment for opportunities for action, depending on the robot's physical capabilities and current "mental" state (representing for instance goals and intentions). The test scenario is a closed room furnished with all sorts of

[75] Perception, Action & Cognition through Learning of Object-Action Complexes (`http://www.paco-plus.org`)

[76] Cognitive Systems using Perception-Action Learning (`http://www.cospal.org/`)

[77] Multi-sensory Autonomous Cognitive Systems Interacting with Dynamic Environments for Perceiving and Using Affordances (`http://www.macs-eu.org/`)

blocks (of different colour, shape, size, surface texture, etc.), beams on the floor, a ramp and some empty space where blocks can be moved to. The arrangement of blocks, beams, et cetera, is dynamic: it may be changed manually while the robot is at work. The robot's task is to discover and internalise the affordances proffered by this environment and to perform certain actions in it (like carrying blocks to the empty space provided). MACS addresses two salient problems, which are in fact also common to the above and many other cognitive robotics projects: (1) the design of a structure for representing affordances (initially MACS envisages a Simple Temporal Network (STN) representation, see [Dechter]); (2) to decide which affordances must be hard-coded, and what should be learned (a variant of the question *"where and how to start"*, raised on page 41). The learning schemes employed by MACS are versions of unsupervised and reinforcement learning.

The concept of environmental affordances is also implicit in GNOSYS[78], the fourth project from the current (mid-2006) IST Cognitive Systems portfolio (see footnote 65, page 46) to be considered here. The GNOSYS robot should be able to navigate safely in an unknown and unpredictably changing territory. It can see and hear, and measure distances (like a bat) as well as its angle from the vertical; it should be able to find out on its own how it can achieve a user-defined goal; for example if it cannot circumvent a pond it should find a plank to put across; it should move a moveable obstacle rather than go around it, et cetera. The project covers the entire chain from perception via knowledge acquisition, abstraction, and reasoning to action and back to perception, based on a largely (human) brain-inspired (and as such bio-mimetic) *cognitive architecture*, with modules implemented in terms of learning artificial neural networks (with reinforcement or Hebbian strategies). Learning also happens on the global level, mimicking child development, as modules "go online" one by one. One of the key ideas underlying GNOSYS is to understand perception as being controlled by *attention*. The project draws on a significant amount of work done by one of its partners on modelling attention as a fundamental control and learning mechanism in neural systems (see for instance [Taylor]).

We note that whatever the projects' approach and their focus on particular sensing and acting modalities, the general objective is always the same:

- to make a machine behave *sensibly*, even in adverse circumstances, independent of direct external control by a human operator;
- to make it adapt to novel situations without having to re-program it from scratch or even adjust a given set of parameters.

The approach is usually a mix of *"top down"* and *"bottom up"*, the former indicating deliberate (and presumably intelligent) design of the robot's body and

[78] An Abstraction Architecture for Cognitive Agents (`http://www.ics.forth.gr/gnosys/`)

its basic operating system, and the latter some sort of self-organising and self-modifying programme, data, and - possibly - microprocessing structure (as discussed in section 5.3). Usually, and in contrast to nature, evolutionary and developmental strategies in robotics are applied to software, data (or soft memory) structures and data sets, and run-time modifiable chips at best. So far, the evolution or growth of macro-scale robot bodies mainly happens in virtual environments (see also the following section), through abstract simulation of the corresponding natural processes (e.g., [Sims], [Bongard]). There are, however, noteworthy exceptions, for instance the work undertaken at the Cornell Computational Synthesis Lab[79] on self-reproducing and self-assembling machines ([Zykov], [White]). Another example is the work on distributed robotics as in project Swarm-Bots[80], which is inspired by the observation of swarming and colonising insects, such as ants and bees. Swarm-bots are collections of physically identical smaller robots that are endowed with communication and (physical) interconnection capabilities. Their behaviour can be subjected to artificial evolution, directed towards a high degree of self-organisation that enables them to carry out tasks jointly that a single "bot" could not do by itself (for instance lugging heavy objects, cf., [Dorigo], [Gross]).

Self-assembly and self-organisation are certainly household terms in micro-scale and still more so in nano-scale robotics (see for instance [Ummat], [Zheng]). It is not entirely clear - yet - to what extent this research will or can contribute to the creation of cognitive machines beyond, say, specific sensor technologies, for instance of the kind referred to in footnote 73 (page 50). Given that in most higher forms of life mental development is inextricably linked to physical growth one might indeed suspect that achieving robust cognition *in vitro* does require this link. We shall return to this issue in section 6.3.2.

The notion of cognitive machine as presented and discussed in this section is certainly an excellent candidate for formulating seemingly hard to meet but concrete targets which may not only spawn leading edge research on key aspects of real-world cognitive systems, but also kindle the interest of relevant industries, the military profession and the general public alike. The Grand Challenge launched by DARPA[81] is a case in point: to have a driverless car travel over a considerable distance across some unknown terrain, on its own, without direct remote control, that is. The race is an annual event, now (in 2006) in its third year, and the rules regarding autonomy or the use of satellite navigation systems are getting stricter every year.

There are of course many challenges imaginable beyond industrial robotics that are of less conspicuous dual use, yet equally competitive. They often re-

[79] CCSL (`http://ccsl.mae.cornell.edu/`)
[80] A European project concluded in 2005, see `http://www.swarm-bots.org/`
[81] (US) Defense Advanced Research Projects Agency (`http://www.darpa.mil/grandchallenge/`)

quire the integration of various sensing and acting modalities, normally present in natural organisms: most importantly perhaps, vision, motion and haptics, for navigation and object manipulation. *RoboCup*[82] is one of them, with the rather long-term aim to develop, by the year 2050, *"a team of fully autonomous humanoid robots that can win against the human world soccer champion team"*. While goals in those playing fields presumably have to be distant and perhaps elusive it is important to set intermediate milestones in terms of the platforms and cognitive competences needed in whatever environment at issue: what do we need to know to achieve these goals and what are the steps to take to make this knowledge operational? We suspect that answering these questions is largely a matter of *piecemeal engineering* rather than of grand design.

6.1.2 Artificial cognition in natural, artificial and hybrid worlds

As pointed out in section 3 the notion of environment encompasses more than we and our animal companions (or the machines of section 6.1.1) can cope with, using eyes, ears and limbs (or the artificial equivalents of these organs), without any technical support. More generally, an environment (a *"world"*) is determined by the (types of) signals and data an entity operating in it is fit to process[83]. Data can be of all sorts. While ultimately stemming from the "real world" they need not necessarily originate directly from sensing (organs or devices).

In section 3 we also explained in what sense we can talk about cognition in *generalised environments*, and what would be expected of a cognitive agent there. *Action*, in particular, may assume forms that are quite different from what robots or robotic devices are poised to do. It may, in fact, come down to an agent's handling external data without physically moving anything. Here we discuss briefly three classes of environments in terms of relevant data, (application) scenarios, and pertinent issues. Many of these issues are of a fairly general nature and will therefore be dealt with in more detail in section 6.2.

(1) There are natural environments not well matched by our senses, but which can nonetheless be perceived and acted upon through devices of our creation. There are for instance the data we generate through observation and experiment, involving matter at the very large and the very small scales: in astrophysics, high energy physics, molecular biology, geophysics, to name but a few areas of science where data sets are becoming so large that new approaches to analysing and interpretating them are called for. New approaches to acquiring these data in the first place may also be needed, for instance by endowing our scientific instruments

[82] `http://www.robocup.org/`, see also [Kitano]

[83] Or the perturbations that impinge on the entity. Strictly speaking, one must add material substances. So far, however, and in spite of paying heed to the importance of a physical body, there is little research on artificial cognitive systems (as for instance in project ICEA, see above) that takes account of the importance of ingesting and processing matter by natural cognitive systems.

with cognitive capabilities. This would be perfectly in line with the analysis-modelling cycle underlying scientific activity in many domains (as explained in the first paragraph of section 5.2). However, cognitive systems created for the express purpose of automating the scientific process, do not seem to be in the offing yet. But many tools do exist that support scientists in analysing and interpreting data generated through experiments and observation. They are largely based on Machine Learning approaches (see section 6.3.1). New ways of organising science have thus become possible (labelled *e-Science* in Great Britain[84]), evolving around the networking of data and processing resources world wide through so called Grids[85].

Cognitive capabilities would also be an asset in instruments used to diagnose and treat diseases of the human body. The whole human body is the most salient part of the environment such devices would have to interact with. To the human cognitive apparatus it is, in a way, the natural environment that emits the interoceptive stimuli which that very apparatus fails to categorise beyond the conscious perception of more or less localised pain. It is incapable of determining for example, whether that pain is caused by an infection or by some cancerous growth.

The desire to have better means of identifying and curing illnesses drives entire industries. Here, one of the basic problems is similar to the one we alluded to in the previous paragraph: to interpret correctly data gained from physical processes, metabolic or neural, and to suggest or - if possible - even apply appropriate remedies. People's interest in good health has long since been a prime motive to develop for example software that visualises and analyses data from electrocardiographs, electroencephalographs, ultrasound and x-ray machines, PET scanners, magnetic resonance tomographs and other diagnostic equipment. Sensors aasigned to all kinds of bodily functions yield a host of data that serve to monitor and control these functions. One of the earliest so called *expert systems*[86], MYCIN[87], had been designed to diagnose certain infectious diseases and recommend a therapy ([Shortliffe], [Buchanan]). Today, micro- and nanodevices, operating as intra-body sensors and actuators[88], are at the leading edge. They can for instance be used to precisely dose and target the delivery of drugs to specific organs or tissues (cf., [Staples]). However, as far as their standing

[84] `http://www.rcuk.ac.uk/escience/`

[85] `http://www.gridforum.org/`, `http://www.gridcomputing.org/`

[86] Expert systems, now considered classical "old AI" products, have became popular in the late 70's of the 20th century. Their principal components are some sort of *knowledge base* (usually set up in terms of factual assertions and rules) and an *inference engine*. A good entry point is `http://www.aaai.org/AITopics/html/expert.html`

[87] `http://smi-web.stanford.edu/projects/history.html`

[88] As investigated in project Mol-Switch (cf., [Firman] and `http://www.nanonet.org.uk/molswitch/`). The FP5-IST project BIOLOCH (Bio-mimetic Structures for Locomotion in the human body, `http://www.ics.forth.gr/bioloch/`) focused on designing endoscope carrying robotic devices operating in human bodies ([Menciassi]).

as *cognitive systems* is concerned our remark on page 53 regarding micro- and nanorobotics applies accordingly.

The *"whole person"* is also prominent in the environment of systems (or machines, or appliances, large or small) that people, individually or jointly, have to interact with closely in order to obtain some service. Ideally, such systems should tune in on the perceived needs of human users, to the extent even of anticipating their wishes. This requires precisely the cognitive capabilities that form the subject matter of this note: for instance the ability to understand natural language in terms of speech acts ([Searle2], [Winograd3]), or to interpret correctly gestures or facial expressions in terms of emotional states (see for instance [Sebe]), and to behave accordingly.

Almost an entire chapter of the 2003-2006 IST programme has been devoted to this kind of problems[89]. Here we mention but a few of the activities and projects supported under this chapter: HUMAINE[90] and ENACTIVE[91], two so called *Networks of Excellence*, federating research groups throughout Europe; and TC-STAR[92], one of the *Integrated Projects* in the area. HUMAINE's thematic agenda features the whole range of research that may lead to viable theories and models of emotion, pertinent to the design and implementation of interactive systems. This concerns for instance the extraction and classification of emotion related cues from user behaviour. The members of the ENACTIVE network share interest in studying the *"role of motor action for storing and acquiring knowledge"* (about users) in *"action driven interfaces"*[93] of robots and other systems; they focus inter alia on haptic technologies for manipulating objects in real and virtual spaces. The idea of joint action between man and machines also underlies the JAST project[94], which is part of the FP6-IST-CogSys portfolio (see footnote 64, page 45). Current research on human-machine interfaces also fits into the bigger picture of *ergonomics*, the discipline of engineering complex technical systems (not necessarily software controlled) that people can easily cope with. It addresses physical, operational, organisational, cognitive and emotional issues[95]. (For specifically cognitive aspects of ergonomics see for instance [Hollnagel] and [Woods].)

[89] *"Interfaces and Interactive Systems"* `http://cordis.europa.eu/ist/ic/index.html` and `http://cordis.europa.eu/ist/ic/projects.htm`

[90] Research on Emotions and Human-Machine Interaction (`http://emotion-research.net/`)

[91] Enactive interfaces (`https://www.enactivenetwork.org`)

[92] Technology and Corpora for Speech to Speech Translation (`http://www.tc-star.org`)

[93] `ftp://ftp.cordis.europa.eu/pub/ist/docs/ic/enactive-brochure_en.pdf`

[94] Joint-Action Science and Technology (`http://www.jast-net.gr`)

[95] In Japan that same discipline has been spiced with an almost philosophical ambition to bring about harmony in the relationship between people and technology. This ambition has given rise to a movement known as *Kansei* engineering (`http://www.jske.org`).

Project TC-STAR builds on a number of precursor attempts[96] to achieve real-time speech-to-speech translation in limited domains of discourse. Parliamentary debates constitute one of its test cases. It aims to extend these attempts to open-ended domains. It is not explicitly preoccupied with the formidable *cognitive dimensions* of translating from one natural language into another natural language but acknowledges the need for developing and applying sophisticated learning methods.

A third category of natural environments comprises those (more or less) immediately around us that we cannot see, hear, smell, taste or touch. Examples are: electromagnetic radiation (from natural sources) beyond the visible light, magnetic fields, barometric pressure, seismic activity, vibrations in the air beyond the audible spectrum, molecules our nostrils or our taste buds do not respond to, and objects that escape our grip. Of course, these environments do not differ in principle from our familiar ones. Just as there are animals that process ultrasound or the faintest traces of pheromones we can equip robots and other machines or systems, with suitably tuned sensors. A key problem with any artificial system operating in this kind of environment consists in communicating with humans about its findings and actions. Like any communication this presupposes an ontology, as explained in Table 2, page 19. This would be a conceptualisation of the world beyond our senses, but one that the machine would nevertheless have to be able to share with us. It is a problem that stimulates a certain amount of research (see [Shurville], [Modayil] and [Goodwin], three more or less representative examples) and which is tractable thanks above all to our well-known capacity for translating alien phenomena into symbolic (and usually visual) representations that can be related to the world as we experience it. We shall come back to this problem in a more general context, in section 6.2.2.

Finally, in certain natural environments, peopled or not, and for certain types of tasks we would welcome an extension of our natural senses (in particular, vision) and means for action, to wider spaces and for longer durations. Cases in point are environmental monitoring (for instance through remote sensing from space) and control, and the surveillance of traffic, buildings, or - increasingly popular[97] - public areas. Usually, these environments pose requirements our perception routinely has to meet, like finding out what is happening in the street, what people are doing, who they are, what they might have in mind, etcetera. With cameras and other sensors we can certainly cover larger areas and for any length of time. However, for a person it would be a rather tedious exercise to scrutinize the output of such devices. Hence the need for artificial systems that can take on that task. The system has to report the results of its inspections in

[96] One of them is *Verbmobil,* a German nationally funded project (`http://verbmobil.dfki.de`).

[97] The reasons are manifold and presumably not always reasonable. Discussing them would be beyond the scope of this note.

our terms - as in the case of the alien sensory worlds of the preceding paragraph. Again, this requires the kind of competences that we shall discuss in section 6.2.2.

(2) The second large class of environments where technical systems would significantly gain from being endowed with cognitive capabilities includes:

(a) man-made environments where people work together and with artificial objects (often: symbolic representations, such as documents of all sorts) and processes (often: manipulation of symbolic representations), for example, in offices and production facilities;

(b) physical infrastructures that need to be tightly controlled, for instance power plants, energy, road and other transportation networks, that contribute to the material wealth of our societies.

In many instances human-machine interaction (as discussed further above) is an important issue in these environments. More generally, however, they can be characterised as a fair mix of natural and artificial (i.e., man-made) data sources where the natural part (especially in type (a) environments) may include data arising from dealings among people. *"Socio-technical"* is a term frequently associated with such *hybrid* environments.

Apart from these commonalities, both subclasses, (a) and (b), are very heterogeneous indeed. There are type (a) environments for instance, where *software agents* mediate people-to-people communication and interaction via networks at various levels: from managing the physical and logical channels and other resources[98], to operating on contents and services in *virtual spaces* such as the World Wide Web. (Operations include: access to, linking, retrieving, filtering, composing, transforming and managing content.) In this context, the physical substrate of an agent can be the entire network or system, but also an appliance (a mobile phone, a "Personal Digital Assistant (PDA)", a personal computer (PC), or some other gadget) that serves to access the network. The agent itself, we recall, can be understood as a structured collection of processes, with a time invariant identity (as explained in section 2, page 8 and section 3, page 16).

Agents of this kind have been popular objects of research for many years (mostly under the heading *Intelligent Agents*, see for instance [Riecken]). Ideas about linking related portions of content in and across repositories actually pre-

[98] In wireless networks for example, a bundle of techniques known as *Cognitive Radio* (see [Mitola]) can be used to manage the dynamic allocation of radio frequencies and a range of parameters that determine the quality of service in relation to a user's needs and preferences. Packet switched techniques are being enhanced by so called *cognitive packets* (see [Gelenbe]), a concept also underlying at least one of the FP6 IST projects on *"Situated and Autonomic Communications"* (`http://cordis.europa.eu/ist/fet/comms-sy.htm` and `http://www.cascadas-project.org`).

date the existence and widespread use of computers, let alone computer networks[99].

Research on *content agents* in particular, has been following tracks that had been laid by classical Artificial Intelligence since its beginnings in the sixth decade of the 20th century (see section 5.2). The spaces in which these agents operate were supposed to be fairly structured and hence amenable to more "logical" approaches. With the explosive proliferation of digital content on the Internet, however, it became clear that these spaces are about as dynamic, changeable and non-deterministic as the real world itself, of which they provide representations, derived from real-world objects, events and processes. Just as physical agents in the real world (such as robots) need some understanding of what is happening around them, so do agents in a hybrid world have to have an understanding of the digital objects and processes that form (part of) their environment.

Usually, this boils down to establishing a formal ontology (see Table 2, page 19) as a basis for content manipulation and for action within the network and at the agent's interface to its users. The way this is done can mark the difference between a mere passive assistant and what we may call a *cognitive agent*. The former would include a *hard-coded* world model and rely on *being explicitly fed* all the information it needs to operate: its ontology would be externally supplied. In contrast, the latter would be capable of developing and continuously adapting its knowledge of its world (its ontology!) through autonomous exploration and analysis of its contents, and through interaction with human users. These are highly non-trivial problems already at the textual level ([Shamsfard]), let alone at the level of multi-modal objects involving digital recordings of (still and moving) imagery, sound and other types of signals. The realisation of a *Semantic Web*[100] effectively hinges on the solution of problems of this kind. We shall come back to them in section 6.2.2. There we shall also meet project CLASS which develops methods for the semantic annotation of multimedia content that make extensive use of Machine Learning techniques (see section 6.3.1).

Cognitive content and service agents can take on an artificial life of their own, as for example in the FP6-IST-CogSys project RASCALLI[101]. According to its authors *"Rascalli combine Internet-based perception, action, reasoning, learning,*

[99] Well known accounts of this are HG Wells' "World Brain" ([Wells]) and Vannevar Bush's "As we may think" ([Bush]).

[100] The term "Semantic Web" refers to a Web whose contents are enhanced by ontology-grounded metadata (see for example [Antoniou] or `http://infomesh.net/2001/swintro/`) that can be used by software agents to render content based as well as computational services. The idea has been around for some time but gained greater momentum only after the year 2000, when a joint EU/US group initiated the process of standardising a web ontology language (`http://www.daml.org/committee/`). Cognitive approaches seem to be called for in order to automate the process of creating both, ontologies and metadata.

[101] *"Responsive Artificial Situated Cognitive Agents Living and Learning on the Internet"*, `http://www.ofai.at/rascalli`

and communication." "They come into existence by creation through the users" who *"train them to fulfill specific tasks"*. The experimental scenarios initially foreseen are a quiz game (where Rascalli assist their masters) and a music portal (where they help retrieve and organise music related information). At the user interface Rascalli are represented as *avatars*. The project largely builds on ideas (of "How the Mind Works") formulated in Marvin Minsky's *"The Society of Mind"* ([Minsky], [Singh]) and developed further into a full-fledged *cognitive architecture* (dubbed *DUAL*) that comprises symbolic and connectionist elements (see section 5.2) and provides for the interplay of such elements ([Kokinov]). ([Hitzler] gives another interesting example of linking symbolic logic and connectionist models.)

Agency in type (b) environments is concerned with monitoring and controlling the processes that are characteristic of these environments, for example the material and energy flows occurring in them. This implies preventing hazardous situations, detecting and diagnosing anomalies wherever they may occur, and taking remedial action if necessary. Typical goals are: to keep road traffic going and distribution networks functioning; or: to maintain manufacturing and power plants in good working order and adjust their output to real needs; to inprove the efficiency of all processes involved, most notably their use of available energy resources. Many of the environments at issue are themselves technical systems operating in a larger context (as explained in section 3, page 16). Hence, from a very general perspective we are dealing with problems that are quite similar (or at least analogous) to the problem of making a robot (i.e., machine or vehicle as in section 6.1.1) do what we want it to do. They are usually dealt with under the heading of *Intelligent Control* ([Antsaklis]) which denotes an engineering discipline that takes up approaches from *Artificial Intelligence* and *Cybernetics* but has its actual roots in the much older tradition of "classical" *Control Theory* and *Control Systems*[102]. We postpone further discussion of pertinent aspects to section 6.2.1.

(3) Lastly, distributed computing systems, for whatever purpose and within whatever organisation they are used, can be considered environments to the components and subsystems they consist of. (One may add: just as an animal's body is environment to its cells and organs.) The more easily new components can be fitted in and the more readily existing components adapt to changing requirements, the less costly is the operation and maintenance of the system, and the more robust and flexible it becomes. This observation may indeed lead to an entirely new understanding of and approach to the engineering of large scale computer systems, impacting on the design and implementation of both hardware and software. The new approach would extend the long history of com-

[102] One of the earliest papers offering a rigorous treatment of a simple control system is James Clerk Maxwell's "On Governors" ([Maxwell]). Not surprisingly, this paper is also often quoted in accounts of the history of Cybernetics.

ing to terms with the problem of managing big hardware and software projects, a history marked by keywords such as: *software crisis*, *modularisation*, *object oriented design and programming*, *plug-and-play, Web and Grid service architectures*. It adds to these the concept of *autonomic component*, a term chosen under IBM's *Autonomic Computing* initiative to denote a new type of system constituent that can manage itself and its integration within the larger system, according to goals set by a human administrator[103]. It will have to raise the level of programming from the compilation of sequences of imperative, declarative and logical statements, to *general goal specification*. [Kephart] succinctly describes the engineering and research issues pertaining to this idea. They form an agenda that at least in abstract terms, overlaps significantly with what has been discussed so far in these notes.

In fact, *Autonomic Computing* systems - while not necessarily analogous to living systems - are supposed to share certain features with living systems; for instance many of the self-X properties listed on page 25. In that regard they can also appropriately be characterised as *Organic Computing Systems*. This is the name of yet another initiative, taken by a group of researchers in Germany[104]. It is potentially broader in scope in that it touches upon the full range of bio-inspired computing, without reference to any particular type of environment, and addresses not only the many problems underlying cognitive ("intelligent") control and cognitive robotics, but also for instance *cognitive middleware*, sensor networking, artificial immune systems and models of computing derived from (bio-)chemistry. Similar, equally broad initiatives, aiming at redefining computing and computer systems engineering, have been started elsewhere in Europe and abroad. We shall come back to some of these in section 6.3.2.

(4) Summing up this and the previous subsection, we distinguish at least four broad classes of environments:

Type1: *Common Sense* environments: with 3+1 dimensions, (to us) visible light, audible sounds, ..., *natural* or (in varying degrees) *civilized*, which may be populated by natural and artificial agents of all sorts; but: in such environments machines may reach, see, hear, smell, ..., and do, more or different things;

Type2: natural environments at various scales, not directly or fully accessible through our own (bodily) senses and actuators (for instance: our own bodies, the deep sea, outer space, etc.);

[103] `http://www-03.ibm.com/autonomic/`

[104] DFG Priority Programme *Organic Computing* (`http://www.organic-computing.de/spp`, see also `http://www.organic-computing.org/`)

Type3: *artificial* environments: external representations of Man's (and machines') perceptions and reflections in Type 1 and Type 2 environments; e.g., *Digital Content/Data spaces;*

Type4: (socio-) technical systems, large and small, embedded in Type 1 or Type 2 env ironments.

6.2 Core competencies

Whether or not a (natural or artificial) system possesses cognitive capabilities becomes apparent through its behaviour, through what it does and how it does what it does. For the sake of definiteness and in line with our understanding of cognition as a fundamentally biological phenomenon (see section 2), we postulate a strong correspondence between the complexity of the behaviour of an agent and the agent's physical and organisational structures which determine what it can do. Although we have to admit that *behavioural complexity* is not a very well defined notion we do contend that the existing varieties of living matter exhibit remarkable differences in the scope and nature of their actions. The behaviour of humans, individually and jointly, is arguably much richer and much more effective in bringing about change in their respective environments than, say, the behaviour of dogs which in turn is more involved than the behaviour of fish, which in turn ... . A similar distinction can be made in the domain of human mental development, linked to physical growth: Before a child talks it walks, and an adult usually differs greatly from a child in what he or she can do in the world. Cognitive competencies in living systems evolve and develop, and hence (or - perhaps: equivalently so) the potential of such systems for acting in their environments.

It is therefore reasonable to assume some sort of layering of the totality of cognitive functions, where functions on a lower layer are prerequisite to (implementing) those above. This is not unlike the situation with artificial software-hardware systems where layering principles have been guiding design since the early days of computer and software engineering[105].

But where can we draw a line between low level and high level functions? What is low, and what is high? Interestingly, a proposal made some 2350 years ago by Aristotle, the towering figure of ancient Greek philosophy, still seems to be popular and widely accepted. He puts it in terms of *faculties of the soul*,

[105] Examples: the structuring of computer software as Operating System plus application software where each of these components in turn consists of various layers (e.g., procedures implementing operations on various datatypes, type definitions in terms of "lower layer" types, etc.) ([Tanenbaum2]); the layering of the functions of computer network software from "physical" to "application" via "transport", "session control" and "representation" ([Tanenbaum1]); or the notion of virtual machines expressed in terms of hierachies of abstract data types ([Guttag]). [Sloman1] gives a somewhat more philosophical account of the analogy.

broadly classified as nutrition (peculiar to plants), movement (peculiar to animals), and reason (peculiar to humans) ([Aristotle][106]). It can be understood as an early attempt to describe what we now call a *cognitive architecture* (a term frequently used in this note and of which aspects are dealt with in many contemporary papers, see for example sections 5.2 and 5.3, as well as footnote 53 on page 40). Further above (in sections 2 and 3, page 16) we have, in a way, already discussed this division which corresponds, in more modern terms, to *materials and energy supply*, *sensory-motor co-ordination*, and *planning and communication* functions. ([Benson] addresses at least the latter two.) Accordingly, we identify a hierarchy of lower layers that comprise competencies covering the basic needs of an autopoietic entity, and a hierarchy of upper layers of functions that use the lower layer functions more effectively to ensure the survival of the entity, for instance by triggering anticipatory action. This distinction is also reflected in the anatomy and physiology of the mammal nervous system, whose main components, the *autonomic* and *somatic* parts, are responsible for automatic and (increasingly) *"deliberate"* control, respectively. We can describe the ensuing behaviours as *reactive* and *proactive*.

In this section we take a closer look at current research on ways and means of designing and implementing analogous structures and functions in artificial systems, based on current technologies.

6.2.1 Maintaining the artificial system's body in its environment

The nutrition (or *material and energy supply*) layer of the Aristotelian hierarchy provides the basic functions that are needed to maintain the integrity and stability of an organism within its environment (as explained in section 2). By the same token, an artificial system can not render any service whatsoever in a sustained way, without "being fed" or "feeding" itself, the energy it requires to keep working, and without the possibility of having faulty parts repaired or replaced. Presumably none of our current technologies yields, at least at macroscopic scales, artificial systems that in this regard are capable of *taking care of themselves,* as living systems do. An *autopoietic system* does not need an external operator *of a different kind* to keep it alive [107]. If it did it would not be autopoietic.

[106] for an overview see `http://www.iep.utm.edu/a/aristotl.htm`

[107] Such systems may well require the presence and support of other systems of the same kind: the extent to which an individual organism can be deemed autonomous (or: able to take care of itself) largely depends on its embedding in a social context that is determined by like organisms. Full autonomy may only be achieved at a *higher organisational level*. It is important to emphasize the distinction between *controllers* and *mates*. Organisms at the same level are each other's mates; they do not control each other, in the sense that an individual's behaviour is causally determined by its fellow individuals' behaviour. Rather, its behaviour is determined by the interactions it is capable of.

The biological evolution has brought forth a variety of autonomous and autonomic mechanisms for controlling life-sustaining processes. They are autonomous because they do not depend on an external operator (or controller), and they are autonomic because they do not require conscious reasoning or symbolic communication (let alone computation!)[108]. In section 2 (page 7) we identified some of them in fairly general terms, mainly in the context of basic behavioural patterns such as foraging and self-defence. They include the exteroceptive, proprioceptive and sensory-motor co-ordination capabilities that are crucial not only for locomotion and obstacle avoidance in 3D environments but also for setting and reaching survival-related goals. Whether such mechanisms are within the *cognitive* domain is of course a matter of debate. (Occasionally, the terms *proto-* and *pre-cognition* are being used.) At the very beginning of this note we have already made our position clear on this. It is justified at least in so far as we know of no natural cognitive system (in any agreed sense of the term) where these mechanisms are not in place. In addition one can argue that they implicitly represent effective interpretations (albeit not *actively* and *willingly* gained) of the world and their host system's situation in it.

The question is: what are analogous mechanisms in technical systems and how far do they carry towards the desired goal of keeping such systems operational? The link between *"being alive"* and *"being cognitive"* that we established (or at least made plausible) in section 2 implies two minimum requirements on these mechanisms: (1) they enable the system to monitor not only its environment but also itself, and itself in relation to its environment; and (2) whatever drives them must be fully integrated with the processes they are supposed to keep running. The latter is, of course, only another way of declaring a given system autonomous, or: to be in control of itself. *Whatever controls an artificial (or natural) cognitive system: it must be, conceptually and physically, an integral part of that system, connecting the system's sensing and actuating elements and processes.*

Historically, control - as in technical control theory - has been rather narrowly defined, limited to keeping a device, machine, or plant (sic!) in a desired state or, more generally, the parameters of a process or collection of processes within certain preset bounds (e.g., position, speed, direction, temperature, or the concentration of a given substance in a reaction vessel). It has been widely recognised though, that (at least) the lower levels of cognitive architectures are, in that sense, largely control architectures. Hence, Control Theory and Systems Control Engineering (disciplines we have repeatedly mentioned in this note)

[108] The term Autonomic Computing (mentioned towards the end of section 6.1.2) has been coined with this property of the autonomic nervous system in mind. It obviously implies computation, though.

must play a key role in finding, analysing and integrating the mechanisms that implement "low level" cognitive functionalities in technical systems[109].

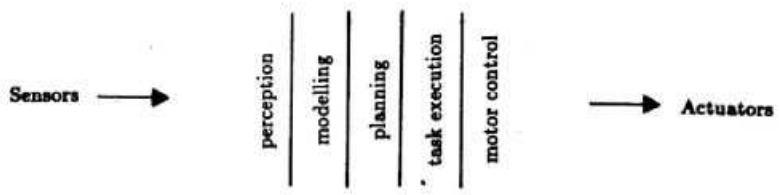

Figure 1: The traditional perspective on robot control (cf., [Brooks1])

Brooks's (one of the main proponents of *New AI*, (cf., Section 6.1.1)) *subsumption architecture* is a well known early example of a "low level" control architecture for mobile robots (see [Brooks1]). It follows from a change of perspective on what links sensors and actuators, input and output, both in terms of structure and content. A traditional perspective would be the one shown in Figure 1: processing sensor input happens through a chain of modules with intermediate results passed on from one module to the next until some result is delivered at the actuator end. Brooks's architecture on the other hand, is based on the insight that certain competencies are prerequisite to others, very much in line with the example ("Before a child talks it walks") cited in the introduction to this section: "before the child walks it stands", "before it stands it crawls" and "before it crawls it sits". Hence, the name "subsumption architecture": competencies needed for walking subsume those needed for standing, and the latter those for crawling. (Apparently, this does not hold for all animal species, at least not in its temporal succession.) In this kind of architecture all layers have access to pertinent sensors and actuators (see Figure 2) but rely for specific action on the function of lower levels in the subsumption hierarchy.

The claim is that in principle, subsumption architectures also extend to the *higher levels* of cognition that form the subject matter of the following subsection. To the best of this author's knowledge, however, this claim remains yet to be proven, in theory as well as in practice.

[109] Ever since the advent of Cybernetics (at least!) control theorists have had a knack for biomimetic approaches. More recently, they have also begun to derive inspiration from Systems Biology, which comprises the study of biochemical pathways and flows in the living cell, the lowest level indeed of the "Aristotelian hierarchy" (see for instance [Sontag]).

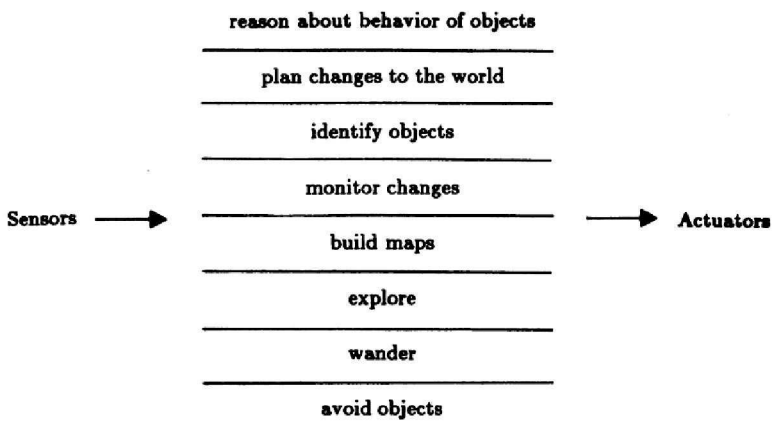

Figure 2: A subsumption architecture for robot control (cf., [Brooks1])

This may be a reason why interest in this kind of architectures seems to have somewhat diminished in recent years. However, the motivating observation remains as valid as ever: implementing *low-level* competencies obviously does not hinge on formal and symbolic reasoning; on the contrary, natural organisms (for instance, insects) equipped with only the scarcest *"computational"* resources, are capable of rather complex behaviours.[110]

Understanding how this works - for instance in terms of neurally based dynamical systems - is considered essential within many projects that deal with robot locomotion. SPARK[111], one of the current FP6-IST-CogSys projects, is a case in point. A collaboration of biologists, cyberneticists, physicists and engineers, it endeavours to build seeing, hearing and moving artefacts, and to link (auditory and visual) perception and action "at the lowest level", based on principles gleaned from comparatively "primitive" animals, such as crickets and other arthropoda ([Frasca]). In that regard it is not limited to co-ordinating the movement of leg-like structures but includes the study of purposive movement in a given terrain (*"what I see is where I go"*, for instance to avoid obstacles). The success of this project hinges on the availability and further development of a CNN (see page 39) based hardware platform called SPARC (Spatial-temporal

[110] Moreover, meaningful (analogue) *"computation"* can be implemented through the way the elementary parts of such organisms are assembled, interconnected and interacting in larger systems, based on their physical properties. In section 6.3.2, under the heading of *"morphological computation"*, we shall briefly come back on this and some of the other issues addressed in the present section.

[111] Spatial-temporal Patterns for Action-oriented perception in Roving robots(`http://www.spark.diees.unict.it/`)

array computer), and of on-chip vision systems, which cater for the analogue nature of the signals to be processed.

Building robots that can climb and walk in irregular, difficult terrains is indeed a major challenge. (Climbing man-made stairs is comparatively easy.) It arouses the interest not only of researchers but - for obvious reasons - also of engineers working in civil or military industry. A European initiative, named CLAWAR[112], which started under the 5th European Research Framework Programme (1998-2002) and is still ongoing, clearly testifies to that interest. So does of course DARPA's already cited Grand Challenge (cf., page 53), albeit here wheels take the place of legs. Autonomous mobility in three dimensions (with wings and rotors) has also long since been added to the agenda, usually with some *military* or *"security"* application in mind (autonomous *Unmanned Aerial Vehicles - UAV*, large and small).

BigDog[113] is another example in the same category: a quadruped of the size of a large dog that has been engineered with input from biologists working at Harvard's Concord Field Station ([Biewener], [Shaw]), who study the anatomy and physiology of animal locomotion. BigDog can climb steep slopes, wade through rough and muddy terrain, carry a substantial payload, recover from stumbling, and keep going while being kicked and pushed. More examples in the league of robots displaying natural modes of locomotion include the bipeds that are being built to mimic human ways of walking and running. Here, the stability problem is even harder (see for instance [Huang] and [Behnke]). The RoboCup initiative, referred to on page 54, would certainly benefit from progress in this domain, not to mention some of the toy and "public relations" robots that are particularly popular in Japan and Korea.

While none of these machines are as yet capable of repairing themselves (see above) some can cope with some degree of mutilation. The SPARK robots for instance can resynchronise their gait and keep walking even if one or two of their legs have been immobilized or if the controller of a leg has been turned off. They owe this resilience not only to the physical properties of their legs but mainly to the way leg co-ordination is managed by suitably switched analogue circuits. More recently, experiments with a moving artefact have been reported that uses *"actuation-sensation relationships to indirectly infer its own structure, and it then uses this self-model to generate forward locomotion; when a leg part is removed, it adapts the self-models, leading to the generation of alternative gaits"*[114].

[112] CLimbing and WAlking Robots. (`http://www.clawar.net/`, `http://www.clawar.com/`)

[113] built by Boston Dynamics Inc. (`http://www.bostondynamics.com`), supported by DARPA with a view to advancing rather obvious military "applications".

[114] quoted from [Bongard1].

Material and energy supply (*"nutrition"*) issues are less frequently addressed under the general theme of artificial cognitive systems. ICEA (see footnotes 71 and 83 on pages 50 and 54 respectively) appears to be the only project in the FP6-IST-CogSys portfolio which seriously considers the energy dimension of autonomy and the need for self-sustenance as potential prerequisites for and drivers of cognition-based behaviours like foraging and self-defence and ultimately, *high-level* processes like planning and reasoning. It comprises experimentation with robots (called *ecobots* by their designers, see [Ieropoulos1]) that are equipped with microbial fuel cells to *"digest"* food (e.g., sugar, insects or fruits, see [Ieropoulos2]) and generate electricity. As in the case of early claims regarding subsumption architectures (see above) it remains to be seen whether these experiments can indeed support the project's main - twofold - hypothesis:

- *"the emotional and bioregulatory mechanisms that come with the organismic embodiment of living cognitive systems also play a crucial role in the constitution of their high-level cognitive processes, and*
- *models of these mechanisms can be usefully integrated in artificial cognitive systems architectures, which will constitute a significant step towards truly autonomous robotic cognitive systems that reason and behave in accordance with energy and other self-preservation requirements."*[115]

In addressing the notion of emotion ICEA is of course not unique. There are numerous research activities, within FP6-IST-CogSys projects and elsewhere, which aim at analysing and modelling emotional phenomena in animals and humans, and implementing in artificial systems what is believed to be the essence of such phenomena. In many of these instances the declared goal (as pointed out in section 6.1.2, page 56) is to have artificial systems interact more naturally with people. This extends in particular to the *low-level* aspects of communication, such as the recognition and expression of emotion (like anger, joy, indifference, etc.) in gestures, facial features or vocal utterances. And thus it is certainly relevant within the context of this section although in expressing emotion, simulation rather than emulation (see section 5.3) seems to be the method of choice for reaching this goal.

In contrast, one of ICEA's alleged goals is to emulate emotions in robotic devices. In pursuing this goal the project's approach is motivated by Antonio Damasio's assumptions regarding emotions and subjective feelings. Damasio explains the latter in terms of brain states representing *"what happens"* in an animal's body ([Damasio]). They arise from the *"homeostatic regulation of bodily activity"* and - we may add - a body's response to external perturbations. They

[115] quoted from `http://www2.his.se/icea/default.htm`; see also section 6.1.1, page 50

are key to the survival of a living organism. Indeed, from this point of view subjective feelings, being based on intricately intertwined physical processes such as the production and consumption of hormones and neurotransmitters, enter the realm of study of process control. This provides an interesting perspective on artificial dynamic systems in general[116].

All the more so as the study of technical process control is not confined to systems with strictly delimited physical bodies like robots, or other moving artifacts, such as cars, ships and planes, but applies to many diverse technical infrastructures with real-time processing needs, including networks of all sorts, as well as large scale systems, distributed or not, such as power and manufacturing facilities. It is pertinent to all systems that need to function flawlessly over long periods of time while making efficient use of their resources, and hence require sensing of and acting on an oftentimes large number of parameters - within or beyond the ranges covered by the natural senses and actuators that occur in the animal world. In a way it would therefore make sense to take an "emotional stance" and to understand and model the internal control of such systems in terms of analogues of some sort of *"feeling what happens"* inside them.

6.2.2 Making the artificial system's mind explicit

As noted in section 2 (page 10) Man, of all animals, is perhaps the only one capable of communicating what he sees, hears, feels, needs, or wants, across *large* distances in space *and time*. (The emphases - on *large* and *"and time"* - are important to distinguish Man's abilities from those of other animals.) He can say what he is doing and, at least to some extent, explain why and how he is doing it.

Clearly, humans, like most other animals, do communicate, in the widest sense, through their behaviour: through what they do and the way they look or move, for example[117]. On the other hand, many animals, like practically all humans, also communicate by using (material) patterns in space and space-time, which, albeit generated by (in, through) their bodies, become external to (or: separate from) them. This concerns in particular vocal utterances (e.g., to indicate pain, aggression, fear, defeat, or readiness to mate), but also various forms of less volatile territorial demarcation.

Such external patterns as well as more general behaviours correspond to and represent neural states or processes that are correlated to what is happening

[116] See also the general discussion in section 2, page 8, on "controlled and controlling processes".

[117] "One Cannot Not Communicate" is one of the five "axioms" Paul Watzlawick established in his seminal work on the "Pragmatics of Human Communication"([Watzlawick]).

in the body itself or the body's environment. In communities of like bodies[118] these patterns and behaviours can trigger (neural or non-neural) processes in other bodies. The meaning (or: informational impact, see footnote 46, page 35) of communicative acts and/or the communicated content can - at least in principle - be expressed in terms of these repercussions. They are likely to be the essence of what we may call *"understanding"* or *consensus*, between different agents: a sense of what some other body is sensing[119].

However, only a human being's actively generated external representations of internal states and processes can have a long lasting effect within communities of fellow humans, through the use of signs, manifested in natural language and other symbolic forms[120]. The study of signs and symbols, and of their usage by humans (in societies and other interhuman relationships), falls within the remit of *semiotics*, a "trans-discipline" inaugurated in the 19th century, by Charles Saunders Peirce, Ferdinand de Saussure and others (for a general introduction to semiotics see for instance [Danesi]). A vast literature has since been created, with contributions from many fields of scholarly and scientific endeavour, and more recently also from the Neurosciences ([Favareau]). This is not surprising given that in a final analysis, all signs and symbols must be grounded in some neural activity relating to a body and its world.

The production and recognition of external signs and symbols is key to the creation of persistent intersubjective and collective memories; it underlies Man's ability to formally plan and reason, individually and/or jointly, and to act accordingly, with all sorts of (beneficial or detrimental) effects on his closer and wider environments. Digital content and virtual worlds - as discussed in section 6.1.2 in terms of specific environments that may host cognitive agents - are perhaps the current epitomes of Man's faculties to create symbolic memories.

Indeed, the biological and cultural evolution has greatly refined the ability of humans to form explicit external representations of what they experience in their worlds, at various scales and levels of abstraction. This has given rise to manifold networks and hierarchies of symbol systems (see section 2, page 10), more or less detached from the immediate needs for sustaining life. So far, the digital electronics based computer is likely the most advanced physical instantiation of such systems.

---

[118] ... bodies that is, which are equipped with neural and non-neural systems of the same or similar kinds, and share the same or similar environments ...

[119] The discovery in the 1980's, of so called mirror neurons in macaque brains ([Rizzolatti]) appears to corroborate this: these are neurons that fire not only on performing a specific action but also on observing such action performed by another body. More recently the "Mirror Neuron hypothesis" has been extended to explaining the evolution of human language ([Arbib]).

[120] ... not the least among them those created for mathematically and algorithmically tractable representations.

Man's growing skills at producing and using symbolic representations appear to be paralleled by the phenomenon of *consciousness* which has emerged from the depths of his animal ancestry. Although a satisfactory scientific explanation of this phenomenon - for instance in terms of predictive models - is still lacking the alleged parallelism alone makes it plausible that both developments are intimately - and possibly causally - connected. While for the time being consciousness can only be fully appreciated from a (subjective) first-person perspective it may be understood - more objectively - as a key property of symbol generating and manipulating brains, based on neural states and processes[121]. As such it enables rational deliberation, internally (in an "inner monologue" or in dialogues with one's own "conscious self") and externally (in direct or indirect communication contexts), based on symbols and, in particular, language. (Before I say a word that word "is in my thought"; before I draw a circle the concept of a circle - with all its attributes - is "on my mind".)

The evolutionary advantages these "high-level" compentencies entail ought to be evident. Homo Sapiens has, after all (and so far!), become the dominant and most powerful species on this planet. Indeed, Man's phylogenetic history as a social animal seems to have favoured brains that can easily recognise and associate sensory patterns and effectively respond to them through explicit communication acts.

But there was obviously no need for Man to do for instance mental arithmetic fast. Consequently, as pointed out elsewhere in this note (towards the end of section 2), humans are not very efficient at formal symbol manipulation. The more abstract the symbols, and the more mechanical the task, the worse it gets - and the faster a suitably programmed computer can perform the task (providing the latter is of manageable computational complexity). On the other hand, the more concrete (i.e., closer to the "real thing", e.g. images) and the more fuzzy symbols are (e.g., handwritten characters), the more difficult it is to turn them into tractable representations and to devise efficient algorithmic procedures that provide the same functionalities as the human brain.

The human brain/body system is most certainly unrivaled when it comes to generating (and, conversely, recognizing and understanding) symbols and concepts: through discriminating, identifying, categorising, classifying and labeling the objects, events and processes that affect its sensory apparatus (from inside and outside of its body), and by putting the representations the brain forms of such entities, in relation to each other and to the representations the brain

[121] While there may be no discernible "neural correlate of consciousness" it stands to reason that consciousness as experienced by humans is due to our ability to "objectify" our own thoughts, i.e., the potential reflexivity of neural processes. Arguments for a strong connection between consciousness and cultural evolution have been advanced by Julian Jaynes ([Jaynes]), in 1976. Not surprisingly, they have been and still are subject to controversial debate.

forms of itself and its body[122]. These cognitive and meta-cognitive competencies are the hallmarks of human ("natural") intelligence, and the very basis of "human-like" behaviour. (In their capacity of finding their way in their environment or avoiding their predators even most (non-human) animal brains are still unsurpassed by man-made signal and data processing machinery.)

We conclude: A human brain/body system can actively make its world explicit to itself (e.g., for acting through planning and deliberate reasoning[123]) and to fellow systems (for joint planning, reasoning, and acting; for deliberate communication, etc.) in an intersubjective consensual domain, spanned by signs and symbols and, in particular, language. A variety of brain functions (or: "mental mechanisms"), collectively known as "consciousness" and supported by the brain's anatomy and physiology, are likely to make these feats possible[124].

The obvious question we have to ask at this juncture is: How can artificial systems (compact machines or distributed systems, a robot or a network of sensors) be made to behave - in this sense - human-like? How should a machine be built so that it tells us in our terms and at to-be-determined levels of detail, what happens to be in front of (or behind) it, what is happening around it, to it, or inside it? What it is doing and why? How can we make a machine understand what we want it to do? How can a system improve its performance (in rendering a service, for instance) by observing what it is doing and how; by taking account of the effects of its actions, and then drawing the right conclusions, based on some sort of symbolic reasoning? How can it make "its mind" explicit to itself and fellow systems?)

---

[122] Recall that these representations are in particulars of the brain's neural states and processes. States of and processes in a brain (human or not) *are* the subjective experience of that brain in its body. Evidently, some states and processes in *human* brains (brains of certain sizes and structures, in certain types of bodies, that is) can give rise to symbolic output. But they should not be considered symbols themselves if we restrict (as in this section and in fact, this note) the term symbol to *"physical representations"* that we generate, *"perceive and manipulate in our environment"* (see W.J. Clancey's contribution to the debate on the symbolic versus non-symbolic nature of human cognition, in [Clancey].)

[123] *Calculation* is considered to be a special form of reasoning.

[124] Note that we link the terms *"consciousness"* and *"explicit"*. Clearly, many if not most of the representations the human brain forms are neither conscious nor can they easily be made explicit. Also, many of the skills humans can acquire are not acquired through conscious learning. Language is a case in point: most of us become experts in the use of our Mother Tongue as we grow up, and few are able to explain why they use it the way they do. On the other hand, practicing those skills that people normally acquire through conscious learning can become entirely unconscious sensorimotor or reflective operations: riding a bike, driving a car, playing the piano, or playing chess (see also Eugen Herrigels account of his venture into the art of archery: [Herrigel]). Last but not least, most likely everyone who has ever seriously been engaged in solving mathematical problems will remember the *"aha"* of finding a solution seemingly out of the blue, not through conscious reasoning. However, making that solution explicit, explaining it to a student for instance, does require conscious work.

Questions like these have been asked ever since the electronic computing machine was invented; and before, although perhaps less seriously and with little prospect of satisfactory answers. In section 3 we have indicated some of the reasons why one may ask these questions in the first place: they address key requirements for artificial cognitive systems. Meeting these requirements would not only enable a wide range of new services but also make the rendering of many services more natural and "human-like" (or, in traditional terminology: more "user-friendly"), more robust and less dependent on frequent human intervention (i.e., more *"autonomous"*).

Admittedly, answers of sorts have already been given - and implemented - in the not so recent past. However, image and video analysis and interpretation, speech recognition and generation, text understanding and classification, in real-world contexts and under real-time constraints, do pose problems that, on a practical level, we can approach only now, with suitable and affordable sensor, memory and processing hardware in place. Powerful hardware apart, solving these problems requires - as in the past - deep research into the principles and methods of symbolification and symbol interpretation: of producing symbolic descriptions, and - conversely - of understanding the symbolic content, of multimodal real-world scenarios.

In the remainder of this section we illustrate some aspects of this research, with examples from IST projects and other initiatives. We confine the discussion to image, complex visual scene, and language understanding. Image and scene understanding should become manifest in symbolic (and ideally, linguistic) descriptions ("symbolification") whereas (spoken or written) language understanding should entail actions (or responses) that are consistent with a speaker's utterances (or a writer's text) and their respective contexts ("interpretation"). While this distinction may be debatable it does serve our purpose to narrow down further the vast scope of the technologies at issue here, to vision and natural language[125].

As far as vision is concerned, we may want a system to look at an image and tell us what or who is in it, or to look out of the window and tell us what is happening in the street. Or we might want a system to find out if a video clip contains material that would not be suitable for children to watch. This sounds rather demanding. Yet there are many projects that aim to provide systems with these and similar capabilities. (See [Forsyth] for a viable approach to detecting obscene nudity in images.)

[125] Symbolification must, after all, start from some sort of interpretation. However, the distinction we make is certainly justified if we view symbols, apart from being external, as essentially man-made (see footnote 122, 72). Of course, one may argue that all perception is symbolic, perhaps as Johann Wolfgang von Goethe did in the closing verses of his Faust II: *"Alles Vergängliche ist nur ein Gleichnis ..."* (*"all things transitory are but parable"*, or: *"all that is transitory is but a symbol"*). But this would probably render useless the very concept of something being a symbol.

Given a real-world scene, as seen for example through a digital (video-) camera, and recorded and represented in digital memory structures, these projects, in more abstract terms, are poised to tackle problems such as:

- find distinguishable objects (which may be living or non-living entities);
- identify and classify objects;
- determine the (relative) positions of objects and other spatial relationships between them;
- identify moving objects;
- determine attributes (e.g., (relative) direction, speed, acceleration) of movement;
- identify and classify the behaviour of objects;
- draw conclusions as to the future (and past) behaviour of objects in dynamic scenes; ... etc.

Clearly, not all of these problems are relevant in all contexts. For still images analysing movement is usually irrelevant, while identifying motion-based behaviour may be of importance. Most of these problems are particularly challenging in scenarios where the observing system itself is not stationary or where lighting and other evironmental conditions change. In any event, solving one of these problems, in principle includes formulating an answer that complies with some human ontology (i.e., a person's conceptual understanding of the scene in question, see Table 2, page 19) and hence can and must be made explicit in human terms. (This is obvious in the case of systems that are supposed to answer to specific queries related to a scene.) This requires at least a subset of the "high-level" cognitive competencies that this section is about. They must be based on suitable representations the system forms of the world around it.

Research with a view to implementing these competencies in machines and thus, to achieving "computer (or: machine) vision", has been undertaken ever since it became feasible to scan and digitise images and dynamic scenes. A noteworthy attempt at boosting it to a new level has been undertaken through a series of projects funded under an IST initiative dubbed "Cognitive Vision", that was launched in 2001, within the European Commission's 5th Framework Programme. Table 5 provides an overview of these projects in terms of functional objectives, key techniques and potential applications[126]. Although none of the systems constructed goes so far as to provide extensive narratives of what they see, many of them are *proto-linguistic* in that they do label objects and describe

[126] Links to the websites of these projects are listed on `http://cordis.europa.eu/ist/cognition/projects.htm`.

Project	Functional objectives	Key techniques	Potential applications
ACTIPRET: Interpreting and Understanding Activities of Expert Operators for Teaching and Education	Robust object detection, tracking, recognition in 3D; hand gesture recognition; semantic interpretation	Bayesian modelling, Hidden Markov Models (HMM), Radial Basis Function (RBF) nets	guiding people in performing complex object manipulations
CAVIAR: Context Aware Vision using Image-based Active Recognition	Improving image-based recognition through visual attention and task, scene and contextual knowledge	Scale-invariant feature transform (SIFT), auto-associative memories, HMM, statistical learning	human behaviour analysis, surveillance
COGVIS: Cognitive Vision System	categorization and recognition of objects and events by mobile agents	Bayesian modelling, developmetal learning	service robotics
COGVISYS: Cognitive Vision System	transformation of video signals into textual descriptions of recorded scenes	Situation Graphs and Situation Graph Trees, Discourse Representation	traffic monitoring and surveillance
LAVA: Learning for Adaptable Visual Assistants	automatic online acquisition of categorical knowledge (of objects, scenes, events)	Kernel methods	content based and context sensitive information retrieval, video description
VAMPIRE: Visual Active Memory Processes for Interactive Retrieval	learning concepts, categories and spatiotemporal relations	neural classifiers, attributed relational graphs, Kernel methods, Bayesian methods	interactive content-based image retrieval, object and action recognition
VISATEC:Vision--based Integrated Systems Adaptive to Task and Environment with Cognitive abilities	locating and identifying fixed objects	mix of appearance and model based methods, multi-cue integration, scene tensors	industrial robotics

Table 5: Cognitive Vision projects in the FP5-IST programme

some of their individual and contextual characteristics in formal terms, based on acquired or preset (i.e., designed in) conceptualisations. While the approaches taken are not explicitly grounded in the analysis of biological systems, some of the methods and techniques used - such as artificial neural networks or Bayesian methods - do call to mind either bio-inspired modelling (in the case of artificial neural networks), or probabilistic reasoning (as in the case of Bayesian methods) that may well be more "natural" than classical deductive logics.

This also holds for CLASS[127] an FP6-IST-CogSys project which, similar to the LAVA project (see Table 5), encompasses research on ways of automatically categorising objects (contained in images embedded in text) and scenes (in video footage). For that purpose it exploits, where possible, textual objects (such as captions) that go with the images. It borrows a technique known as *Latent Semantic Analysis* (LSA, see [Landauer]), developed in the early nineties (of the 20th century) for indexing text, and combines it (or rather its probabilistic extension, see [Hofmann]) with *Support Vector Machine* (SVM) based image analysis and face recognition methods. Machine Learning (see section 6.3.1) is indeed the hallmark of this project. Its results will be demonstrated through an image interrogator, a video commentator and a news digest (summarising semantic content derived from video and text). These are just three of many possible applications; they would certainly suit well the kind of content agents discussed in section 6.1.2.

For any embodied agent that moves in a "common sense environment" it is of prime importance to understand its visible world. Advances in robotics therefore also crucially depend on progress in mastering artificial vision. Hence, of the projects cited in the previous sections, several of those dealing with robotic platforms do aim to improve for instance navigation and manipulation, through better vision. BACS (Bayesian Approach to Cognitive Systems[128]), yet another FP6-IST-CogSys project, applies Bayesian techniques not only to robot navigation in 3 dimensions but also to face and human behaviour recognition in real-world environments.

Although in this section we focus on vision we should not fail to mention that problems similar to the ones listed above can also be posed with regard to audition and haptics. Describing and reasoning on acoustic scenes for instance is no less a challenge than making sense of visual scenes. The FP6-IST-CogSys project DIRAC (Detection and Identification of Rare Audio-visual Cues[129]) partly addresses these problems in the auditory as well as in the visual domains.

If robots and other artificial systems are to render useful services communication with them should go both ways; they should also be able to understand

[127] Cognitive-Level Annotation using Latent Statistical Structure, `http://class.inrialpes.fr/`

[128] `http://www.bacs.ethz.ch/`

[129] `http://www.diracproject.org/`

instructions given by people, by translating them into desired actions. Traditionally this has been done - from the very beginning of the computer age - through carefully crafted formal command languages that human operators had to use. However, the challenge is for a machine to infer the need for action from unconstrained human vocal utterances. It should understand spoken language in all its nuances, not to mention ordinary texts in writing. (Remember, we want our artificial systems to be *natural*!)

The latter has been on the agenda ever since it was realised that computers could be used for more than calculating ballistic curves. The initial interest in Machine Translation for example, has never ebbed off although it has taken a long time for its concrete results to become acceptable input to human post-editors. Yet, serious applications of Machine Translation are limited to rather narrow domains, such as technical documentation. But it is nevertheless speech-to-speech translation that is now at centre stage, as in the TC-Star project (see footnote 92, 56) mentioned in section 6.1.2.

Clearly, research with a view to achieving the more general goal of speech understanding also contributes to making spoken dialogue a viable option for designing man-machine interfaces. Typical applications of this technology would for instance be to all sorts of services requested and specified via the telephone.

For these applications to work satisfactorily speech understanding must go way beyond mere speech recognition; it must include the interpretation in unambiguous formal terms of what is being said. To achieve this a system may have to engage its human user in a clarifying and disambiguating dialogue. The quality of such dialogues largely hinges on the availability of comprehensive ontologies (see Table 2, page 19) of the application domain at issue. They provide the semantic ground for understanding a human user (by inferring her wishes, for instance, from ontological relations and rules) and triggering the appropriate response, be it the passing of parameters to an application program, posing further questions or formulating an error or exception message (see for example [Milward]).

Where do these ontologies come from? In principle, the situation here is similar to the one agents in digital content spaces have to deal with (as explained in section 6.1.2): the way these ontologies originate and develop makes all the difference between a cognitive and a non-cognitive system. Systems on which they are externally imposed and maintained can not be considered cognitive. In contrast, a natural language dialogue system for instance may be deemed cognitive to the extent it builds (and modifies) its ontologies on the fly while conversing with its human users. This is usually referred to as *ontology learning* (which so far appears to be applied mainly to purely textual environments, see for example [Maedche]).

Research with a view to enabling artificial natural language dialogue has been undertaken for nearly as long as on advancing Machine Translation. It encompasses many fields and approaches, with statistics and Machine Learning (see section 6.3.1) figuring prominently. It has come a long way from Joseph Weizenbaum's 1966 parody of a psychiatrist (the famous ELIZA programme described in [Weizenbaum1]), to dialogue based tutoring systems (see for instance [Benzmüller] and [Jordan]), spoken language interfaces to databases (with commercial applications, e.g., [Dupriez]) and planning systems (e.g., [Allen]). COMPANIONS, one of the FP6-IST projects in the area of *multimodal interfaces*, ambitiously attempts to provide people with lifelong personal access agents (*"companions"*) on/to the Internet, virtual creatures that converse with their users in spoken natural language[130] ([Wilks]).

The challenges in this domain are indeed formidable; they also include language identification, robust speaker independent speech recognition, speaker identification and context dependent vocal response generation, with natural (and perhaps emotional) pitches and modulation. In fact, since the very dawn of the computer age natural language dialogue has been considered a yardstick for "machine intelligence", through the famous test scenario that Alan Turing already proposed in 1950 ([Turing2])[131]. Integrating so called *Common Sense* is perhaps one of the most difficult problems that need to be solved before a dialogue system can stand a chance of passing that test: to formalise and codify the myriad of snippets of trivia that make up nearly everyone's everyday knowledge. In an enterprise called *Cyc*, started in the 1980s, Doug Lenat and Ramanathan Guha, set out to do just that ([Lenat]). Cyc is supposed to be an en***cyc***lopedic ontology (it is, in fact, a collection of ontologies) of common sense knowledge, that for instance man-machine dialogue systems could draw on. While largely handcrafted originally, its proponents have more recently turned to automating its "growth" (see for instance [TaylorM]), thus in a way joining the ontology learning band waggon.

Most conventional approaches to endowing artefacts with linguistic capabilities (such as speech recognition, speech generation, natural language dialoguing) encounter problems not only in dealing with communicative context; they also seem not to take proper account of the uncertainty inherent in a human user's communicative behaviour. Refined statistical methods, probabilistic ontologies ([Costa]) and even more powerful hardware may not lead to satisfactory solu-

[130] `http://www.companions-project.org/`

[131] In 1991 this scenario has been set on the stage of an annual (rather controversial) *"Turing-test"* competition sponsored by the New York "philanthropist" Hugh Loebner (`http://www.loebner.net/`) with US$ 100.000 prize money to win. Turing allegedly meant this test to prove or disprove the presence of "intelligence". However, apart from the fact that there appears to be no agreed definition of intelligence, this test can at best only determine whether or not a system behaves more or less in a human-like fashion.

tions. A deeper understanding of the human (biological and embodied!) language capacity would indeed be required as a basis for building language-endowed artificial systems.

Why do we speak? What makes us speak? How did language evolve? What were its advantages in the ecological niches that primates and eventually Homo Sapiens thrived in? What is the body's role in forming (a) language? What is the role of the social environment? Why do humans speak so many different languages? How do their languages evolve? How are they structured (syntax)? How does the meaning of words, phrases, and other vocal utterances come about (semantics)? How is language used, depending on context and a speaker's intentions (pragmatics)? How does it develop in a child?

Some of these and similar questions have been raised by philosophers ever since our ancestors (e.g., in ancient Greece) began to take a rational and systematic approach to understanding their worlds. More recently, following the diversification and specialisation (or fragmentation?) in science and the humanities, philosophers have been joined by linguists, psychologists, anthropologists, ethologists and representatives of other pertinent disciplines. Together, they have created a vast body of literature and established a large number of empirical results. It goes without saying that even the slightest attempt at reviewing these in this note would be utterly futile.

Language and "How the Mind Works" ([Pinker]) have nearly always presented closely related sets of (oftentimes controversially debated) issues. Even more recently, these issues began to attract computer scientists and engineers with an interest in questions like the above enumerated ones (see for instance [Winograd1, Winograd2], [Schank]). They are now teaming up with language experts, developmental psychologists, and neuroscientists, in projects aimed at demonstrating ideas about language evolution and development through the construction of software-driven machines ("robots") that (need to) communicate among themselves (in "swarms" or "teams") and/or with people. The Paco-Plus project, mentioned in section 6.1.1, falls into this category. But there are many more, setting up and carrying out experiments designed to test (currently popular) hypotheses that claim sensory-motor capabilities and experiences to be a possible basis for concept formation, *symbol grounding* ([Harnad]) and - eventually - language development. (Accounts of this type of work can be found for instance in [Cangelosi], [Roy], [Steels], [Weng], to name but a few.)

In fact, language as we know it can be understood as only the ultimate (most sophisticated?) way of interacting in an evolving social world. However, as stated at the beginning of this section, language as we know it, is not the only way of making one's worldview explicit. Categorisation of phenomena in a given environment for instance, can easily be made explicit through speechless action. But then this kind of action also becomes a sort of language (just as speech, in a

final analysis, is muscular action) and it should be possible to study its syntax, semantics and pragmatics ([Guerra]). Several of the other (robotics oriented) projects discussed in section 6.1.1 which set out to explore the connection between perception, action, concept formation and symbol grounding (e.g., Gnosys and Cospal) are preparing the ground for doing just that.

It remains to be seen whether the results and further developments emerging from these experiments can become substantial contributions towards advancing the language technologies mentioned (or alluded to) in this section. These technologies are certainly still a far cry from what may have been promised or expected in the early days of Artificial Intelligence research. The ways people communicate and the contents of human communication are, after all, dauntingly complex, biased by people's history as individuals, members of social groups and, last but not least, evolved biological organisms. This makes it particularly difficult to be clear about the meaning of *human-likeness* or *human-like communication behaviour*, in connection with artificial systems. When it comes down to it, dogs, horses and parrots may still be more human-like than any of the machines we make.

6.3 Closing the gaps

A large part of the research prompted by some of the projects mentioned in the previous sections (and especially those referred to in section 6.2.2) in one way or another appears to be committed to the Aristotelian hierarchy (outlined briefly in sections 6.2 and 6.2.1): it contributes to explaining the transition in the evolution and development of biological organisms from supporting life (or: metabolism) sustaining functions via sensory-motor processes to (conscious) symbolic expression and reasoning. The projects themselves, however, are clearly motivated by the desire to apply knowledge gained through this kind of research to the design and construction of all sorts of artificial systems so that they live up to the qualities described in Table 1 (page 3). Many of the questions that have been in the realm of philosophy (and theology!) at least since Aristotle's days have become scientific and technical problems, to be tackled in analysis-modelling-synthesis loops, as discussed in section 5 (see also our reference to [Harvey], towards the end of section 3). Explanation thus boils down to actually building ("Aristotelian") systems that are robust, versatile, adaptive and autonomous in delivering all sorts of (more or less) sophisticated services *because* their high-level capabilities are firmly grounded in low-level structures and functions.

Seen from our current vantage point as human beings nature has obviously succeeded in making this transition (of which we are living proof). And we are reasonably certain as to the general principles and mechanisms employed. An

important class of such mechanisms is usually subsumed under the term "bio-evolution", denoting a process believed to have started at the molecular level under conditions existing on our planet not long after its birth within the cosmic dust that surrounded our sun (see for instance [MillerS] for early experimental evidence and [Eigen] for an early theoretical treatment). It has been going on for billions of years, unfolding the vast richness of life that we encounter today and in the fossil record. Allegedly, at all times those life forms have prevailed that were most capable of securing the resources they needed in order to maintain themselves and reproduce in a given environment. This capacity includes the ability to adapt to changing environmental conditions (e.g., through various ways of economising on the resources available, by changing the temperature of the body, the colour of the skin, etc.) and eventually, also the ability of organisms of a given species to actively transform and adapt, individually or jointly, their environment according to their needs.

This requires solving problems posed by the environment they live in. Often, as part of the solution to a given problem (related to, for instance, obtaining food or building a nest), they create and use tools as extensions of their own bodies. In fact, there is a growing record of observations of all sorts of animals showing that Man is not the only creature endowed with this talent[132]. However, and notwithstanding some animals' exceptional skills, there is no doubt that Man's problem solving capacity by far exceeds that of all other animals.

Coping with change by changing is perhaps the most basic principle underlying both evolution as an adapting process and the functioning of the (adapted and adapting) organisms brought forth by it. Evolution (or, rather, life as a whole) applies this principle *mindlessly*, without any particular plan or goal, whereas many of its creatures do it - consciously or unconsciously (deliberately or instinctively) - in pursuit of very concrete goals. Of course, the time scales on which this is happening are as different as can be. But both types of processes and systems have in common that the changes they have to cope with are often caused by themselves and not only by forces external to them. (Examples abound: the atmosphere is a case in point; it allegedly owes its oxygen to the earliest life forms, and its current increase in carbondioxide to unfettered human consumption of the enormous amounts of fossil fuels that accumulated over many millions of years in the earth's crust.) It is worth noting here that an organism's activity in its environment - whether deliberate or not - may result in changes that make future changes (of the environment) more *predictable*, *less harmful* and more *manageable*. Man's own history as a *homo faber* demonstrates

[132] Studies of tool-making primates have been commonplace ever since Wolfgang Köhler's seminal work on apes in the early 20th century ([Köhler]); more recently corvids have been observed making tools as well (see for instance, [Bluff]). As tools can be considered external knowledge representations Man is also not the only creature who produces such representations.

this most clearly (although the above example - and many others, of course - seem to testify to the contrary, at least as far as a reduction of potential harm to our planet, habitats, and to ourselves is concerned).

Key lessons we can learn from the history of life on earth can thus be summarised as follows:

(i) The overall composition of the *biosphere,* as well as the structures and functions of its constituent organisms, are subject to long-term *random processes* (these are based, in standard parlance, on the reproduction, mutation and limited life spans of individual biological entities, and on competition among biological entities for limited resources); they induce and are influenced by more or less rapid changes in life's own environment[133];

(ii) an organism's chances of reproduction hinge on its ability to make the most of its environment (for that purpose);

(iii) an organism's environment usually contains many other organisms (of the same or different species) with which it interacts;

(iv) most environments are *non-deterministic* (in the short and medium term) which makes exploitation problematic;

(v) there are at least three ways for an organism to cope with non-determinism: (a) to change itself *in reaction to* environmental changes (i.e., to adapt itself to the environment); (b) to *predict* changes and act accordingly (e.g., to change location, to seek shelter, etc.); or (c) to change the environment with a view to making it more deterministic (i.e., to adapt the environment to its needs); all three options must be based upon capabilities for *sense-making* and *acting*;

(vi) options (b) and (c) are likely to be more advantageous in the long run, especially if an organism can *improve their effectiveness* during its lifetime;

(vii) whether or not it is possible for an organism to improve the effectiveness of its sense-making and acting capabilities (under any of the options (a), (b) and (c)) while living in some environment, depends on the organism's evolved structures and functions.

[133] In Faust's study Goethe's pantheistic spirit succinctly describes these processes in beautiful verses: *"In Lebensfluten, im Tatensturm / Wall' ich auf und ab, / Wehe hin und her! / Geburt und Grab, / Ein ewiges Meer, / Ein wechselnd Weben, / Ein glühend Leben, / So schaff´ ich am sausenden Webstuhl der Zeit/ Und wirke der Gottheit lebendiges Kleid."* (rough translation: *"In the tides of life, in actions' storm / Up 'n down I wave, / Waft I to and fro! / Birth and grave, / An eternal flow, / With change so rife, / A glowing life. / Thus I am toiling at Time's whirring loom / And weaving the deity's living gown."*) (Faust, Erster Teil, Nacht)

Given these lessons, how can Artificial Cognitive Systems research and engineering close the perceived explanatory gaps between neighbouring levels of the Aristotelian hierarchy? In particular: how can the key ability of natural processes and systems to *cope with change by changing* be translated into the artificial domain? What are the implications for the design and implementation of hardware and software, depending on the specific environment of operation? We shall address these questions in the following two sections, thus resuming the discussion of some of the issues raised in sections 5.2 and 5.3, and highlighting some of the research activities that might provide (more or less satisfactory) answers.

6.3.1 Adaptation and learning in artificial systems

As argued above, to *"cope with change"* is but an instance of the more general notion of *"problem solving"* through sense-making and acting. No matter which of the basic strategies (a), (b), or (c) (see item (v), in the above list) or any mix of these a biological or artificial agent adopts, it has to solve a problem in its real world. And there are innumerable problems indeed, as encountered in previous sections, derived from the supreme goals of staying intact and (certainly for artificial agents and systems) rendering a high quality service. They range from distinguishing between different visual or acoustic patterns, via moving towards or avoiding some object, to speaking and understanding a (foreign) language, or designing an artificial cognitive system (!).

For humans and animals the ability to solve basic, survival related problems is usually linked to the developmental stage of the animal, child or adolescent. In fact, there are many animals that demonstrate remarkable problem solving capacities at very early stages of their development, and often from the moment they are born or hatch. Some animals, such as foals or baby deer, are able to synchronize their gait and walk, right after they leave their mother's womb. Others, like fish and some fowl, can swim from the very beginning of their life. By contrast, human infants and children must learn to walk and to swim (which usually requires a major effort on their part). With regard to a specific competence or set of competencies one may therefore distinguish between precocial and altricial species, the former owing their competencies in large part to the (long-term) mechanisms of evolution whereas the latter acquire them thanks to their individual *learning* abilities. (See for instance [Starck]) for a discussion of the rather broad altricial-precocial spectrum pertaining to the class of birds (*aves*).)

Many animals and, of course, humans *learn* - in the more common sense of this term - through some form of interaction with their environment: their fellow-inhabitants and the objects they come across. They can change, within narrower or wider limits, their reasoning and behavioural patterns as they go along. Interaction may be limited to mere observation and imitation. Humans

can learn consciously (knowingly) or unconsciously (unknowingly) (see footnote 124, page 72); we are less informed about other animals, at least as far as their being conscious (as experienced by the author of this note) or not is concerned.

However, no matter whether cognitive faculties have *evolved*, *develop* in the course of individual growth, or are being *learned* (in the above sense), in either case we are dealing with the *acquisition of competencies* allowing an individual (animal or human) to solve problems as they come up. (Obviously, evolution has equipped Homo Sapiens with learning mechanisms to the extent that he can not only solve more problems than any other species but also create more.) We shall therefore use the term *"learning"* with this wider interpretation in mind: we say a system learns if it autonomously gains and improves pertinent problem solving capabilities. In this context, autonomy comes down to the system's ability to modify and, possibly, to extend itself. *Self-modification* and *self-extension* thus also become key characteristics of cognitive systems, natural or artificial ones[134].

The projects so far discussed in this note, aim at creating artificial systems that are supposed to solve problems of the above mentioned sorts (and more, of course). In many (if not all) cases these problems arise (or could arise) in the context of the system's (or machine's) interactions within a given environment. The particulars of the environment and the problems to be solved are unknown at the design stage. It is therefore not surprising that the term *"learning"* has been fairly extensively used to denote one of the most salient features of these systems namely, to adapt to changes in their environment and/or to improve their performance over time. *"Learning"* is perhaps their greatest common divisor, their core functionality, so to speak. Indeed, if a system were not meant to learn and yet be flexible and adaptive its designers would have to think of the minutest details of its envisaged environment, and all eventualities occurring during its operation.

Systems of the types that interest us here hinge for the most part on digital structures which, in terms of human cognition, correspond to percepts, concepts and forms of thinking (i.e., ways of generating, associating and processing percepts and concepts). Hence self-modification and self-extension in artificial systems must take place at the level of these structures. In essence then, artificial learning is a matter of generating ever more useful and efficient structures (with respect to the class of problems at issue), and ever more efficient algorithms operating on them.

As already mentioned in various instances in previous sections: approaches to and methods of making artificial systems learn are within the remit of *"Machine Learning"*, a discipline drawing on various sub-fields of Mathematics, with Mathematical Statistics figuring prominently. As a branch of Theoretical Com-

[134] These characteristics are necessary but not sufficient. They most certainly apply to the (terrestrial) biosphere as a whole and to the human (and other animals') brain. Clearly, there are significant differences, not only as far as time scales are concerned.

puter Science it also has roots in early Artificial Intelligence research on game strategies and general problem solving ([Newell1]). Tom Mitchell describes Machine Learning as *"a natural outgrowth of the intersection of Computer Science and Statistics"* and continues:

> *"Machine Learning focuses on the question of how to get computers to program themselves (from experience plus some initial structure). Whereas Statistics has focused primarily on what conclusions can be inferred from data, Machine Learning incorporates additional questions about what computational architectures and algorithms can be used to most effectively capture, store, index, retrieve and merge these data, how multiple learning subtasks can be orchestrated in a larger system, and questions of computational tractability."* ([Mitchell])

Statistics (and hence probability theory) plays a dominant role because, after all, we are dealing with environments that are characterised by uncertainty. Coping with uncertainty requires us to make guesses and estimates based on assumptions about the statistical laws governing the events around us. The more realistic these assumptions and the more precise the estimates are, the better we fare. (In [Chater], Chater et al make a very strong case for probabilistic approaches to modelling cognition.)

Machine Learning also has firm roots in neuro-biology. Indeed, a computational architecture quite popular among builders of artificial cognitive systems, has been mentioned time and again in this note: *Artificial Neural Networks* (ANNs) that have been modeled after their "real life" counterparts more than six decades ago ([McCulloch]), and become significantly more varied and practicable since. An ANN can learn, as explained in section 5.2, by changing the weights of connections between its artificial neurons and hence its firing patterns, as determined by an algorithm implementing some learning strategy and rules. Its internal structure keeps changing in accordance with the data it "experiences", until the system hosting it (its "body") is good enough at carrying out a given task. And it may change again (or, in other words: adapt) if either the task specification or the operating conditions, or both, change.

ANNs learn - as all learning machines do - a mapping from an input data space to an output data space. In general, the output affects not only the hosting system's environment but also the system itself. In some applications of Machine Learning the output space is quite simple. In others, both input and output spaces can be richly structured. For *classification problems*, which come up in many applications, the output is usually an element of a finite set of labels: given a set of features of a certain object or phenomenon and a finite number of classes, what is the most likely class that object or phenomenon belongs to? Another case in point is *regression* where the output space can be as simple as the set of real numbers (insofar as that set can be considered "simple"). Both,

classification and regression, should allow to predict with some confidence, the output that corresponds to a given input. In *time series* regression for example, the output may be a probability distribution over a set of possible values of some random variables, at some future point in time.

Learning through *supervised training* is the standard approach to solving classification and regression problems. In its learning phase a classification or regression algorithm is presented with a set of known input-output pairs, the *training set*. Assuming a plausible parameterised mechanism (e.g., an ANN) underlying the input-output relation, the algorithm then fits the parameters of that mechanism (e.g., the connection weights of an ANN) to the given data. In order to reliably process previously unseen inputs the training set often needs to be fairly large. This can give rise to hard tractability problems.

Other learning scenarios call for algorithms that discover some structure (patterns, features, regularities, irregularities, distribution densities, cause-effect relationships, etc.) in a given data set. Such algorithms are *unsupervised* in the sense that there is no training set to start from. They would typically yield data describing or representing the structures found. Unsupervised learning can inter alia be applied to clustering data derived from an image or 3-dimensional scene, by determining the shape, size, location and other features of salient patterns. These features could then be used to define input classes to be further dealt with by some classification algorithm. They can also be used in concept formation.

We also note that *"self-organisation"* of a system in response to environmental conditions (see section 5.2) can be viewed as an unsupervised learning process. A classical example is Kohonen's Self-Organising Map (SOM, see also page 48), a representation (and visualisation!) of highly dimensional input spaces in a two-dimensional grid of nodes ([Kohonen]). The latter could be interpreted as modelling cortical structures of neurons that are subject to lateral inhibition.

Perhaps the most general way of describing a system's (or agent's) operation in a data-defined environment is the following: whatever the system does at any given time (its "choice") probabilistically depends on its present internal state, the currently available and the anticipated data. In other words: the system chooses from a number of options according to some probability distribution which may be biased by what it expects to happen in the future. (In degenerate cases it may have but one choice.) Its future internal states depend on its present internal state and what it is doing now. The further into the future that latter dependency persists the more memory can be ascribed to the system. Here, "what it expects" is shorthand for "what future data and states it infers from its current internal state (its memory!), the current data, and its potential actions". (Mathematically, this sort of description could be formalised in terms of controlled stochastic processes.)

This type of scenario is rather typical of an agent that seeks to survive in some environment (or simply achieve a certain objective, like rendering some service at a sustained level of quality). It would require a learning mode which takes account of the positive or negative reward associated with actions, as determined by their putative future effects, such as getting closer to or further away from a given goal. What needs to be learned here - through reward induced *reinforcement* - is a policy, a set of rules, that can be applied to selecting an action whenever this is called for. Preferably, the strategy should improve over time in terms of increasing the agent's chance to reach its goal or, to put it more generally, to gain as much as possible through its actions.

Although the particulars of the problems to be solved by a learning algorithm are not known at design time some assumptions have to be made as to the most likely or most suitable mechanism that produces the data which characterise a problem and specify its instances. Such assumptions, usually called *"models"*, (occasionally also *"hypotheses"*) are needed regardless of which learning mode - *supervised*, *unsupervised* or *reinforcement* - is being applied. Once a model has been selected learning becomes a matter of tuning its parameters in such a way as to satisfy the performance criteria pertaining to the problem at issue.

Models usually reflect the statistical and/or functional properties of the phenomena or objects that are being studied, as understood by their human observer. In some cases the structure of the set of parameters is itself parameterised, giving rise to a richer class of models to choose from and making the algorithm more flexible. In any event it is the designer who sets the basic *"rules of the game"*. In view of the *altricial-precocial* spectrum observed in nature this should not be dismissed as too unrealistic. In fact, nature, through evolution as a *meta-learning* process, endows its creatures with mechanisms so that they may act "successfully" in their respective environments. These mechanisms are more or less powerful and flexible in terms of their learning capacity; in any event they do represent the models (in the above sense) that are reasonably well suited to solving the problems a given organism needs to solve in its lifetime (and sometimes more). The "algorithms" and their physical substrates, that we (like any other animals) are born with, represent the knowledge nature has been amassing for eons. ([Sloman2] discusses some of ramifications of this - rather obvious - insight and its consequences for the design and implementation of artificial systems, such as robots.)

The areas of Statistical and Computational Learning abound with sophisticated mathematical concepts and tools geared to formulating, analysing and reasoning about learning algorithms, based on all sorts of computational structures (neural networks, kernel representations, decision trees, Bayesian networks, finite automata, etc.) and used in all sorts of applications. (Two approaches to "learning in artificial systems" have become increasingly popular in recent years:

Kernel methods are based on representations of sets of objects (taken from some superset: images, text strings, sound patterns, protein sequences, etc.) in terms of matrices, that express numerically the *"similarity"* between pairs of objects. The numerical values are obtained through so-called *kernel functions* which can be interpreted as scalar products (sometimes termed *dot products*) in suitable *vector spaces*. This allows for efficient algorithms (in particular for solving classification and regression problems, but others as well) involving only scalar products in vector spaces; they can be applied to many different types of objects, without (major) changes. (For an excellent elementary introduction to Kernel methods see [Vert].)
Bayesian methods rest upon the interpretation of **Bayes' theorem as a rule** for updating *belief* following some experience. The theorem is a straightforward consequence of the fact that the joint probability $P(A, B)$ of two events (or situations) A and B can be written as $P(A, B) = P(A \mid B)P(B)$ and also as $P(B \mid A)P(A)$. Hence $P(A \mid B)P(B) = P(B \mid A)P(A)$, or: $$P(A \mid B) = P(A)\frac{P(B \mid A)}{P(B)}$$ Then **the rule** is the following: "If $P(A)$ is someone's (or: a system's) *'strength of belief* (a real number between 0 and 1) *in A to be the case', prior* to observing B then, after observing B (or: *posterior* to B), the strength of his (or: its) belief in A should be $P(A \mid B)$, that is, the prior strength times the *normalised likelihood* of observing B given that A is indeed the case."
As a result of this "change of mind", actions can be taken – which takes us straight back to some of the learning scenarios outlined in this section. (See e.g., [Wolpert] for a general explanation of Bayes' rule, and [D'Agostini] for a more elaborate Bayesian Primer.)

Table 6: Kernel Methods and Bayes' rule

Kernel methods and Bayesian methods; see Table 6, page 88.) Presenting in this note only some of these concepts and tools in any detail would carry us way too far afield. The literature (even on individual topics) is vast and the number of relevant journals and conferences keeps growing[135]. Introductory texts, such as [Alpaydin] can only cover a relatively limited scope. Much has been made ac-

[135] See for instance: `http://jmlr.csail.mit.edu/`(Journal of Machine Learning Research), `http://www.kernel-machines.org/`, `http://www.machinelearning.org/`(The International Machine Learning Society), `http://nips.cc/`(Neural Information Processing Systems Foundation)

cessible on the Internet though, for instance through PASCAL[136], a *Network of Excellence* funded by the European Commission under FP6. PASCAL connects more than fifty research groups in Europe. Its website gives access to one of the richest repositories of publications pertaining to relevant domains, and to a large number of video lectures on various themes of interest to the Machine Learning community and beyond.

There are of course other research areas, not fully covered by this initiative, that are also of considerable relevance to the general topic of this section. They include *evolutionary* and *genetic* algorithms and *"complex adaptive systems"*, as already mentioned in section 5.2. Unlike statistical and computational learning the methods and approaches subsumed under these headings are more or less directly inspired by observing and studying long-term processes of self-modification and adaptation in biological systems. They have a long history, some dating back to the pioneering work on Cybernetics in the 40's and 50's of the previous century. Although there are many far reaching theoretical results major contributions to mainstream engineering of data and signal processing systems are still relatively scarce.

The spectrum of learning methods and applications, covered by the set of currently funded FP6-IST Cognitive Systems projects, in many ways also extends beyond the compass of classical Machine Learning, as a more detailed analysis would readily reveal. (A single project alone may involve many instances where some form of learning is required; see for example the Robot-Cub scenarios in Table 4, page 49.) There are modes of natural learning still to be captured in formal mathematical terms, and to be transposed to the domain of artificial systems: learning by observation and imitation for instance, the incorporation of meaning that happens as animals and humans begin to master the repertoire of actions they can safely perform, the socially conditioned construction of meaning when humans acquire their first language (but also at later stages), and so on. Learning object and environment affordances, the creation of symbols and their association with signals (emanating from some environment, involving all sensory modes), most certainly requires more than one method and more than one representational structure. Human learning may also take the form of constructing and accumulating retrievable propositions about the world (as in encyclopedic ontologies of the type mentioned in section 6.2.2, page 78) although, of course, no one would ever be able to list the entire world knowledge he or she might have gained in this way. Yet, our individual memories and histories and, indeed, our selfs, are the results of lifelong learning and adaptation processes.

"Can machine learning theories and algorithms help explain human learning?" is one of the questions Mitchell poses in the above mentioned White Pa-

[136] Pattern Analysis, Statistical Modelling and Computational Learning, `http://www.pascal-network.org`

per, proposing to expand and augment the Machine Learning research agenda ([Mitchell]). And with an eye on learning in artificial systems he asks: *"What is the relationship between different learning algorithms, and which should be used when?", "How can we transfer what is learned for one task to improve learning in other related tasks?", "Can we build never-ending learners?", "Will computer perception merge with machine learning?", "Can we design programming languages containing machine learning primitives?"*

There are many more questions that require answers which go beyond what Machine Learning theory currently has to offer. *"How can an artificial system improve the quality of its service(s) through interaction with its users?"*, for example, comprises a variety of problems which have yet to be solved through suitable models and theories. Ultimately, we would like to know *"What dynamic structures can learn (or adapt to) what, and at what cost, in a given dynamic environment?"*. In spite of early (and highly influential!) attempts to formulate a *"Theory of the Learnable"* ([Valiant]) we are still a long way from having sufficiently general, yet meaningful, (mathematical, formal) frameworks at our disposal in which to address this question; and we certainly are still a long way from having machines that *"program* (let alone, *enact!*) *themselves"*.

6.3.2 Beyond discrete patterns and computation

In this section we revisit some of the questions raised in section 3, but with a focus on concrete current attempts to find more or less satisfactory answers. We thus also continue and deepen the discussion in sections 5.2 and 5.3, of models and their implementations respectively.

Can artificial mechanisms generate lifelike behaviour? Does matter matter? Is cognition at its lowest level computation? Can it be emulated? Does a living cell compute? And if so, what and how? Or are there fundamental differences between living organisms (or "life as we know it") and artificial mechanisms, no matter how sophisticated the latter? Is organisation different from computation?

We do not know (yet) the answer to the first question. But we now do believe that matter matters. Or this is what the proponents of new AI (see section 6, page 43) keep telling us: at least the body matters, the system's body with all its parts that determine what the system can do and the sort of knowledge it can attain - if any.

Further to our somewhat polemical remarks in section 5.2, we may deem this insight so obvious that it is indeed surprising that it took so long to become seriously mainstream (or almost) - in the sense that it does not put academic careers, in Computer Science, for example, in jeopardy any more. Computer Scientists above all should be familiar with the notions of "function depending on structure", "structure allowing for function", and "function (re-)building structure". Examples from this and other disciplines abound, involving either material

or abstract entities or both: processor and memory chips supporting computation; data structures for efficient sorting, searching, and arithmetics; random access versus serial access storage; roman numerals versus place systems; brains enabling perception, reflection, and action (and different brains doing it differently); proteins and enzymes catalysing chemical reaction, depending on their 3-dimensional foldings, and on and on. Structures may be abstractly conceived or described (e.g., in terms of relations between components): they are functional in the real world if and only if they are instantiated in a physical substrate that affords them. Function results from the interplay of material components in material structures which in turn are contingent on what their components admit.

Of course, within the context of discussing the possibility of artificial cognitive systems, the above questions only make sense if we assume that "true" cognition can only be present in living systems. This assumption seems to underly the enactivist's approach. But, as already pointed out in section 5.2, the enactivist's agenda, largely based on the alleged interdependence of material structure and concrete function, is not easy to implement. One of the reasons (if not the main reason) may be our lack of knowledge as to how to create lifelike artificial organisations and mechanisms or, rather, to emulate life.

Indeed, there appears to be no way of emulating life in an artificial substrate unless we gain a better understanding of life itself. What distinguishes life from non-life? Robert Rosen, the theoretical biologist whose (M,R)-Systems[137] can be considered a formalisation of Maturana and Varela's concept of autopoiesis[138] suggested the following necessary (differentiating) condition: *"a biological system (an organism) is closed with respect to efficient causation"*. In other words: whatever happens within the system is caused (or: entailed) by some function of, by some process inside the system itself. Clearly, "closure" does not mean that the system is not in contact with its environment - all organisms are. Rather, the term "closure" refers to the above mentioned closed loop which inextricably connects the structure and function of individual living entities. Rosen distinguishes between *analytic* and *synthetic* models of function-performing systems. Essentially, synthetic models are derived from externally observable input-output relations whereas analytic models are based on probing processes internal to the system. To make his argument more precise Rosen phrases it in category theoretical terms[139]. These same terms can be used to model biological systems on a highly abstract level[140] . Rosen and his interpreters (e.g., [Wolkenhauer1]) arrive at the conclusion that living organisms have *impredicative, non-simulable* (or: non-computable) analytic models (i.e., models for which no synthetic equiv-

[137] *"Metabolism, Repair"*, [Rosen1]; see also [Joslyn]
[138] see section 5.2, and [Maturana]
[139] see for instance [MitchellB], an introductory text on Category Theory
[140] see for instance [Wolkenhauer1] for models of cells and their metabolic pathways

alent can be constructed), this being their distinguishing mark with respect to non-living matter (presumably including all man-made artificial machines). The latter is (are) called *"simple"*; in contrast, living organisms are *"complex"*, by virtue of not being fully simulable. *"Life"*, according to Rosen, is not computable, and the only fully functional synthetic *"model"* of a living entity is that entity itself.

Rosen's approach and conclusions (in particular regarding the "non-computability of life", and the possibility or impossibility of artificial life) are not uncontested and remain subject to debate ([Chu], [Louie], [Wolkenhauer2]). Yet one may argue that, irrespective of its level of sophistication and refinement, any analytic or synthetic (in the above sense) model of a living system - and hence any possible implementation based on it - must ignore details (for instance through discretisation of intrinsically analogue phenomena, through preconceived ways of dealing with non-determinism, etc.) that may well be crucial. Whether on theoretical principle or for practical limitations: models of physical phenomena seem to be necessarily incomplete (Physics is rife with examples, one of the commonest being the *point mass*). Numerical modelling in particular - as required for computer simulation - is necessarily approximative.

But we do not have to model everything and it would probably be preposterous to equate artificial life to "life itself" (just as it seems rather inappropriate to equate human intelligence to artificial intelligence). Moreover, we must bear in mind that the discipline of Artificial Life is not limited to simulating *"in silico"* life-like processes but may well extend to the fabrication of (not necessarily discrete) structures that can support and (at least) emulate such processes[141]. This leaves the possibility of artificial life-like material structures, resulting from environmentally conditioned self-organisation, squarely on the research agenda.

This also indicates a strong link between Artificial Life research and the embodiment paradigm. Hence, if we postulate (whatever form of) embodiment as a prerequisite for cognition then two obvious questions to ask are: At what bodily level should we start modelling and implementing? At the level of bones, tendons and muscles, at the level of cells and tissues, or even at the molecular level? And: At what bodily level should we start seeking for cognitive features? These questions concern both, the body in its entirety as well as the body's sub-structures (e.g., brain, CPU, sensors, ...) that control its actions. A third question arises, possibly related to the previous two: Are there degrees of life-likeness, corresponding to physically instantiated cognitive capabilities?

Questions of this sort, including those asked at the end of section 5.3, are now defining research areas, such as *morphological computation* ([Paul], [Pfeifer1]) and *amorphous computing* ([Abelson]), that may have a considerable impact on future developments in (industrial) robotics across a wide range of scales, and

[141] cf., the discussion in section 5.3

in *"smart" materials*. However, the terms *"computation"* and *"computing"* seem somewhat unfortunate here as both areas centre on the *physical configuration* of *physical objects*, either on the process itself or its results, or on both.

Morphological "computation" is about exploiting the structural and physical properties of composite objects in order to make them behave – for instance move – in a desired way. It is, in other words, about offloading any *"algorithmic toil"* that would otherwise be needed, onto physical processes that are peculiar to "the shape of things", the material things are made of, and the way they are put together[142]. This approach is clearly motivated by the observation that certain capabilities (in particular at the sensory-motor level) of living creatures are largely determined by the anatomy and basic physiology of their sensor and actuator organs. The *"Yokoi hand"* [143] is a particularly nice example of how morphology can – biomimetically, so to speak – be taken advantage of in order to reduce the computational effort in controlling the grasping of all sorts of objects of different shape and size[144].

Ideas underlying *"amorphous computing"* can be traced back (at least) to the early fifties of the 20th century when John von Neumann introduced his cellular automata[145]. These automata are bound to a rigid array with well-defined neighbourhood constraints. By contrast, Abelson, Beal and Sussman describe the goal of amorphous computing as follows:

> *"... to identify organizational principles and create programming technologies for obtaining intentional, pre-specified behavior from the cooperation of myriad unreliable parts that are arranged in unknown, irregular, and time-varying ways. The heightened relevance of amorphous computing today stems from the emergence of new technologies that could serve as substrates for information processing systems of immense power at unprecedentedly low cost, if only we could master the challenge of programming them.* (cf., [Abelson])*"*

We came across similarly far reaching general ideas and ambitions in previous sections, in particular at the end of section 6.1.2, under the label of *autonomic* and *organic computing*. We may include others, such as those underlying *swarm robotics*, mentioned in sections 3 and 6.1.1. While these ideas differ in many ways, their common thrust is the quest for engineering principles that guarantee system robustness at the hardware and connectivity level, through several of the

[142] Soap film experiments make for beautiful examples of "physical computation"; minimal surfaces can be found in this way, as well as solutions to so called *Steiner problems* (finding minimal length trees spanning a given number of points). [Hildebrandt] provides a luxuriously illustrated account of such experiments, describing them in terms of the mathematical calculus of variations.

[143] described in [Pfeifer1]

[144] see also footnote 110, page 66, and the examples in section 6.2.1

[145] cf., section 5.2, page 34

self-X capabilities that are characteristic of even the most basic living - and hence, to some degree, cognitive - entities. The capabilities in question here are self-configuration and -reconfiguration, self-healing, self-defence (as per immune subsystems, for example) and, to the extent implied by these processes, self-control, i.e., systems ought to have a certain level of autonomy. Programming such systems or rather, the evolution of such systems and the emergence of desirable properties and functions, comes down to defining the interactions its components can engage in, including of course, the constraints these interactions are subjected to.

BIOLOCH (FP5)	**Bio-mimetic Structures for Locomotion in the human body** Micro-robots with wormlike morphology for applications to minimally invasive endoscopy (see also footnote 88, page 55, and [LaSpina]), `http://www.ics.forth.gr/bioloch/`
SWARM-BOTS (FP5)	**Swarms of self-assembling artefacts** A set of components that self-assemble into a shape-changing swarm-bot navigating on rough terrain, `http://www.swarm-bots.org/`
HYDRA (FP5)	**"Living" Building Blocks for Self-designing Artefacts** Self-(re)configuring robots consisting of adaptive components modelled after living cells, `http://hydra.mip.sdu.dk/`
POEtic (FP5)	**Reconfigurable POEtic Tissue** *"... development of a flexible computational substrate inspired by the evolutionary, developmental and learning phases in biological systems."* ([Tyrrell]) `http://www.poetictissue.org/`
I-Swarm	**Intelligent Small World Autonomous Robots for Micro-manipulation** Micro-robots exhibiting biologically realistic swarm behaviour (like ants or bees) through self-organisation `http://i60p4.ira.uka.de/tiki/tiki-index.php?page=I-Swarm`
PACE	**Programmable Artificial Cell Evolution** Nanoscale emulation of cell-like structures, exploiting programmable molecular mechanisms `http://www.istpace.org/`

Table 7: (Some relevant) "Beyond Robotics" and "Complex Systems" projects

There are and have been many research projects in Europe and beyond, aiming to reach goals of the above described kind. Most of the larger European ones are - one might say: traditionally - being funded under the FET part (*"Future and Emerging Technologies"*) of the European Union's research programmes. They mainly figure under the *"Beyond Robotics"* and *"Complex Systems"* sub-programme headings (of the FP6-FET agenda). Table 7 gives an

overview of some of these projects. Some of them demonstrate a close relation between *amorphous* and *morphological* "computation": a specific morphology may arise from the interplay between "amorphously", spatially distributed parts. Our selection spans the whole range from molecular (or: nano-size) to macroscopic components.

PACE, at the low end of this range, applies microfluidic (nano-)techniques to "programming" molecular reactions and priming an evolution of artificial cells and cell assemblies. It is, in that regard, an Artificial Life project *par excellence*[146]. As such it is in the long tradition of research into the most basic mechanisms of life's self-organisation ([Eigen]). Neighbouring - and partly overlapping - research areas are *"Supramolecular Chemistry"* ([Lehn]), *"Chemical Computing"* ([Matsumaru], [Zauner]; see also footnote 56, page 41, as well as the by now classical work on DNA computing, first documented in [Adleman]) and, somewhat more distant, *Molecular Electronics*, *Nanoelectronics* (both mentioned earlier, in section 5.3, page 39) and *Nanorobotics*[147], where (controlled, targeted) self-organisation is becoming a key engineering technique ([Lu], [Balzani]).

On the more theoretical side, the study of cellular processes has given rise to new computational models such as *"Membrane Computing"* (also known as P-systems, [Paun2]), which can be considered abstractions from the physico-chemical action observed in natural cells and cell assemblies. They are in the tradition of what may be called *"Natural Computing"* ([Paun]) of which Lindenmayer-systems (L-systems), modelling the growth of plants (see [Prusinkiewicz]), are earlier representatives.

Accepting the embodiment paradigm takes us beyond the level of cells, cell assemblies and the specificities of the organs (or tools) that make (or help) us sense and act. Proponents of that paradigm maintain - and presumably rightly so - that the qualities and scope of cognitive processes do not only depend on the material characteristics and the affordances of the body in which they are running but also on the way these processes are physically instantiated within that body. In virtually all living organisms that exhibit intricate behavioural traits and "intelligence" (humans and some animals in particular) this instantiation is the *Central Nervous System*, with a *brain* as its principal component. It implements mechanisms for establishing and recognising patterns pertinent to a given environment. In humans it brings about the well known *"high-level"* cognitive capabilities, such as conceptualisation, reasoning, planning, symbolic communication, intelligent control, goal-oriented and social behaviour, and ulti-

[146] ECLT, the European Center for Living Technology, is a spin-off of this project, `http://www.ecltech.org/`.

[147] See the discussion in section 5.3, page 53. BIOMACH (Design and nano-scale handling of biological antetypes and artificial mimics, `http://www.biomach.org/`) is one of the largest FP6-IST projects in this domain.

mately, some form of self-understanding and, through various modes of learning, self-extension. It drives the most advanced (natural) cognitive system to date.

It is therefore only reasonable to ask what properties of the nervous system make it so powerful in terms of cognitive performance. What is so peculiar about it? To what extent do the qualities of cognition hinge on these properties, for instance on our brains' inherent plasticity or on endocrinal activity? Can correspondences be identified between types of behaviour or capabilities (e.g., language, self-awareness) and neural processes? We have discussed some of these questions in section 5.1. Here we may ask again: "What use is actually being made in designing and implementing cognitive architectures, of insights into the way (mammalian and other types of) brains work?"

It is most likely (and luckily!!) true that *"current and future machines for detecting and measuring physical and chemical processes in brains will no more reveal the contents of virtual machines running on brains than electronic probes can reveal the contents of virtual machines running in computers"*[148]. It is, however, quite a different kettle of fish to learn more about the workings of the brain's "*wetware*", about the functional "circuits" and "electronics" of the brain (rather than the content of neural processes), in order to apply that knowledge to derive computer implementable cognitive architectures or to construct neuromorphic hardware that may directly support cognitive capabilities such as memory and learning. After all, none of our current hardware technologies can offer anything close to the properties that are typical of the physical substrate of cognition in natural organisms (e.g., energy efficiency, compactness, plasticity, and fault tolerance). Moreover, given that traditional hardware does not seem to lend itself particularly well to implementing capabilities like vision, we may assume that it is precisely the way the brain is built and the way it works that make our (and our fellow animals') much acclaimed cognitive feats possible[149].

Table 8 lists ten FP6-IST projects that pursue goals along the above lines. The first seven projects on that list have been or are being funded under the Cognitive Systems part of the FP6-IST programme. They have been mentioned and commented on elsewhere in this note (see section 6.1.1). They fit into the present context as they comprise work that largely relies on neuro-modelling at the (relatively) high level of functional architectures, and its translation into

[148] Aaron Sloman at the AINC Seminar at the University of Birmingham's School of Computer Science, 5th March 2007 (`http://www.cs.bham.ac.uk/research/projects/cosy/presentations/sloman-evolang.pdf`). We prefer to substitute "processes" for "virtual machines".

[149] A strong case has been made in [Trehub], in favour of studying the ways of the brain with a view to deriving implementable models - especially as far as vision is concerned.

BACS	**Bayesian Approach to Cognitive Systems**, Bayesian-directed analysis and implementation of neural computation, http://www.bacs.ethz.ch/
GNOSYS	**An Abstraction Architecture for Cognitive Agents**, A computational architecture abstracted from observable brain processes, http://www.ics.forth.gr/gnosys/
ICEA	**Integrating Cognition, Emotion and Autonomy**, Robot control based on an architecture derived from the anatomy and physiology of the rat brain, http://www.iceaproject.eu/
PACO-PLUS	**Perception, Action and Cognition through Learning of Object-Action Complexes**, http://www.paco-plus.org/PACO-PLUS
POP	**Perception On Purpose**, Implementation of an architecture for active perception based on behavioural and neuroimaging studies, http://perception.inrialpes.fr/POP/
ROBOT-CUB	**Robotic Open-architecture Technology for Cognition, Understanding and Behaviours**, A platform for implementing cognitive architectures, including brain-based ones, http://www.robotcub.org/
SENSOPAC	**SENSOrimotor structuring of Perception and Action for emerging Cognition**, Neuromodelled computational architecture realising haptic cognition, http://www.sensopac.org/
SPARK	**Spatial-temporal patterns for action-oriented perception in roving robots**, Insect-brain inspired hardware and software architecture realising perception-action loops in robots, http://www.spark.diees.unict.it/
CILIA	**Customized Intelligent Life-Inspired Arrays**, Neuromorphic sensing hardware mimicking various forms of cilia based systems occurring in nature, http://www.cilia-bionics.org/
DAISY	**Neocortical Daisy Architectures and Graphical Models for context-dependent Processing**, Hybrid (digital/analogue) hardware mimicking the "daisy architecture" of the neocortex, http://daisy.ini.unizh.ch/
FACETS	**Fast Analog Computing with Emergent Transient States in Neural Architecture**, Emulation in hybrid neuromorphic hardware, of processes observed in biological nervous systems, http://www.facets-project.org/

Table 8: Neuro-modelling and neuromorphic hardware in FP6-IST projects

software. The sole exception here is SPARK, which utilises specific hardware implementing CNNs[150] for vision-based sensory-motor control and navigation.

The remaining three projects in Table 8 are being funded under the *"Bio-I3"* (*Biologically Inspired Intelligent Information Systems*) theme of FET (see above). They are all strongly hardware oriented. FACETS is the largest and perhaps most ambitious one. One of its stated goals is to physically emulate (or: reproduce) in silico, through hybrid VLSI ASIC[151] technologies, aspects of the physiology of real neurons and their synaptic interconnections in real brains ([Grübl]). In FACETS and similar projects these techniques are being employed to create neuromorphic structures that adapt and learn [152]. This clearly differs from attempts to simulate the activity of brain structures, as undertaken for instance in the *Blue Brain*[153] project. The Blue Brain simulation relies on the computational power of an IBM *Blue Gene* supercomputer with 8192 processors, to run a software model of neocortical columns ([Brette]). It does not aim to design and build specific brain-like hardware.

The technologies underlying and developed further by the FACETS project, are certainly a far cry from reaching the "high level" of human cognition; yet they have – like the above mentioned CNNs - already served to improve greatly the capacities of professional and experimental computer vision systems. They date back to the early eighties of the 20th century when Carver Mead, one of the dominating figures in electronics hardware design, together with Lynn Conway, inaugurated the *"VLSI revolution"* ([Mead]), and subsequently used these techniques to produce chips that integrate the sensing and processing of audiovisual signals[154]. Implantable electronic devices taking the role of retinas or cochlears have since become possible.

Just as we considered the term *"computation"* somewhat inadequate in connection with qualifiers such as *"morphological"* and *"amorphous"*, one might have similar reservations regarding its use within the context of neuromorphism – given the mental bias that term connotes. Traditional computational paradigms are epitomised by Turing's (and equivalent) formalisms and the classical *von Neumann computer architecture*; they focus on (discrete) abstraction rather than on concrete (analogue) physical processes and, ultimately, on software rather than hardware. Even if Rosen were wrong and "life *is* computable" (i.e., reducible to a synthetic model that in principle could be run on an abstract Turing Ma-

[150] Cellular Neural/Nonlinear Networks, introduced by Leon O. Chua in the late eighties of the 20th century (see section 5.3, page 39, and [Chua])

[151] Very Large Scale Integration, Application Specific Integrated Circuit

[152] Analogue VLSI circuits have long since also enabled the implementation in silico of learning algorithms ([Mead2], [Arthur], [Hsu], [Cauwenberghs]), including Support Vector Machines ([Genov]).

[153] `http://bluebrain.epfl.ch/`; both projects are in rapport with each other.

[154] A company founded by Carver Mead that specialises in vision hardware is Foveon: `http://www.foveon.com/`

chine) we nonetheless could question (as in section 5.2) whether the prevailing computational metaphors are well suited to capturing the ways natural organisms incorporate their environment and act in it. Indeed, as we have seen, there is a growing consensus, exemplified through many of the projects discussed or mentioned in this note, on the all-importance of the physical structures and properties of machines when it comes to interacting with a physical environment that is characterised by nondeterministic analogue processes and asynchronous events. Many of the complexity issues inherent in abstract computation schemes seem to become irrelevant in "physical, analogue computation", as for instance in the example (Steiner problems) referred to in footnote 142 (page 93), or if performed by muscles, molecules, or spiking neurons[155].

Bearing in mind that brains are primarily controllers of the behaviour of their host bodies rather than stand-alone computers, it would seem more appropriate to view their external (observable) activity in terms of mappings between functions of a real variable (representing continuous time), with values in highly-dimensional event and action spaces respectively. Natural brain-body systems represent and realise such mappings. Whether this happens in principle through computation as we know it[156], is quite a different issue, and subject to partly philosophical, partly scientific, but always lively debate[157].

For *artificial "brain-body" systems* (i.e., cognitive systems) one may not wish to give up, harum-scarum, the idea of achieving behaviour control through some sort of computation, as already pointed out elsewhere in this note (see section 5.3, page 42). One should, however, replace models more suited for *"off-line"* computation (such as *Turing-Machines* and *"Random Access Machines"*) with frameworks that more readily accommodate *"on-line"* and *"real-time"* processing of environmental input streams, as argued for instance in [Maass]. In fact, Maass proposes such a framework, calling it the *"Liquid State Machine (LSM)"*, which is essentially a generalisation of classical finite state machines to continuous input streams and state spaces - hence the term *"liquid"* (ibid.). A similar approach, under the name of *"Echo State Networks (ESN)"* has been suggested independently in [Jaeger]. Both types of approaches are meant to capture key characteristics (plasticity, fault-tolerance, etc.) of biological neural networks. The LSM model in particular has been applied not only to understanding the dynamics of neocortical microcircuits, as called for in the DAISY and FACETS projects, but also to the design of the specific biomimetic (hybrid, fault-tolerant) VLSI chips the latter project is poised to explore. Moreover, references in both, [Maass]

[155] Another example is Watt's "Governor" - see footnote 102, page 60 - which essentially "embodies" the solution of a differential equation.

[156] i.e., through algorithms as captured by the Church-Turing thesis

[157] for some contributions to this debate see for example [Penrose], [Landau], [Calude], and [Siegelmann1]

and [Jaeger], point to applications to speech recognition, grammar learning, and robot control, inter alia.

More recently, LSMs and ESNs have been subsumed under the more general concept of *"Reservoir Computing (RC)"* ([Schrauwen]). RC encapsulates the key idea underlying both models: the separation of producing output streams from processing the input stream. The latter *"fills a reservoir"* (which can be represented as a sparsely connected random network of different kinds of *"neurons"*) that reverberates (*"echos"*!) the input world and makes it possible for *"readout functions"* to take account, at least to a certain extent, of the input history. The readout mechanism (usually also represented by specific *"neurons"*) can be (and usually is) subjected to efficient learning algorithms.

Finding out what constitutes a "good reservoir" and how to obtain one, is one of the current "hot" topics in this very active research area. It is here where RC, as and when applied to hardware design, could likely meet with other attempts to provide more malleable substrates for creating representations and generating behaviour. Amorphous computing and the approach taken by the above mentioned POEtic project (cf., Table 7) are but two of them. The latter, like other approaches, draws on the rich body of research on genetic and evolutionary algorithms and design strategies (for all sorts of ends) that has repeatedly been alluded to in this note[158].

This should not be too astonishing given that natural brain-body systems are themselves the results of the mechanical (mindless?) application of like strategies. Whether they work equally well in the world of artefacts still remains to be proven, in spite of many, sometimes immodest, claims that they in fact do work well. What appears to be certain, however, is the indissoluble unity of mind and body. Dualistic (*"Cartesian"*) approaches, strictly separating hardware from software, so characteristic of "old AI" paradigms, may indeed be fundamentally flawed – at least if and when they are being applied to understanding the phenomena of life and natural intelligence (in animals and humans). Taken to an extreme they nourish the *extropian* dream of transplanting ("uploading") an individual's mind onto some artificial substrate ([Minsky2]), a modern version perhaps of the immortal soul hovering in heaven, waiting for a new body to inhabit[159].

[158] The British *"Grand Challenge 7, Unconventional Computing"* (`http://www.cs.york.ac.uk/nature/gc7/`) also focuses on new computing paradigms, including those that are inspired by observing biological phenomena (see also [Paun]).

[159] Alternatively, pending further progress in various nanotechnologies, people might want to replace their natural brain with an artificial one. Which all leads to endless discussions about the intricate nature of Self (e.g., in [Hofstadter]).

Contrary to [Sloman2][160] it is more likely that *"evolution's discovery"* of how to grow bodies, and its discovery of how to grow minds pertaining to bodies, were indeed concomitant: Nature as we know it has no way of growing great minds except in the brains (and hosting bodies) of some of its creatures, commonly known as animals. Minds (great or small) cannot exist in abstracto. (Note that we took the liberty of translating *"virtual machines"* into *"minds"*.) This is not to say that computers cannot have minds; on the contrary: they do run the virtual machines software engineers build for them, and that can be very useful because they have a "body" (keyboard and screen, for instance). But computers and brains as we know them cannot run the same kinds of minds. Admittedly, this statement, as speculative as its negation, is not a theorem (unless Rosen, see above, is right). However, a brain, if only for practical reasons, cannot run a chess machine (but it can invent one!), and a computer will most likely never be able to run a chess player let alone be the inventor of a chess machine, or the *game of chess* in the first place[161]. On a somewhat different note we recall (cf., the discussion of ANNs in section 5.2, and [Siegelmann2]) that formally describable dynamic structures (e.g., *Analogue Recurrent Neural Networks*, ARNNs) do exist whose *computational* (or, rather, *"mapping"*) power goes provably beyond that of Turing Machines. The brain may well be an instantiation of such a structure. Whether some of its *functions* can be *approximated* or even fully *replicated* by computers (more or less as we know them) is indeed a separate question which can, nobody would seriously doubt it, be answered in the affirmative. (For more philosophically flavoured but probably no less convincing arguments see [Searle3].) To question the "computer-likeness" of brains does, by the way, not mean to fall into the old trap of wanting to keep Man at the centre of the universe. It means just the opposite: it puts Man right back where he belongs - to physical Nature, not to some ethereal realm of logics and abstractions.

[160] *"Application domains where tasks and environments are fairly static and machines need to be functional quickly, require precocial skills (possibly including some adaptation and self-calibration), whereas others require altricial capabilities, e.g. where tasks and environments vary widely and change in complex ways over time, and where machines need to learn how to cope without being sent for re-programming. Architectures, mechanisms, forms of representation, and types of learning may differ sharply between the two extremes. And the end results of altricial learning by the same initial architecture may differ widely.*

If all this is correct, it seems that after evolution discovered how to make physical bodies that grow themselves, it discovered how to make virtual machines that grow themselves. Researchers attempting to design human-like, chimplike or crow-like intelligent robots will need to understand how. Whether computers as we know them can provide the infrastructure for such systems is a separate question." (quoted from [Sloman2])

[161] In spite of some optimistic and possibly over-hyped allegations it is hard to imagine a Turing Machine that would make its substrate (1) create the symbol systems of the kind Nature has created through human brains (including Turing Machines), or (2) wonder what makes the sun, moon, and stars move.

Resuming a recurring theme of these notes we do agree that in order to *design human-like, chimplike or crow-like intelligent robots* we do need to know how Nature manages to co-evolve and simultaneously grow bodies and minds; but given the discussion and references especially in this last section we are - more than ever[162] - inclined to answer the question in the negative, *"whether computers as we know them can provide the infrastructure for such systems"*.

7 Postscript

So it may turn out that in order to make *substantial* progress in the design and implementation of artificial cognitive systems we may indeed have to reinvent life itself, as in the above mentioned PACE project, or in Synthetic Biology ([EUR]). This is risky.

We should be aware of an old fairy tale recounted by the Grimm Brothers ([Grimm]): *The Fisher and his Wife*.

The enchanted flounder the fisher had one day encountered in the placid sea and returned to its element, thankfully gave in to the wife's every demand, promoted her from the filthy shack where she used to live, all the way to the papal throne, and made her richer and richer, and more and more powerful. Then the woman wanted to be God. Upon hearing this the flounder told the fisherman: *"Go home. She is sitting in her filthy shack again."* And the sea was roaring like hell[163].

Acknowledgements: These notes result from almost four years involvement in the preparation and operation of the *Cognitive Systems* (and robotics) part of the European Commission's IST/ICT programmes. They draw on a number of workshops I had the pleasure to (co-)organise[164], on projects I came in contact with[165], on many discussions (on various occasions) with workers in pertinent fields and, last but not least, with my colleagues in the *Cognition Unit* of *Directorate General Information Society and Media* of the European Commission.

References

[Abelson] Abelson H. , Beal J. & Sussman G. J. (2007) Amorphous computing. Technical Report MIT-CSAIL-TR-2007-030

[Adleman] Adleman L.M. (1994) Molecular Computation of Solutions to Combinatorial Problems. Science 226, pp. 1021–1024

[Akyildiz] I.F. Akyildiz I. F., Su W., Sankarasubramaniam Y. & Cayirci E. (2002) Wireless sensor networks: a survey. Computer Networks 38, pp. 393-422

[162] see the remarks towards the end of section 2.

[163] `http://www.pitt.edu/~dash/grimm019.html`

[164] `http://cordis.europa.eu/ist/cognition/presentations.htm`

[165] `http://cordis.europa.eu/ist/cognition/projects.htm`

[Allen] Allen J. F., Byron D. K., Dzikovska M., Ferguson G., Galescu L. & Stent A. (2001) Toward Conversational Human-Computer Interaction. AI Magazine 22 (4), Winter 2001, pp. 27-38

[Alpaydin] Alpaydin E. (2004) Introduction to Machine Learning. The MIT Press, Cambridge, Massachusetts

[AndersonJR] Anderson J. R. (1996) ACT: A simple theory of complex cognition. American Psychologist, 51, pp. 355-365

[AndersonML] Anderson M. L. (2003) Embodied Cognition: A field guide. Artificial Intelligence 149, pp. 91-130

[Antoniou] Antoniou G. & van Harmelen F. (2004) A Semantic Web Primer. Boston: MIT Press

[Antsaklis] Antsaklis P. J. & Passino K. M., eds. (1993) An Introduction to Intelligent and Autonomous Control. Norwell, MA: Kluwer Academic Publishers

[Arbib] Arbib M. A. (2002) Beyond the mirror system: imitation and evolution of language. In: Imitation in Animals and Artifacts (C. Nehaniv, K. Dautenhan, editors). pp. 229-80. Cambridge MA: MIT Press

[Aristotle] Aristotle (350 BC) De Anima (On the Soul. Translated by J. A. Smith, `http://classics.mit.edu/Aristotle/soul.html`

[Arthur] Arthur J.V. & Boahen K. (2006) Learning in Silicon: Timing is Everything. Advances in Neural Information Processing Systems 17, B. Schölkopf and Y. Weiss, Eds, pp. 75-82, MIT Press

[Asada] Asada M., MacDorman K.F., Ishiguro H. & Kuniyoshi Y. (2001) Cognitive Developmental Robotics As a New Paradigm for the Design of Humanoid Robots. Robotics and Autonomous Systems, 2001. 37: pp. 185-193

[Ashby1] Ashby W. R. (1947) Principles of the Self-Organizing Dynamic System. Journal of General Psychology, Volume 37, pages 125–128

[Ashby2] Ashby W. R. (1952) Design for a Brain: The Origin of Adaptive Behavior. London: Chapman&Hall

[Axelsson] Jakob Axelsson J. (2001) Unified Modeling of Real-Time Control Systems and their Physical Environments Using UML. In Proc. 8th International Conference on the Engineering of Computer Based Systems, pp. 18-25, Washington, April 17-20, 2001

[Balzani] Balzani V., Credi A. & Venturi M. (2007) Molecular devices and machines. Nanotoday, Vol 2 No 2, pp. 18-25

[Bar-Yam] Bar-Yam Y. (2003) When Systems Engineering Fails — Toward Complex Systems Engineering. International Conference on Systems, Man & Cybernetics, 2003, Vol. 2, 2021- 2028, IEEE Press, Piscataway, NJ

[Bateson] Bateson G. (1972) Steps to an Ecology of Mind. Chandler Publishing

[Bauckhage] Bauckhage C., Hanheide M., Wrede S., Käster T., Pfeiffer M. & Sagerer G. (2005) Vision Systems with the Human in the Loop. EURASIP Journal on Applied Signal Processing 14, pp. 2375-2390

[Beckett] Beckett P. & Jennings A. (2002) Towards Nanocomputer Architecture. Seventh Asia-Pacific Computer Systems Architecture Conference (ACSAC'2002), Melbourne, Australia. Conferences in Research and Practice in Information Technology, Vol. 6., Feipei Lai and John Morris, Eds. (`http://citeseer.ist.psu.edu/beckett02towards.html`)

[Beer1] Beer R. D. (2004) Autopoiesis and cognition in the game of Life. Artificial Life 10(3), pp. 309-326

[Beer2] Beer R. D. (2006, in press) Beyond control: The dynamics of brain-body-environment interaction in motor systems. To appear in D. Sternad (Ed.), Progress in Motor Control V: A Multidisciplinary Perspective. Berlin: Springer (`http://vorlon.cwru.edu/~beer/Papers/PMCChapter.pdf`)

[Behnke] Behnke S. (2006) Online trajectory generation for omnidirectional biped walking. In Proceedings of IEEE International Conference on Robotics and Automation (ICRA'06), Orlando, Florida (`citeseer.ist.psu.edu/behnke06online.html`)

[BenJacob] Ben-Jacob E., Becker I., Shapira Y. & Levine H. (2004) Bacterial linguistic communication and social intelligence. Trends in Microbiology, Vol. 12, No. 8

[Benson] Benson S. & Nilsson N. (1995) Reacting, Planning and Learning in an Autonomous Agent. In Furukawa, K., Michie, D., and Muggleton, S., (eds.), Machine Intelligence 14, Oxford, The Clarendon Press

[Benzmüller] Benzmüller C., Horacek H., Kruijff-Korbayová I., Pinkal M., Siekmann J. H. & Wolska M. (2005) Natural Language Dialog with a Tutor System for Mathematical Proofs. Cognitive Systems 2005, Lecture Notes in Computer Science 4429 Springer 2007, pp. 1-14

[Biewener] Biewener A.A. (2002) Future directions for the analysis of musculoskeletal design and locomotor performance. J. Morph. 252, pp. 38-51.

[Bluff] Bluff L A, Weir A A S, Rutz C, Wimpenny J H & Kacelnik A (2007) Tool-related cognition in New Caledonian crows. Comparative Cognition & Behavior Reviews 2: 1-25

[Boahen] Boahen K. (2005) Neuromorphic Microchips. Scientific American, vol 292, no 5, pp. 56-63

[Bongard] Bongard J. C. & Pfeifer R. (2001) Repeated Structure and Dissociation of Genotypic and Phenotypic Complexity in Artificial Ontogeny. In: Spector, L. et al (eds.), Proceedings of The Genetic and Evolutionary Computation Conference, GECCO-2001. San Francisco, CA: Morgan Kaufmann publishers, pp. 829-836

[Bongard1] Bongard J. C., Zykov V. & Lipson H. (2006) Resilient machines through continuous self-modeling. Science, Vol. 314. no. 5802, pp. 1118 - 1121

[Breazeal] Breazeal C. (2005) Cynthia Breazeal: Socially intelligent robots. Interactions 12(2): 19-22

[Brette] Brette R., et al (21 authors) (2006) Simulation of networks of spiking neurons:A review of tools and strategies. Journal of Computational Neuroscience 23, pp. 349–398

[BromBryson] Brom C. & Bryson J. (2006) Action selection for Intelligent Systems. euCognition White Paper (`http://www.eucognition.org/asm-whitepaper-final-060804.pdf`)

[Brooks1] Brooks R. A. (1986) A Robust Layered Control System for a Mobile Robot. IEEE Journal of Robotics and Automation, Vol. 2, No. 1, pp. 14-23

[Brooks2] Brooks R. A. (1991) The Role of Learning in Autonomous Robots. Proceedings of the Fourth Annual Workshop on Computational Learning Theory (COLT '91), Santa Cruz, CA, Morgan Kaufmann Publishers, pp. 5-10.

[Brooks3] Brooks R. A., C. Breazeal, M. Marjanovic, B. Scassellati, M. Williamson (1999) The Cog pro ject: building a humanoid robot. in: Computation for Metaphors, Analogy, and Agents. C. Ne- haniv (ed), Lecture Notes in Artificial Intelligence 1562. New York, Springer, pp. 52-87

[Buchanan] Buchanan B. G. & Shortliffe E. H. (1984) Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project. Reading, MA, Addison-Wesley

[Burks] Burks A. W. (1970) Essays on Cellular Automata. University of Illinois Press

[Bush] Bush V. (1945) As We May Think. Atlantic Monthly

[Butterfass] J. Butterfaß J., Grebenstein M., Liu H. & Hirzinger G. (2001) DLR-Hand II: Next Generation of a Dextrous Robot Hand. Proc. IEEE Intern. Conf. on Robotics and Automation; Seoul, Korea, 21-26 May 2001

[Calude] Calude C., Campbell D. I., Svozil K., Stefanescu D. (1995) Strong determinism vs computability. In: W. Depauli-Schimanovich, E. Koehler, F. Stadler (eds.). The Foundational Debate, Complexity and Constructivity in Mathematics and Physics, Kluwer, Dordrecht, pp. 115-131

[Cangelosi] Cangelosi A. & Riga T. (2006), An embodied model for sensorimotor grounding and grounding transfer: Experiments with epigenetic robots. Cognitive Science, 30(4), pp. 673-689

[Capek] Capek Karel (1921) R.U.R (Rossum's Universal Robots), Pocket Books (see also: `http://capek.misto.cz/english/robot.html`)

[Cauwenberghs] Cauwenberghs G. (1996) Adaptation, Learning and Storage in Analog VLSI. Proc. 9th Annual IEEE International ASIC Conference, pp. 273-278

[Chaitin] Chaitin G. J. (2004) Leibniz, Information, Math and Physics. In: Wissen und Glauben / Knowledge and Belief. Akten des 26. Internationalen Wittgenstein-

Symposiums 2003. Herausgegeben von Löffler W. / Weingartner P. pp. 277-286 Wien: ÖBV & HPT (http://www.cs.auckland.ac.nz/CDMTCS/chaitin/kirchberg.html)

[Chater] Chater N., Tenenbaum J. B. & Alan Yuille A. (2006) Probabilistic models of cognition: Conceptual foundations. Trends in Cognitive Sciences Vol.10 No.7

[Chu] Chu D., Ho W. K. (2006) A Category Theoretical Argument against the Possibility of Artificial Life: Robert Rosen's Central Proof Revisited. Artificial Life 12.1, pp. 117–134

[Chua] Chua L.O. & Yang L. (1988) Cellular neural networks. IEEE Trans. On Circuits and Systems, 35 (10), pp. 1257-1290

[Clancey] Clancey W..J. (1993) Situated action: A neuropsychological interpretation (Response to Vera and Simon). Cognitive Science 17(1): 87-107.

[Clark1] Clark A. (1998) Where Brain, Body, and World Collide. Daedalus, Journal of the American Academy of Arts and Sciences, issue "The Brain", Vol. 127, No. 2

[Clark2] Clark A. (1998) Being There: Putting Brain, Body, and World Together Again. Boston: MIT Press

[Costa] Costa P. C. G. & Laskey, Kathryn B. (2006) PR-OWL: A Framework for Probabilistic Ontologies (ABSTRACT). Proceedings of the International Conference on Formal Ontology in Information Systems (FOIS 2006)

[Crowder] Crowder R. (2006) Toward Robots That Can Sense Texture by Touch. Science 312 (5779) pp. 1478-1479

[D'Agostini] D'Agostini G. (2003) Bayesian reasoning in data analysis - A critical introduction. World Scientific Publishing

[Damasio] Damasio A. (1999) The Feeling of What Happens: Body, Emotion and the Making of Consciousness. Heinemann, London

[Danesi] Danesi M. (1994) Messages and Meanings: An Introduction to Semiotics. Toronto: Canadian Scholars' Press.

[Dechter] Dechter R., Meiri I. & Pearl J. (1991) Temporal constraint networks. Artificial Intelligence 49, 61-95.

[de Duve] de Duve C. (1995) The Beginnings of Life on Earth. American Scientist September-October 1995 (http://www.americanscientist.org/template/AssetDetail/assetid/21438?fulltext=true&print=yes)

[di Primio] di Primio F., Müller B. S. & Lengeler, J. W. (2000) Minimal Cognition in Unicellular Organisms. In: J.-A. Meyer, A. Berthoz, D. Floreano, H. L. Roitblat and S. W. Wilson (Eds). SAB2000 Proceedings Supplement, International Society for Adaptive Behavior, Honolulu, pp. 3-12 (http://www.ais.fraunhofer.de/~diprimio/publications/diprimio_MinCog.pdf)

[Dorigo] Dorigo M. (2005) SWARM-BOT: An experiment in swarm robotics. In: Proceedings of 2005 IEEE Swarm Intelligence Symposium, Arabshahi P. & Martinoli A. (edt.), IEEE Press, Piscataway, NJ, pp. 192-200

[Dreyfus] Dreyfus H. (1979) What Computers Can't Do: The Limits of Artificial Intelligence (2nd edition). New York, Harper and Row / (1992) What Computers Still Can't Do: A Critique of Artificial Reason. Cambridge, Mass., MIT Press

[Dupriez] Dupriez C. (2003) Multilingual Vocal Access to Databases: the Central Role of Ontology Management in VocaBase. Sixth International Protégé Workshop, England, 2003

[Eigen] Eigen M. (1971) Selforganization of Matter and the Evolution of Biological Macromolecules, Die Naturwissenschaften, Vol. 58, pp. 465-522

[EUR] EUR21796 (2005) Synthetic Biology - Applying Engineering to Biology. Report of a NEST High-Level Expert Group. Office for Official Publications of the European Communities, Luxembourg

[Favareau] Favareau D. (2002) Beyond self and other: On the neurosemiotic emergence of intersubjectivity. Sign Systems Studies 30.1

[Felsberg] Felsberg F., Forssén P.-E., Moe A. & Granlund G. (2005) A COSPAL Subsystem: Solving a Shape-Sorter Puzzle. AAAI Technical Report Series FS-05-05

[Firman] Firman K. (2005) A Molecular Magnetic Switch that links the Biological and Silicon Worlds. IST-FET Newsletter, Vol. 1, p. 4.

[Floridi] Floridi L. (2004) Open problems in the philosophy of information. Metaphilosophy, Vol 35, No 4

[Forssén] Per-Erik Forssén P.-E., Johansson B. & Granlund G. (2006) Channel Associative Networks for Multiple Valued Mappings. Proc. 2nd International Cognitive Vision Workshop, Graz, Austria, pp. 4-11

[Forsyth] Forsyth D.A. & Fleck M. M. (1999) Automatic Detection of Human Nudes. International Journal of Computer Vision , 32, 1, pp. 63-77

[Frasca] Frasca M, Arena P. & Fortunato L. (2004) Bio-inspired Emergent Control of Locomotion Systems. World Scientific Series on Nonlinear Science

[Freeman] Freeman W. J. (2004) How and Why Brains Create Meaning from Sensory Information. International Journal of Bifurcation & Chaos 14, 513-530

[Gärdenfors] Gärdenfors P. (1999) Cognitive science: from computers to anthills as models of human thought. In: World Social Science Report, UNESCO Publishing/Elsevier, pp. 316-327

[Gelenbe] Gelenbe E., Lent R. & Xu Z. (2001) Design and performance of cognitive packet networks. Performance Evaluation, Volume 46, Issues 2-3, 155-176

[Genov] Genov R. & Cauwenberghs G. (2003) Kerneltron: Support Vector "Machine" in Silicon. IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 14, NO. 5, pp. 1426-1434

[Gibson] Gibson J. .J. (1977) The theory of affordances. In: R. Shaw & J. Bransford (eds.), Perceiving, Acting and Knowing. Hillsdale, NJ: Erlbaum

[Gödel] Gödel K. (1931) Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme. Monatshefte für Mathematik und Physik, Vol. 38, pp. 173-198

[Goodwin] Goodwin C. & Russomanno D. (2006) An Ontology-Based Sensor Network Prototype Environment. Fifth International Conference on Information Processing in Sensor Networks (IPSN 2006) (`http://www.cs.virginia.edu/~ipsn06/WIP/goodwin_1568983444.pdf`)

[Granlund] Granlund G. (2005) A Cognitive Vision Architecture Integrating Neural Networks with Symbolic Processing. Künstliche Intelligenz, Heft 2/05

[Griffiths] Griffiths T. () A reading list on Bayesian methods `http://cocosci.berkeley.edu/tom/bayes.html`

[Grimm] Grimm J. and W. (1857) Von dem Fischer un syner Fru. In: Kinder- und Hausmärchen (Children's and Household Tales – Grimms' Fairy Tales). Berlin

[Gross] Groß R., Tuci E., Dorigo M., Bonani M. & Mondada F. (2006) Object Transport by Modular Robots that Self-assemble. In: Proc. of the 2006 IEEE Int. Conf. on Robotics and Automation, IEEE Computer Society Press, Los Alamitos, CA

[Gruber] Gruber T. R. (1993) A translation approach to portable ontologies. Knowledge Acquisition 5(2), pp. 199-220

[Grübl] Grübl A. (2007) Implementation of a Spiking Neural Network. PhD Dissertation, Heidelberg

[Grush] Grush R. (2004) The emulation theory of representation: Motor control, imagery, and perception. Behavioral and Brain Sciences. 27 (3), pp. 377-442. Postprint available free at `http://repositories.cdlib.org/postprints/723`

[Guerra] Guerra-Filho G., Fermuller C. & Yiannis Aloimonos Y. (2005) Discovering a language for human activity. AAAI Workshop on Anticipation in Cognitive Systems

[Guttag] Guttag J. V., Horowitz E. & Musser D. R. (1978) Abstract data types and software validation. Communications of the ACM, Vol. 21, 12; pp. 1048-1063

[Harnad] Harnad S. (1995) Grounding symbols in sensorimotor categories with neural networks. In Proceedings of "Grounding Representations: Integration of Sensory Information in Natural Language Processing, Artificial Intelligence and Neural Networks", 103

[Harvey] Harvey I. (2000) Robotics: Philosophy of Mind using a Screwdriver. In: Evolutionary Robotics: From Intelligent Robots to Artificial Life, Vol. III, T. Gomi (ed). AAI Books, Ontario, Canada, pp. 207-230 (`ftp://ftp.cogs.susx.ac.uk/pub/users/inmanh/screwdriver.ps.gz`)

[Haynie] Haynie D. (2001) Biological Thermodynamics. Cambridge University Press

[Heath] Heath J. R. & Ratner M. A. (2003) Molecular Electronics. Physics Today, May 2003

[Herrigel] Herrigel E. (1953) Zen in the Art of Archery. Pantheon Books, New York.

[Hildebrandt] Hildebrandt S., Tromba A. (1995) The Parsimonious Universe - Shape and Form in the Natural World. Springer Verlag, New York (Copernicus imprint)

[Hitzler] Hitzler P., Hölldobler S. & Seda A. K. (2004) Logic Programs and Connectionist Networks. Journal of Applied Logic 2(3), pp. 245-272

[Hofmann] Hofmann T. (1999) Probabilistic latent semantic analysis. In Proc. of Uncertainty in Artificial Intelligence, UAI?99, Stockholm, 1999

[Hofstadter] Hofstadter D. (2007) I am a Strange Loop. Basic Books, New York

[Holland] Holland J. H. (1992) Adaptation in Natural and Artificial Systems: 2nd edition. Boston: MIT Press

[Hölldobler] Hölldobler S. (1992) On Deductive Planning and the Frame Problem. In: Logic Programming and Automated Reasoning, International Conference LPAR'92, St. Petersburg, Russia, July 15-20, 1992, Proceedings, Andrei Voronkov (Ed.), Lecture Notes in Computer Science 624, Berlin: Springer

[Hollnagel] Hollnagel E. (Ed.) (2003) Handbook of cognitive task design. Mahwah, NJ: Erlbaum

[Hsu] Hsu D., Figueroa M. & Diorio C. (2000) A silicon primitive for competitive learning. Proceedings NIPS, pp. 713-719

[Huang] Huang Q., Yokoi K., Kajita S., Kaneko K., Arai H., Koyachi N., Tanie K. (2001) Planning walking patterns for a biped robot. IEEE Trans. on Robotics and Automation, 17(3), pp. 280, 289

[Ieropoulos1] Ieropoulos I., Melhuish C. and Greenman J. (2004) Energetically Autonomous Robots. Proceedings of the 8th Intelligent Autonomous Systems Conference (IAS-8), Amsterdam, pp. 128-35.)

[Ieropoulos2] Ieropoulos I., Greenman J., Melhuish C. and Hart J. (2005) Comparison of three different types of microbial fuel cell. J. Enzyme and Microbial Technology, 37(2):238-245

[Illich] Illich I. (1973) Tools for Conviviality. New York: Harper & Row, Publishers. (`http://todd.cleverchimp.com/tools_for_conviviality/`)

[Jaeger] Jaeger H. (2001) The "echo state" approach to analysing and training recurrent neural networks. GMD Report 148, GMD - German National Research Institute for Computer Science

[Jaynes] Jaynes J. (1976, 2000) The Origin of Consciousness in the Breakdown of Bicameral Mind. Boston: Houghton Mifflin.

[Jirsa] Jirsa V. K. & Kelso J. A. S. (2000) Spatiotemporal pattern formation in neural systems with heterogeneous connection topologies. Phys.Rev. E, 62, pp. 8462-8465

[Joslyn] Joslyn C. (1993) Book Review: 'Life Itself'. Int. J. of General Systems 21, pp. 394-402

[Jordan] Jordan P. W., Makatchev M., Pappuswamy U., VanLehn K. & Albacete P. L. (2006) A Natural Language Tutorial Dialogue System for Physics. Proceedings FLAIRS Conference 2006: 521-526

[Kaelbling] Kaelbling L. P. & Littman M. L. (1996) Reinforcement Learning: A Survey. Journal of Artificial Intelligence Research 4, 237-285

[Kasderidis] Kasderidis S. & Taylor J. G. (2004) Attention-based Learning. Proceedings, 2004 IEEE International Joint Conference on Neural Networks

[Kelso] Kelso J. A. S. (1995) Dynamic Patterns: The Self Organization of Brain and Behavior. Cambridge: MIT Press

[Kephart] Kephart J. & Chess D. (2003) The Vision of Autonomic Computing. IEEE Computer Magazine

[Kitano] Kitano H., Asada M., Kuniyoshi Y., Noda I. & Osawa E. (1997) RoboCup: The Robot World Cup Initiative. Proceedings of the First International Conference on Autonomous Agents (Agents'97)

[Koch] Koch C. (2004) The Quest for Consciousness - a Neurobiological Approach. Robert&Company Publishers

[Köhler] Köhler W. (1921) Intelligenzprüfungen an Menschenaffen. Berlin (rev. ed. of Intelligenzprüfungen an Anthropoiden of 1917) (English: The Mentality of Apes. New York: Harcourt and Brace, 1925)

[Kohonen] Kohonen T. (1995) Self-Organizing Maps. Series in Information Sciences, Vol. 30. Springer, Heidelberg. 3rd ed. 2001.

[Kokinov] Kokinov B. (1994). The DUAL cognitive architecture: A hybrid multi-agent approach. In: A. Cohn (Ed.), Proceedings of the Eleventh European Conference on Artificial Intelligence. London: John Wiley & Sons, Ltd.

[Lakoff] Lakoff G. & Núñez R. E. (2000) Where Mathematics Comes From: How the Embodied Mind Brings Mathematics into Being, New York: Basic Books

[Landau] Landau L.J. (1997) Penrose's Philosophical Error. In: Concepts for Neural Networks, L.J.Landau and J.G.Taylor eds., Springer

[Landauer] Landauer T. K., Foltz P. W. & Laham, D. (1998) Introduction to Latent Semantic Analysis. Discourse Processes, 25, pp. 259-284

[LaSpina] La Spina G., Hesselberg T., Williams J. and Vincent J.F.V. (2005) A biomimetic approach to robot locomotion in unstructured and slippery environments. Journal of Bionics Engineering, 2.1, pp. 1-14

[Lehn] Lehn J.-M. (2002) Toward complex matter: Supramolecular chemistry and self-organization

[Lem] Lem S. (1971) Non Serviam. In: The Perfect Vacuum. Orlando: Harcourt Brace Jovanovich (English translation 1978, 1979, 1983)

[Lenat] Lenat D. B., Guha R. V. (1990) Building Large Knowledge Based Systems. Reading, Massachusetts: Addison Wesley

[Louie] Louie A. H. (2007) A Living System Must Have Noncomputable Models. Artificial Life 13.3, pp. 293-297

[Lu] Lu W. , Lieber C.M. (2007) Nanoelectronics from the Bottom-up. Nature Materials 6, pp. 841-850

[Lungarella] Lungarella M., Metta M., Pfeifer R. & Sandini G. (2004) Developmental robotics: a survey. Connection Science, 0(0), pp. 1-40

[Maass] Maass W. (2007) Liquid computing. In: Proceedings of CiE'07, COMPUTABILITY IN EUROPE 2007. Lecture Notes in Computer Science, Springer, Berlin

[Maedche] Maedche A. & Staab S. (2004) Ontology learning. In: S. Staab and R. Studer, editors, Handbook on Ontologies, pp. 173-189. Berlin: Springer

[Matsumaru] Matsumaru N., Centler F., Speroni di Fenizio P., Dittrich P. (2007) Chemical Organization Theory as a Theoretical Base for Chemical Computing. International Journal of Unconventional Computing, 3(4), pp. 285-309

[Maturana] Maturana H. R., & Varela, F. J. (1980) Autopoiesis and cognition: The realisation of the living. London: Reidel

[Maxwell] Maxwell J. C. (1868) On Governors. Proceedings of the Royal Society, no. 100 (1868); or, slightly easier of access, in The Scientific Papers of James Clerk Maxwell, vol. II, pp. 105-120

[McCulloch] McCulloch W. & Pitts W. (1943) A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, 7: pp.115 - 133.

[McMullin] McMullin B. (2004) 30 Years of Computational Autopoiesis: A Review. Artificial Life, Vol. 10, Issue 3, MIT Press

[Mead] Mead C. & Conway L. (1980) Introduction to VLSI Systems. Addison-Wesley, Reading

[Mead2] Mead C. (1989) Analog VLSI and Neural Systems. Addison-Wesley, Reading

[Menciassi] Menciassi A. and Dario P. (2003) Bio-inspired solutions for locomotion in the gastrointestinal tract: background and perspectives. Philosophical Transactions of the Royal Society - Series A, vol. 361, pp. 2287-2298

[MillerJ] Miller J. F. (2003) Evolving Developmental Programs for Adaptation, Morphogenesis, and Self-Repair. In: Advances in Artificial Life - 7th European Conference on Artificial Life (Banzhaf W. et al., ed.), Lecture Notes in Artificial Intelligence, Berlin: Springer

[MillerS] Miller S. L. (1953) Production of Amino Acids Under Possible Primitive Earth Conditions. Science 117: 528.

[Milward] Milward D. & Beveridge M. (2003) Ontology-based dialogue systems. Proc. 3rd Workshop on Knowledge and Reasoning in Practical Dialogue Systems (IJCAI03), pp. 9-18

[Minsky] Minsky M. L. (1986) The Society of Mind. New York: Simon & Schuster

[Minsky2] Minsky M. L. (1994) Will Robots Inherit the Earth? Scientific American 271(4), pp. 108-113

[Mitchell] Mitchell T. M. (2006) The Discipline of Machine Learning. Technical Report CMU-ML-06-108

[MitchellB] Mitchell B. (1965) Theory of Categories. Academic Press, New York and London

[Mitola] Mitola III J. (2000) Cognitive Radio: An Integrated Agent Architecture for Software Defined Radio. Dissertation, Royal Institute of Technology (KTH), ISSN 1403-5286, ISRN KTH/IT/AVH-00/01–SE

[Modayil] Modayil J. & Kuipers B. (2004) Towards Bootstrap Learning for Object Discovery. AAAI-2004 Workshop on Anchoring Symbols to Sensor Data

[Moran] Moran D. (2000) Introduction to Phenomenology. Oxford: Routledge

[Morgenstern] Morgenstern L. (1999) Nonmonotonic Logics. In MIT Encyclopedia of Cognitive Science

[Newell1] Newell A., Shaw J.C. & Simon H.A. (1959) Report on a general problem-solving program. Proceedings of the International Conference on Information Processing. pp. 256-264.

[Newell2] Newell A., Rosenbloom P. S., & Laird J. E. (1989) Symbolic architectures for cognition. In: M. I. Posner (Ed.), Foundations of Cognitive Science. Cambridge, MA: Bradford Books/MITPress.

[Nguyen] Nguyen H. T. & Walker E. A. (1999) A First Course in Fuzzy Logic. Second Edition. Boca Raton, Florida: CRC Press

[Nilsson] Nilsson N. J. (2005) Human-Level Artificial Intelligence? Be Serious! AI Magazine, 25th Anniversary Issue, American Association for Artificial Intelligence

[Nolfi] Nolfi S. & Floreano D. (2000) Evolutionary robotics: the biology, intelligence, and technology of self-organizing machines. Cambridge, Mass.: MIT Press,

[Nomura] Nomura T. (2002) Formal Description of Autopoiesis for Analytic Models of Life and Social Systems. In: Proc. 8th International Conference on Artificial Life (ALIFE VIII), 15-18

[Norman] Norman D. O. & Kuras M. L. (2004) Engineering Complex Systems. Technical Report, The MITRE Corporation (`http://www.mitre.org/work/tech_papers/tech_papers_04/norman_engineering/`)

[Passino] Passino K. M. (1995) Intelligent control for autonomous systems. IEEE Spectrum, June 1995

[Paul] Paul C. (2004) Morphology and Computation. Proceedings of Simulation of Adaptive Behaviour, pp. 33-38

[Paun] Paun G. (2005) Bio-inspired computing paradigms (natural computing). Unconventional Programming Paradigms, Springer LNCS 3566, Berlin, pp. 155-160

[Paun2] Paun G. (1998) Computing with membranes. Turku Center for Computer Science, TUCS Technical report 208

[Pearl] Pearl J. & Russell S. (2003) Bayesian Networks, in: M. A. Arbib (Ed.), Handbook of Brain Theory and Neural Networks, pp. 157-160, Cambridge, MA: MIT Press

[Penrose] Penrose R. (1994) Shadows of the Mind: A Search for the Missing Science of Consciousness. Oxford University Press

[Pfeifer1] Pfeifer R. & Fumiya Iida F. (2005) Morphological computation: connecting body, brain, and environment. Japanese Scientific Monthly, Vol. 58, No. 2, pp. 48-54

[Pfeifer2] Pfeifer R. & Scheier C. (1999) Understanding Intelligence. Boston: MIT PressSeries, Bradford Book Series

[Piaget] Piaget J. (1937) La construction du réel chez l'enfant. Delachaux & Niestlé

[Pierce] Pierce J. R. (1980) An Introduction to Information Theory: Symbols, Signals and Noise. 2nd edition, New York: Dover Publications

[Pinker] Pinker S. (1997) How the Mind Works. WW Norton & Company, London & New York

[Portillo] Portillo I. A. & Atkins E. M. (2002) Adaptive trajectory planning for flight management systems. Proceedings of the AIAA Aerospace Sciences Conference, Reno, Nevada, January 2002 (AIAA 2002-1073)

[Prusinkiewicz] Prusinkiewicz P., Lindenmayer A. (1990) The Algorithmic Beauty of Plants. Springer, Berlin-Heidelberg-New York (available online at `http://algorithmicbotany.org/papers/`)

[Pylyshyn] Pylyshyn Z. W., ed. (1987) The Robot's Dilemma: The Frame Problem in Artificial Intelligence. Norwood, NJ: Ablex Publishing Corporation

[Rechenberg] Rechenberg I. (1973) Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution. Stuttgart: Frommann-Holzboog.

[Reid] Reid, D. (1996) Enactivism as a methodology. In L. Puig & A Gutiérrez (Eds.), Proceedings of the Twentieth Annual Conference of the International Group for the Psychology of Mathematics Education, (Vol. 4, pp. 203-210). Valencia, Spain.

[Riecken] Riecken D. (1994) Intelligent Agents. Communications of the ACM 37.7, pp. 18-21

[Rizzolatti] Giacomo Rizzolatti G. & Craighero L. (2004) The Mirror Neuron System. Annu. Rev. Neurosci. 27, pp. 169-192

[Rocha1] Rocha L. M. (2000) Syntactic autonomy, cellular automata, and RNA editing: or why self-organization needs symbols to evolve and how it might evolve them. In: Closure: Emergent Organizations and Their Dynamics. Chandler J.L.R. and G, Van de Vijver (Eds.) Annals of the New York Academy of Sciences. Vol. 901, pp. 207-223

[Rocha2] Rocha L. M. & Hoedijk W. (2005) Material Representations: From the Genetic Code to the Evolution of Cellular Automata. Artificial Life 11: 189-214

[Rosen1] Rosen R. (1991) Life Itself. A Comprehensive Inquiry into the Nature, Origin, and Fabrication of Life. New York: Columbia University Press.

[Rosen2] Axiomathes, 16 (2006) Special issue dedicated to the work of Robert Rosen. Berlin: Springer

[Roth] Roth G. & Menzel R. (1996) Neuronale Grundlagen kognitiver Leistungen. In: J. Dudel, R. Menzel & R. F. Schmidt (eds). Neurowissenschaft - Vom Molekül zur Kognition, Berlin: Springer, 539-558

[Roy] Roy D. (2005) Grounding words in perception and action: computational insights. Trends in Cognitive Sciences Vol. 9, No.8

[Rutishauser] Ueli Rutishauser U., Joller J. & Douglas R. (2005) Control and Learning of Ambience by an Intelligent Building. IEEE Tansactions on Systems, Man, and Cybernetics - Part A: Ssystems and Humans, Vol. 35, No. 1, 121-132

[Sandini] Sandini G., Metta G. & Vernon D. (2004) RobotCub: An Open Framework for Research in Embodied Cognition. In: IEEE-RAS/RJS International Conference on Humanoid Robotics. 2004. Santa Monica, CA: IEEE (to appear in: International Journal of Humanoid Robotics)

[Schank] Schank R. (1973) Identification of Conceptualizations Underlying Natural Language, in R. Schank and K. Colby (eds.), Computer Models of Thought and Language, chapter 5, W.H. Freeman and Company, pp. 187-248

[Schrauwen] Schrauwen B., Verstraeten D. and Van Campenhout, J. (2007) An overview of reservoir computing: theory, applications and implementations. Proceedings of the 15th European Symposium on Artificial Neural Networks. pp. 471-482

[Schrödinger] Schrödinger E. (1944) What is Life. New York: Macmillan (also: Cambridge University Press, 1967-2003)

[Searle] Searle J. R. (1980) Minds, brains and programs. Behavioral and Brain Sciences 3: 417-424 (`http://www.bbsonline.org/documents/a/00/00/04/84/bbs00000484-00/bbs.searle2.html`)

[Searle2] Searle, J. R. (1969) Speech Acts - An Essay in the Philosophy of Language. Cambridge University Press

[Searle3] Searle, J. R. (1990) Is the Brain a Digital Computer? Proceedings and Addresses of the American Philosophical Association, Vol. 64, No. 3, pp. 21-37

[Sebe] Sebe N., Yafei S., Bakker E., Lew M. S., Cohen I. & Huang T. .S. (2004) Towards authentic emotion recognition. IEEE International Conference on Systems, Man and Cybernetics 2004, Volume 1, pp. 623 - 628

[Sekanina] Sekanina L. (2004) Evolvable computing by means of evolvable components. Natural Computing 3: 323-355, Kluwer Academic Publishers

[Shamsfard] Shamsfard M. & Barforoush A. A. (2004) Learning ontologies from natural language texts. International Journal of Human-Computer Studies. Volume 60 , Issue 1, pp. 17 - 63

[Shanahan] Shanahan M. & Baars B. (2005) Applying global workspace theory to the frame problem. Cognition, pp. 1-20

[Shaw] Shaw J. (2002) Hop, Skip, and Soar - Studying animal locomotion at Harvard's Concord Field Station. Harvard Magazine, Jan-Feb 2002, pp. 27-28

[Shortliffe] Shortliffe E. H., Rhame F. S., Axline S. G., Cohen S. N., Buchanan B. G., Davis R. W., Scott A. C., Chavez-Pardo R. & va. Melle W. (1975) MYCIN: A computer program providing antimicrobial therapy recommendations. Clinical Medicine,Vol 34 (`http://smi-web.stanford.edu/auslese/smi-web/reports/SMI-75-0005.pdf`)

[Shurville] Shurville S. (1993) Symbol Grounding Problems and Sensor Design. In: Proc. of IASTED/IEEE International Conference on Robotics and Manufacturing, September 23-25, 1993, Christ Church, Oxford

[Siegelmann] Siegelmann H. T. (1999) Neural Networks and Analog Computation: Beyond the Turing Limit. Boston: Birkhäuser

[Siegelmann1] Siegelmann H.T. and Sontag E.D. (1994) Analog Computation via Neural Networks. Theoretical Computer Science, vol 131, pp. 331-360

[Siegelmann2] Siegelmann H. T. (1999) Neural Networks and Analog Computation: Beyond the Turing Limit. Birkh¨auser, Boston 1999

[Sifalakis] Sifalakis M. & Hutchison D. (2004) From Active Networks to Cognitive Networks. In: Proceedings of ICRC Dagstuhl Seminar 04411 on Service Management and Self-Organization in IP-based Networks. Schloss Dagstuhl, Wadern, Germany, October 2004.

[Simon] Simon H. A. (1969) The Science of the Artificial. (The Karl Taylor Compton lectures) Cambridge, MA: The MIT Press.

[Sims] Sims K. (1994) Evolving Virtual Creatures. Computer Graphics 7/ 1994, pp. 15-22 (Siggraph '94 Proceedings)

[Singh] Singh P. (2003) Examining the Society of Mind. Computing and Informatics, 22(5), pp. 521-543.

[Sipper] Sipper M., Sanchez E., Mange D., Tomassini M., Pérez-Uribe A. & Stauffer A. (1997) A Phylogenetic, Ontogenetic, and Epigenetic View of Bio-Inspired Hardware Systems. IEEE Transactions on Evolutionary Computation, Vol. 1, No. 1, pp. 83-97

[Sloman1] Sloman A. (1994) Semantics in an intelligent control system. Philosophical Transactions of the Royal Society: Physical Sciences and Engineering 349, pp. 43-58

[Sloman2] Sloman A. & Chappell J. (2005) The Altricial-Precocial Spectrum for Robots. Proceedings IJCAI05, pp. 1187-1192

[Someya] Someya T., Sekitani T., Iba S., Kato Y., Kawaguchi H. & Sakurai T. (2004) A large-area, flexible pressure sensor matrix with organic field-effect transistors for artificial skin applications. PNAS Vol. 101, No. 27 (`www.pnas.orgcgidoi10.1073pnas.0401918101`)

[Sontag] Sontag E. D. (2004) Some new directions in control theory inspired by systems biology. Systems Biology, Vol. 1, No. 1

[Staples] Staples M., Daniel K., Cima M. J. & Langer R. (2006) Application of Micro- and Nano-Electromechanical Devices to Drug Delivery. Pharmaceutical Research

[Starck] Starck J.M. & Ricklefs R.E. (1998) Patterns of Development:The Altricial-Precocial Spectrum. In: Avian Growth and Development. Evolution within the altricial precocial spectrum. J. M. Starck and R. E. Ricklefs (eds). Oxford University Press, New York

[Steels] Steels L. (2005) The Emergence and Evolution of Linguistic Structure: From Lexical to Grammatical Communication Systems. Connection Science. Vol 17(3).

[Stork] Stork H.-G. (2004) Cognition and (Artificial) Cognitive Systems - explanatory & exploratory notes. ftp://ftp.cordis.europa.eu/pub/ist/docs/dir_e/cognition/cognotes1_en.pdf

[Stork2] Stork H.-G. (2005) Report on the Workshop "Future Trends in Artificial Cognitive Systems". Frankfurt Airport Conference Centre, 9th and 10th of December, 2004. ftp://ftp.cordis.europa.eu/pub/ist/docs/dir_e/cognition/acs_workshop_report_hgs2b_en.pdf

[Swinson] Swinson M. L. & Bruemmer J. (2000) The expanding frontier of humanoid robotics. In: IEEE Intelligent Systems Special Issue on Humanoid Robotics, July/August 2000

[Tanenbaum1] Tanenbaum A. S. (2002) Computer Networks (4th edition). Prentice Hall

[Tanenbaum2] Tanenbaum A. S. (2001) Modern Operating Systems. Prentice Hall

[Taylor] Taylor J. G. (2001) Attention as a Neural Control System. Proc. International Joint Conference Neural Nets. 2001, IEEE Press

[TaylorM] Taylor M. E., Matuszek C., Klimt B. & Witbrock M. (2007) Autonomous Classification of Knowledge into an Ontology. In: Proc. 20th International FLAIRS Conference

[ThompsonA] Thompson A. (1998) Hardware Evolution: Automatic design of electronic circuits in reconfigurable hardware by artificial evolution. Berlin: Springer-Verlag

[ThompsonE] Thompson E. (2004) Life and mind: From autopoiesis to neurophenomenology - A tribute to Francisco Varela. Phenomenology and the Cognitive Sciences 3, pp. 381-398

[Trehub] Trehub A. (1991) The Cognitive Brain. MIT Press, Cambridge

[Turing1] Turing A. M. (1936) On Computable Numbers, with an application to the Entscheidungsproblem. Proc. Lond. Math. Soc. (2) 42 pp. 230-265; correction ibid. 43, pp. 544-546 (1937) http://www.emula3.com/docs/OnComputableNumbers.pdf

[Turing2] Turing A. M. (1950) Computing Machinery and Intelligence. Mind 49, pp. 433-460

[Tyrrell] Tyrrell A. M., Sanchez E., Floreano D., Tempesti G., Mange D., Moreno J.-M., Rosenberg J. & Villa A. E. P. (2004) POEtic Tissue: An Integrated Architecture for Bio-Inspired Hardware. In: Lecture Notes in Computer Science Volume 2606, Springer Berlin -Heidelberg

[Ummat] Ummat A., Dubey A., Sharma G. & Mavroidis C. (2006) Bio-Nano-Robotics: State of the Art and Future Challenges. Invited Chapter in: Tissue Engineering and Artificial Organs (The Biomedical Engineering Handbook, M. L. Yarmush (edt.)), CRC Press, ISBN: 0849321239

[Valiant] Valiant L. (1984) A theory of the learnable. Communications of the ACM, Vol. 27, No. 11, pp. 1134-1142

[Varela] Varela F., Thompson E. & Rosch E. (1993) The Embodied Mind - Cognitive Science and Human Experience, MIT Press

[Vert] Vert J.P., Tsuda K. & Schölkopf B. (2004) A Primer on Kernel Methods. In: Kernel Methods in Computational Biology. MIT Press , pp. 35-70 ([http://eprints.pascal-network.org/archive/00000487/])

[von Neumann] von Neumann J. (1966) Theory of Self-Reproducing Automata. University of Illinois Press

[Wagner] Wagner S., Lacour S. P., Jones J., Hsu P. I., Sturm J. C., Li T. & Suo Z. (2004) Electronic skin: architecture and components. Physica E 25, 326-334

[Watzlawick] Watzlawick P., Jackson D. & Beavin J. (1967) Pragmatics of Human Communication: a study of interactional patterns, pathologies and paradoxes. New York: W.W.Norton

[Weizenbaum1] Weizenbaum J. (1966) ELIZA - A Computer Program For the Study of Natural Language Communication Between Man And Machine. Communications of the ACM, 9(1)

[Weizenbaum2] Weizenbaum J. (1976) Computer Power and Human Reason. San Francisco, WH Freeman and Company

[Wells] Wells HG (1937) World Brain: The Idea of a Permanent World Encyclopaedia. In: Contribution to the new Encyclopédie Française

[Weng] Weng J. (2003) Developmental robots: theory and experiments. International Journal of Humanoid Robotics

[Whitaker] Whitaker R. (1998) Encyclopaedia Autopoietica: Autopoiesis & Enaction Compendium (`http://www.cybsoc.org/EA.html#enactive%20cognitive%20science`)

[White] White P., Zykov V., Bongard J. & Lipson H. (2005) Three Dimensional Stochastic Reconfiguration of Modular Robots. Proceedings of Robotics: Science and Systems, Cambridge, MA: MIT Press, pp. 161-168

[Wiedermann] Wiedermann J. (2005) Autopoietic Automata. Technical report No. 92, Institute of Computer Science Academy of Sciences of the Czech Republic

[Wiener] Wiener N. (1948) Cybernetics or Control and Communication in the Animal and the Machine. New York: Wiley & Sons

[Wilks] Wilks Y. (2006) Artificial Companions as a new kind of interface to the future Internet. Oxford Internet Institute, Research Report 13

[Winograd1] Winograd T. (1972), Understanding Natural Language. Academic Press

[Winograd2] Winograd T. (1983) Language as a Cognitive Process: Volume I: Syntax. Reading MA: Addison-Wesley, 1983.

[Winograd3] Winograd T. (1987) A Language/Action Perspective on the Design of Cooperative Work. Human-Computer Interaction 3:1, pp. 3-30

[WinogradFlores] Winograd T. & Flores F. (1986) Understanding Computers and Cognition: A New Foundation for Design. Norwood NJ: Ablex Publishing Corporation

[Wolkenhauer1] Wolkenhauer, O., Hofmeyr, J.-H.S. (2007) An abstract cell model that describes the self-organization of cell function in living systems. J. Theore. Biol., doi:10.1016/j.jtbi.2007.01.005

[Wolkenhauer2] Wolkenhauer O. (2007) Interpreting Rosen. Artificial Life 13.3, pp. 291-292

[Wolpert] Wolpert D. M. & Ghahramani Z. (2005) Bayes rule in perception, action and cognition. In: Gregory, R.L. (ed) The Oxford Companion to the Mind. Oxford University Press (`http://eprints.pascal-network.org/archive/00001354/`)

[Woods] Woods D. D. & Hollnagel E. (2006) Joint cognitive systems: Patterns in cognitive systems engineering. Boca Raton, FL: Taylor & Francis

[Wynne] Wynne C. D. L. (2001) Animal Cognition - The Mental Lives of Animals. Basingstoke: Palgrave

[Zauner] Zauner, K. P. (2005) Molecular Information Technology. Critical Reviews in Solid State and Material Sciences 30(1) pp. 33-69

[Zheng] Zheng W. & Jacobs H. O. (2005) Fabrication of Multicomponent Microsystems by Directed Three-Dimensional Self-Assembly. Advanced Functional Materials 15, p. 732

[Ziemke] Ziemke T. (2001) The Construction of 'Reality' in the Robot: Constructivist Perspectives on Situated Artificial Intelligence and Adaptive Robotics. In: Foundations of Science, special issue on "The Impact of Radical Constructivism on Science", edited by A. Riegler, vol. 6, no. 1-3: 163-233

[Ziemke2] Ziemke T. (2003) "What's that Thing Called Embodiment?" In: Proceedings of the 25th Annual Meeting of the Cognitive Science Society, pp. 1305-1310, Lawrence Erlbaum

[Zykov] Zykov V., Mytilinaios E., Adams B. & Lipson H. (2005) Self-reproducing machines. Nature Vol. 435 No. 7038, pp. 163-164