# Understanding Brains and Minds: A Truly Grand Challenge for Information Science

Aaron Sloman
School of Computer Science, The University of Birmingham,
Birmingham. B15 2TT, UK
http://www.cs.bham.ac.uk/~axs/

February 13, 2009

**Abstract**

This position-paper argues that we still have much to learn from the information processing systems produced by biological evolution since many of their capabilities are *vastly* superior to competing artificial systems. The first information-processing systems were produced (directly or indirectly) by biological evolution, but not for the purposes for which most computers are now used. Biological information-processing systems are control systems for machines that are embedded in a complex and changing 3-D environment, build themselves, control themselves in different ways and on different time scales, in some cases grow their own information-processing architectures as a result of interacting with the environment, in some cases represent, reason about, and interact with other control systems, and are usually engaged in different tasks of different sorts at the same time. Many researchers assume that it is obvious *what* biological systems do (the requirements) and the only problem is to find out *how* they do it (develop the designs). On the contrary much of what they do and how it relates to the features of the environment is not yet understood. So, for those systems, requirements analysis remains a major challenge, as much as work on designs. I shall try to show that the requirements for human-like systems suggest a need for types of multi-level virtual machines forming networks of dynamical systems of different sorts, which grow themselves over time. An implication is that current mechanisms, architectures and designs are grossly oversimple, compared with what is required. Some steps forward are suggested. This work relates closely to several of UKCRC's grand challenges, especially Grand Challenge 5: Architecture of brain and mind.

**Keywords:**
Architectures, Brains, Control, Development, Diversity of designs, Dynamical systems, Environment, Evolution, Minds, Virtual machines.

## 1 Introduction: Visions of Computing

There was a time when "computing" referred to a mathematical activity done by humans, and "computers" were the highly trained human experts doing it. Those computers were products of biological evolution, as well as cultural evolution and individual learning and development. In contrast, nowadays, for most people, the words "computer", "computing", "computation", and phrases containing them, imply a reference to the artifacts we now call computers, which are nothing like biological organisms. They form a class of machines that change as technology develops, but share some common features, both regarding how they work, and what they are used for. To a first approximation, these artificial computers are suited to tasks that can be achieved by manipulating bit-patterns, or more abstract discrete structures implemented in bit-patterns (e.g. trees, graphs, and arrays), though some of the tasks also require use of transducers connected to other things, including motors, sensors, other computers, or networks containing any combination of these devices.

There is an increasingly widely used notion of "information-processing system" that subsumes both the old and the new concepts of computer, and also includes many products of biological evolution that are in various ways far more sophisticated than anything produced by human designers, although some artificial systems surpass natural ones in speed and reliability (e.g. arithmetical calculators). The superiority of much biological information-processing should not surprise us. Mechanical engineering, which has a much longer history than computer science or software engineering, also lags behind the achievements of biological evolution in several respects, including power-weight ratio, miniaturisation, self-construction, self-repair, and in many cases exquisite matches between designs and requirements.

Some researchers have begun to acknowledge the importance of biological information-processing and there are many computing projects attempting to use biologically inspired mechanisms. This theme is part of several of the UKCRC Grand Challenges: Grand Challenge 1: In Vivo – In Silico,[1] Grand Challenge 7: Journeys in Nonclassical Computation,[2] and Grand Challenge 5: Architecture of Brain and Mind.[3]

Despite the growing recognition of the importance of biological information-processing systems, there are kinds of sophistication in the processing supported by brains that have not generally been noticed, and which will prove extremely difficult to implement using existing hardware and software techniques. Whether the task is merely difficult, or impossible without replacing the "bottom level" information-processing devices with something quite unlike current computers (e.g. molecular computers, which play many important roles in biological organisms, including construction, repair and control), is an open question.

## 2 Conceptions of computation: general and specific

What the common features of computation are assumed to be differs from one community to another, and has evolved in many small steps. I remember encountering ridicule from an astronomer around 1980 in response to the suggestion that computers could be used for word-processing. It is very likely that the protester now uses computers for word-processing every day.

Some relevant ideas were developed long before what we now call computers existed: they can be found in many devices designed to control behaviours of physical machines in a systematic way, including music boxes, player pianos, mechanical clocks, toys with more or less complex behaviours, programmable looms, governors controlling the speed of machines by using feedback loops and card-sorting machines. In all these cases the information-processing

[1] http://fizz.cmp.uea.ac.uk/Research/ivis/index.jsp

[2] See http://www.cs.york.ac.uk/nature/gc7/ and the two issues of *International Journal of Unconventional Computing* edited by Susan Stepney, Vol 3 numbers 2 and 3, 2007

[3] http://www.cs.bham.ac.uk/research/projects/cogaff/gc

was very closely tailored to the requirements for controlling specific sorts of machine in specific environments. Such *application specificity* characterises most, perhaps all, of the information-processing systems produced by evolution. Certain human brain functions may be the only exceptions.

In contrast, the "core" study of computing now mostly focuses on *generality*, including general-purpose machines, general-purpose programming formalisms, general models of computation and general-purpose tools. An example of this generality, sometimes called "universality" is found in a universal Turing machine, which is universal in the sense that any other Turing machine can be modelled on it, as can several other very general forms of computation. The view of this sort of machine as universal may be illusory, if there are forms of information-processing that are not capable of being accommodated within that framework. It is unlikely that Turing universality subsumes all the forms of biological information-processing described below. (Ref: report on GC7-BCS08)

Because evolution generally selects useful design changes for a particular type of organism (specified by a genome) in a particular type of context (a niche) the results are often far from general purpose. Control-systems for different sorts of machine have different sets of requirements that are satisfied by different designs. AI researchers often rediscover a trade-off between generality and power. The most effective systems for performing a particular type of task are often those that are designed to use forms of representation and architectures suited to that task. In contrast more general designs often spend more time searching through large search-spaces.

## 3 A broader vision of computing

This paper is an attempt to promote a broader vision, one in which the current narrow (but general!) notion of computing is a special case of something that has been far more widespread on this planet for millions of years. I am referring to the many and varied forms of information-processing produced by biological evolution long before humans started to make computers, themselves an indirect product of biological evolution, like everything else created by humans and other animals. (Compare Dawkins on the "extended phenotype" (Dawkins 1982).)

Like most of the deep concepts of the physical and biological sciences, the concept "information" used in this context cannot be defined explicitly without circularity, as explained in Section 13. In that sense all organisms do information-processing: they acquire information, about their own state or about aspects of the environment, and use it, either immediately, e.g. to control external behaviours or to change some enduring information structure, or store it for future use (possibly after transforming it or combining it with other information). In animals very many kinds of information are acquired, and many different uses are made of that information, using different forms of representation, different processing mechanisms and different architectures combining many mechanisms of different sorts. However, there are some researchers who, while drawing attention to the importance and interest of biological information-processing systems, restrict their attention to a narrow subset (e.g. popular choices are neural nets that omit most of the features of biological neural nets, or simplified models of evolution that omit the complexities of epigenesis and the complex interactions between diverse species evolving in an ecosystem). So many who promote biologically inspired computing ignore important forms of biological information-processing.

Some features of computers, for instance high speed arithmetical operations, surpass anything found in nature. Yet evolution not only produced information-processing machines long before we did, but also produced examples which, in their domain (for example, self-assembly, self-repair, and functions like human vision[4]), far outstrip anything we can now design. I try to characterise some aspects of that gap below. Whether we can reduce it, or even eliminate it, is central to a broad study of computing, construed as information-processing, which, may make future scientists smile at the limited preoccupations of computer scientists at the beginning of the 21st Century.

## 4 Virtual Machines

Although computer scientists have important lessons to learn from biological investigations, the converse is also true. The use of computation not merely to process biological data but also to run models that articulate theories about biological processes is rapidly increasing. However there is one feature of computer science whose importance has not been widely understood – perhaps not even by most computer scientists, though it is of profound importance both for those who design, build and maintain complex computing systems and those who wish to understand natural information processing. I am referring to the fact that we use physical computers to run *virtual machines*, whose properties and behaviours are very different from the properties of the physical machines in which they are implemented. This aspect of computation has largely been ignored by many who discuss the nature and significance of information-processing, including philosophers, psychologists and neuroscientists trying to understand the implications of computing ideas for their own fields.

Virtual machines are important for several reasons. One is that by formulating designs at a virtual machine level, engineers allow the possibility of making use of new technology to implement improved (faster, cheaper, more compact) physical systems running pre-existing virtual machines. Perhaps evolution also benefited from separation of improvements in design and improvements in implementation, which we could call *vertical modularity*. A more important, but related, reason for making use of virtual machines is that they are typically much simpler and easier to design, extend, understand, test, and debug than the physical processes that occur when they run. When a word processor, or email program starts up on a computer, that causes many changes in the memory of the computer and alters the sequence of machine instructions that the CPU executes, though previously started programs will typically continue running. Although instructions for different processes are interleaved, a process does not cease to exist when another runs, since most of the implementation of each process is in enduring memory. If process A accesses some of the memory of process $B$ it is influenced by $B$ even though $B$ is not "running". So enduring virtual machines have enduring causal powers, even though their instructions run only intermittently. In multi-processor implementations, e.g. brains, influences can go in both directions simultaneously. The use of virtual machines provides a kind of vertical modularity in the system that is very different from the separation into spatially distinct physical modules (*horizontal modularity*), but nevertheless supports "separation of concerns" for designers, maintainers, and so on.

When several virtual machines run concurrently, the set of physical processes that can occur at the implementation level, is enormously complex and enormously varied, since what

[4] Illustrated in this online demonstration http://www.cs.bham.ac.uk/research/projects/cogaff/misc/multipic-challenge.pdf

exactly goes on in all the central registers, in the buses transferring bit patterns around the system, and in the various hardware interfaces, depends on which *collection* of virtual machines happens to be running on the physical machine and also on the operation of memory management systems. By abstracting away from most of that detail we can understand and reason about the virtual machines whose invariant features and causal relationships allow huge variations in physical event sequences that implement them.

Not only is this useful for the humans who design and build information processing systems: it can also be important for sub-systems *within* the machine, that monitor and control other systems – for example a scheduler that ensures fair allocation of resources to different processes, or a file-system manager that monitors and constrains reading and writing of files, for instance preventing some processes from accessing files they are not entitled to access. If those functions had to be performed by controlling processing at the physical level instead of using virtual machine interfaces, they would be far more complex, and would have to be re-designed whenever the hardware technology changed, and also as the set of programs running on the computer changed.

Those advantages of virtual machines, and possibly others, appear to have been "discovered" and exploited by biological evolution long before human engineers thought of using them, and long before humans existed on earth, though details are very different in biological systems. One example, apparently unique to humans, is the ability of brains using very different human languages (e.g. English and Chinese) to learn and use the same mathematical, physical, and geographical facts. The physical neural processes involved in using different languages, will be very different even when thought contents are the same.

## 5 Philosophical issues concerning virtual machines

Events in virtual machines can be *causes*, not only of other virtual machine events, but also of physical processes, such as events on a screen or in a chemical plant. Some philosophers think that talk of virtual machines is not to be taken seriously: it is just a useful fiction that can be used to summarise what is "really" being done by physical machines. But that view fails to do justice to the processes whereby virtual machines are designed, tested, debugged, extended and used in many areas of engineering, science, commerce, entertainment, and domestic usage. I shall not pursue this point now except to note that similar disputes are possible regarding the reality and causal powers of other states, processes and events that we normally regard as real and having effects, for example: economic inflation, poverty, spread of rumours, changing fashions, voter apathy, education, religion, and so on. These are all aspects of biological virtual machines, implemented in the brains, social institutions and physical environments of large numbers of humans. We would not be able to learn and reason about social and economic systems if we thought only about the underlying *physical* causal processes, though many people feel that *only* physical events and processes can be causes: they have an incorrect understanding of what causation is. The correct understanding will not be discussed here, except to note that it allows the same event to have different causes at different levels of abstraction.[5] Claiming that what current physics models and explains is the only kind of cause, will turn out to be problematic if physics is extended to include

[5]See http://www.cs.bham.ac.uk/research/projects/cogaff/talks/wonac for presentations (with J. Chappell) discussing differences between Kantian and Humean causation, in humans, animals and robots. There is also a relevant discussion in (Sloman 2008a).

deeper levels of processing, as has happened in the past. For more on virtual machines in biology, engineering and philosophy, see http://www.cs.bham.ac.uk/research/projectsx/cogaff/talks/#wpe08

## 6 Biological Information Processing Systems

The first information-processing systems, and by far the most complex ones on earth, were produced (directly or indirectly) by biological evolution, but not for the purposes for which most computers are used by humans. Biological information-processing systems are *control systems* for machines that (a) are embedded in a complex and changing 3-D environment, (b) build themselves, (c) control themselves in different ways and on different time scales, (d) in some cases grow their own information-processing architectures as a result of interacting with the environment, (e) in some cases represent, reason about, and interact with other control systems, (f) are usually engaged in different tasks of different sorts at the same time.

Understanding a control system requires understanding its environment. Researchers often assume they know what biological systems do (the requirements) and believe the only problem is to find out how they do it (develop the designs). In reality, much of what organisms do and how their behaviour depends on features of the environment is not yet understood. So the "requirements analysis" task remains to be done. We'll see that requirements for human-like systems suggest a need for types of multi-level virtual machines forming networks of dynamical systems of diverse sorts, which grow themselves over time. So current artificial mechanisms, architectures and designs are grossly inadequate. Some steps forward are suggested. This work relates closely to the aims of UKCRC's Grand Challenge 5: Architecture of brain and mind.

## 7 Kinds of Dynamical System

I shall discuss a collection of competences of humans and other animals that appear to develop over time under the control of both the environment and successive layers of learning capability that build on previously learned capabilities. For example after an infant has learnt to control some of her movements, she can use that competence to perform experiments on both the physical environment and nearby older humans. After new forms of representation have been learnt they can be used to learn how to form and test new plans, goals, hypotheses, etc., by acting in the environment. After new concepts have been acquired they can be used to formulate new theories.

All of the above can help drive the development of linguistic and other communicative competences, but languages for *internal* use must develop first (Sloman 1979; Sloman and Chappell 2007): contradicting some common assumptions about human language being required for thinking about thinking. On the contrary, pre-verbal communication is essential for leaning a human language.

Being able to communicate with others makes it possible to learn things that others have already learnt. These layered learning processes start in infancy, and, in humans, can continue throughout adult life, extending the information-processing architecture.

Such competences are usually studied separately, in different disciplines. Likewise people who build working AI models or robots consider only a small subset of the competences shown by humans and other animals, and different researchers focus on different subsets. It is not obvious that models or explanations that work for a narrowly focused set of tasks can be extended to form part of a more general system: systems that "scale up" do not always "scale out".

Analysis of combinations of different sorts of competence, including perceiving, reasoning, planning, controlling actions, developing new ontologies, playing, exploring, inventing explanations, and interacting socially, provides very demanding requirements to be met by both: (a) human-like robots that develop through interacting with a rich and complex 3-D world and (b) an explanatory theory of how humans (and similar animals) do what they do. Tasks (a) and (b) both need to take account of the distinctive features of 3-D environments in which objects of very varied structure, made of many different kinds of materials with different properties, can interact, including objects manipulated by humans, animals or robots. The evolutionary niches associated with such environments posed combinations of problems for our biological predecessors that need to be understood if we wish to understand or replicate the products of evolution.

## 8 Interacting designs and niches

Consideration of a space of niches (possible sets of requirements) and a space of designs for different sorts of animal or machine reveals many nature/nurture tradeoffs, and exposes problems that AI researchers, psychologists and neuroscientists have not addressed. A robot or animal that learns through play and exploration in a complex, changing, 3-D environment needs certain competences that also support human abilities to do mathematics (including geometry and topology), for instance the ability to perceive and reason about both action affordances and epistemic affordances, as explained in (Sloman 2008c). Viewing mathematical competence as side-effects of evolutionary processes meeting biological needs can shed new light both on old philosophical problems about the nature of mathematical knowledge and on problems in developmental psychology and education, especially mathematical education, as explained in (Sloman 2008b) and `http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#math-robot`.

We need to understand much better the kind of self-extending, multi-functional, virtual machine information-processing architecture required to explain such human capabilities; a requirement that no current AI systems or current neural theories come close to addressing.

## 9 Dynamical systems, some simple and some very complex

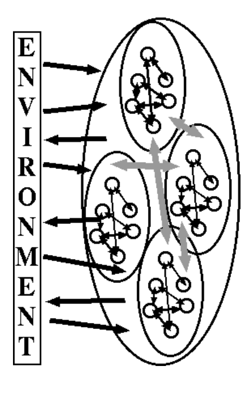



Figure 1: *Dynamical systems as often conceived of: closely coupled with an environment through sensors and effectors.*

Figure 1 illustrates, in a sketchy fashion, the relatively simple dynamical systems investigated in robots and other "biologically inspired computational systems in the last two decades, e.g. influenced by Brooks and others (Brooks 1990; Brooks 1991), who emphasise close coupling with the physical environment. In contrast, consideration of many human competences, especially sophisticated human visual competences,[6] suggests a kind of machine sketched in Figure 2, consisting of a much more complex dynamical system, itself composed of a network of dynamical sub-systems of different sorts, which grows itself over time. We hypothesise that such complex multi-layered self-extending dynamical systems exist in humans and some other animals as *virtual* machines that run on their brains, and which will also be needed for future human-like robots, and other intelligent control systems.

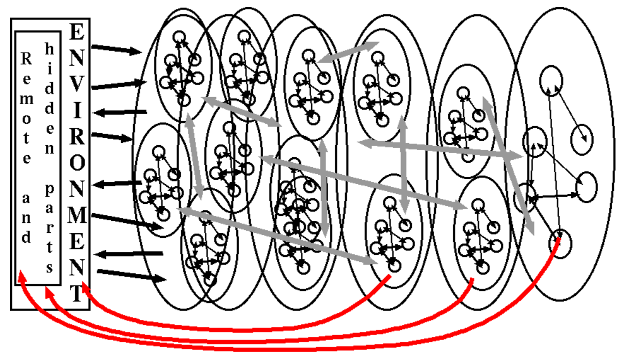



Figure 2: *A more complex kind of dynamical system, composed of sub-systems of different sorts, concerned with different levels of abstraction, some continuous and some discrete, with sub-systems grown at different stages of development. Long arrows indicate references to inaccessible (e.g. past, remote, future, microscopic) entities.*

For more detail see (Sloman 2003).

Those who emphasise *embodied* cognition, often assume, mistakenly, that only dynamical systems closely coupled with the environment are required, failing to notice that something very different is needed to support more abstract, more loosely coupled, more discrete, processes, such as enable you to think about past events, wonder about possible futures, form hypotheses about what is going on out of sight, plan journeys to remote locations. do algebra in your head, discuss philosophy, read music, or read these notes. To what extent these occur in other species is not so clear. Elephants may not play chess, study transfinite ordinals, or wonder how to design intelligent systems, but simple sensori-motor based architectures may not suffice for the things they can do, such as helping calves overcome physical obstacles.

Just as their opponents ignore some requirements, people who think symbolic AI will suffice for everything fail to attend to the kinds of intelligence required for controlling continuous actions in a 3-D structured environment, including maintaining balance, drawing a picture with a pencil, and playing a violin. Many such activities require both sorts of mechanism operating concurrently.

The ability to think about mental contents (in oneself or others) requires the use of symbolic *meta-semantic* competences involving the ability to cope with referential opacity – a problem discussed by some AI researchers but with no agreed solution. Some propose notational solutions, e.g. using special logics (e.g. McCarthy), while others propose architectural solutions, in which the same formalisms can be used for different purposes, with different constraints in different parts of the architecture (see (Sloman 2008c)). Moreover, different dynamical systems operating concurrently at different levels of abstraction need to be accommodated in any theory of motivation and affect, since motives, attitudes, evaluations, preferences, hopes, fears and the like can exist at many levels.[7]

## 10 Types of Architecture

We can crudely decompose the variety of sub-processes that occur in biological organisms in two dimensions: one concerned with whether the processes are (1) perceptual/sensory, or (2) central or

[6]Illustrated in this online demonstration http://www.cs.bham.ac.uk/research/projects/cogaff/misc/multipic-challenge.pdf

[7]For a critique of some shallow theories of emotion see http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#cafe04 and http://www.cs.bham.ac.uk/research/projects/cogaff/04.html#200403 “What are emotion theories about?”

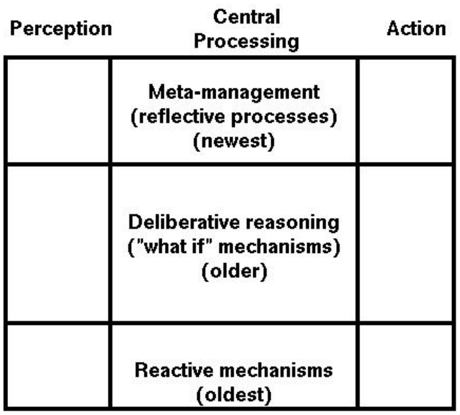



Figure 3: The CogAff schema specifying, to a first approximation, types of (concurrently active) components that can occur in an architecture.

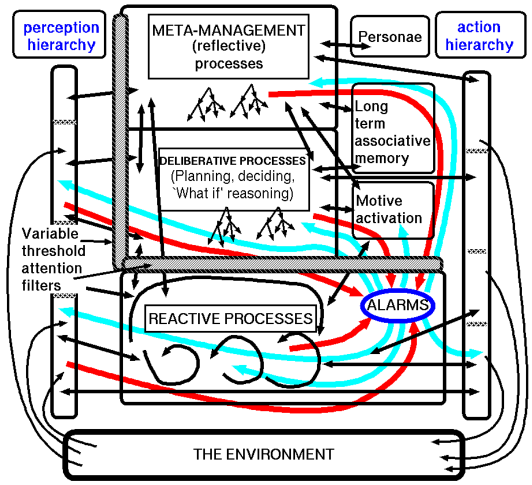



Figure 4: The H-CogAff architecture – an instance of the CogAff schema that seems to be required to account for many human competences described in previous work. For more detail see (Sloman 2003).

(3) concerned with effectors/actions. and another dimension concerned whether the processes are based on (1) reactive (usually evolutionarily old) mechanisms, or (2) deliberative mechanisms or (3) meta-management mechanisms (concerned with self-monitoring or control, or using meta-semantic competences in relation to other agents). Note that 'reactive' here does not imply close coupling with the environment.

Division in those two dimensions produces a 3 by 3 grid of types of sub-system as illustrated in Figure 3. The grid is only an approximation – more subdivisions are to be found in nature than this suggests, in both dimensions, but especially the vertical dimension. It is often forgotten that biological evolution can produce only discrete changes, not continuous changes. The discrete changes vary in size and significance: e.g. duplication is often more dramatic than modification, and can be the start of a major new evolutionary development.

A special case of the CogAff schema that we have been investigating for some years is the H-CogAff (Human-CogAff) architecture specification depicted crudely in Figure 4, (which partly overlaps with Minsky's ideas in (Minsky 2006)). There is a lot more detail regarding these ideas in presentations here: `http://www.cs.bham.ac.uk/research/projects/cogaff/talks/` and papers in the Birmingham CogAff and CoSy projects.[8]

## 11 Co-evolution of designs and niches

There is a space of possible designs for working systems (*design space*) and a space of possible niches within which systems can function, learn, evolve, etc. (*niche space*). There are complex mappings between those spaces, with no simple one to one relationship (Figure 5). It is misleading to think of every relationship between a design and a niche as a numerical fitness value, or a vector of fitness values. Rather (as in consumer reports), there will be *structured* relationships between features of designs and features of niches, e.g. specifying what the consequences are of a particular design limitation in a particular class of niches. Compare the type of analysis of bugs in a program that helps a good

[8] `http://www.cs.bham.ac.uk/research/projects/cogaff/` `http://www.cs.bham.ac.uk/research/projects/cosy/papers/`

programmer think about how to improve the program. A vector of numerical values will not usually characterise what needs fixing: descriptions of flaws are needed.

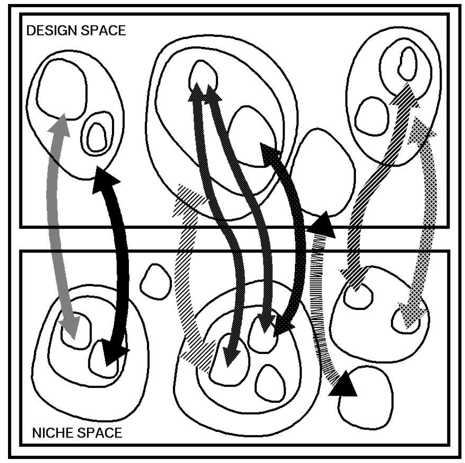



Figure 5: Design space and niche space and structured mappings between them.

Since the niche of a particular type of organism depends in part on the design features of a set of *other* organisms (influencing its physical characteristics and its behaviours, learning abilities, etc.) we can say that the niche inhabited by instances of a particular design is partly produced by other coexisting designs. Such a niche will, in turn, tend to produce evolutionary pressures on the designs selected for it. So designs will change, producing evolutionary trajectories through design space. However, since changes in the designs will lead to changes in the niches they produce there will also be evolutionary trajectories through niche space (e-trajectories in Figure 6) — feedback in ecosystems. The individuals that are instances of evolving designs may also undergo learning and development, producing individual trajectories (i-trajectories in Figure 6).

At every stage the designs and individuals must be biologically viable. Contrast that with the case of a human designer tinkering with a design: during some of the discontinuous changes there may be no working instances. The corresponding trajectories (which may be called repair trajectories (r-trajectories in Figure 6) will be discontinuous, unlike i-trajectories and e-trajectories.

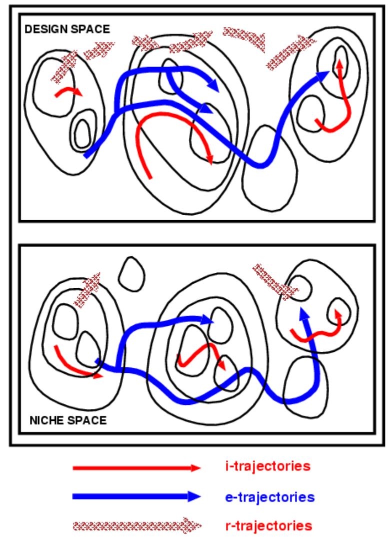
Insofar as individual members of a design or class of designs communicate, and can acquire information that is passed on to off-spring, we can talk about *cultures* being added to the virtual machinery. There will then also be cultural or social trajectories in the spaces (not depicted).

## 12 Multiple virtual machines everywhere

The concurrent evolution of both multiple designs and multiple (often co-located) niches forms an enormously complex dynamical system, a virtual machine that explores several spaces of possibilities at different levels of abstraction concurrently on different time-scales. An ecosystem, or even the whole biosphere, can be thought of as a complex *virtual* machine with multiple concurrent threads, producing multiple feedback loops of many kinds (including not just scalar feedback but structured feedback – e.g more like sentences than forces or voltages). This may be the main truth in the Gaia hypothesis! A new kind of mathematics for multi-level inhomogeneous dynamical systems will be needed for a more precise characterisation.

Individual organisms, insofar as they implement some of the designs indicated above, will also include multiple virtual machines. I assume all these virtual machines are ultimately implemented in physics and chemistry, though there are some people who don't like that idea. So far we don't know enough about what variety of virtual machines can be implemented on the basis of physical mechanisms that we currently understand. If it turns out that there is something that cannot be implemented in that sort of physics because something more is required, this is more likely to support

Figure 6: Various trajectories can link the two spaces, with very complex feedback loops.




the conclusion that our physical theories need to be extended rather than the conclusion that physics is not enough.

## 13 What is information?

There is a highly technical notion of "information" introduced by Shannon which has caused enormous confusion because it is only indirectly related to our ordinary notion of information, which involves meaning, reference, semantic content, truth, falsity, relevance, implications, contradictions, and so on. We could call this concept, which I have been using, the *referential* concept of information, in contrast with Shannon's *numerical* concept of information.

The pre-Shannon, referential concept of "information" compares, in depth, generality and relevance to understanding the universe, with concepts of "matter" and "energy". None of these can be defined explicitly without circularity. Rather they are all mostly defined *implicitly* by the explanatory theories in which they occur. Residual ambiguities are reduced by the ways in which we relate those theories to observations, measurements, predictions and explanations of facts about the world. This use of implicit definitions is common for deep theoretical terms in explanatory science.

How do we (and physicists) manage to understand the concept 'energy'? Answer: by making use of it in a rich, deep, widely applicable theory (or collection of theories) in which many things can be said about energy, e.g. that in any bounded portion of the universe there is a scalar (one-dimensional), continuously variable amount of it, that its totality is conserved, that it can be transmitted in various ways, that it can be stored in various forms, that it can be dissipated, that it flows from objects of higher to objects of lower temperatures, that it can produce forces that cause things to move or change their shape, etc. etc.

The crucial point is that the logical structure of the theory determines which portions of reality can be models of the theory, and which aspects of those models could be referred to by the undefined symbols in the theory. The richer and more precise the implications of the theory, the more constrained the models are, and so the more precise the meanings of implicitly defined theoretical terms. Additional constraints come from 'bridging rules' linking aspects of the theory to kinds of prediction and measurement. This amounts to a theory of meaning through "theory tethering" rather than through "symbol grounding".[9]

Likewise, despite the lack of any *explicit* definition of the word 'information', we understand it insofar as it is *implicitly* defined by its use in a rich, deep, and widely applicable theory (or collection

[9]As explained in http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#models

of theories) in which many things can be said about information, e.g. that it is not conserved (I can give you information without losing any), that instead of always having a scalar value, items of information have a structure (e.g. there are replaceable parts of an item of information such that if those parts are replaced the information changes but not necessarily the structure), that it can be transmitted in various ways, that it can vary both discontinuously (e.g. adding an adjective or a parenthetical phrase to a sentence, like this) or continuously (e.g. visually obtained information about a moving physical object), that it can be stored in various forms, that it can influence processes of reasoning and decision making, that it can be extracted from other information, that it can be combined with other information to form new information, that it can be expressed in different syntactic forms, that it can be more or less precise, that it can express a question, an instruction, putative matter of fact, and in the latter case it can be true or false, known by $X$, unknown by Y, while Z is uncertain about it, etc. etc.[10]

Some items of information allow infinitely many distinct items of information to be derived from them. (E.g. Peano's axioms for arithmetic, in combination with predicate logic.) Physically finite, even quite small, objects can therefore have infinite information content, e.g. systems containing brains or computers, as argued in (Sloman 1996).

## 14 Relations between information items

Although there is no generally useful scalar concept of "amount" of information there is a partial ordering of containment. Thus one piece of information $I1$ may contain all the information in $I2$, and not vice versa. In that case we can say that $I1$ contains more information. But not every partial ordering determines a linear ordering, let alone a scalar measure. Even the partial ordering may be relative to an information user: Giving information $I1$ to an individual $A$, may allow $A$ to derive $I2$, whereas another individual $B$ may not be able to derive $I2$, because the derivation depends on additional in formation, besides that in $I1$.

I suspect it is one of the basic laws of the universe that operations in which the information content of some bounded system increases require energy to be used in the physical machine in which the information processing is implemented. This may be a special case of something more general.

There are informationally equivalent (i.e. mutually derivable) rearrangements of information-bearing structures where one arrangement is more useful for certain purposes than others. Learning such things is an important aspect of human development.

## 15 Potential information content for a user, or type of user

Whereas energy and physical structures simply exist, whether used or not, information is only information for a type of information-user. The very same physical structure can contain different information, or refer to something different for another user $U'$, with different potentialities. Thus a structure $S$ refers to $X$ or contains information about $X$ for $U$, a user of $X$. $U$ need not actually do anything with the information, it may be stored for a rainy day that never comes. All that is required is the *potential* to use it.

The information in $S$ can be potentially usable by $U$ even though $U$ has never encountered $S$ or anything with similar information content. That is obviously true when $U$ encounters a new sentence, diagram or picture for the first time. Even before $U$ encountered the new item, it was potentially usable as an information bearer. This can also be true of other environmental structures (e.g. those with epistemic affordances (Sloman 2008c)). In some cases, the potential cannot be realised without $U$ first learning a new language, or notation, or even a new theory within which the information has a place.

[10]For more on this, see http://www.cs.bham.ac.uk/research/projects/cogaff/misc/whats-information.html

You cannot understand the information that is potentially available to others in your environment if you have not yet acquired all the concepts involved in the information. For example, it is likely that a new born human infant does not have the concept of a metal, i.e. that is not part of its ontology. So it is incapable of acquiring the information that it is holding something made of metal even if a doting parent says "you are holding a metal object". The information-processing mechanisms (forms of representation, algorithms, architectures) that are required to think of things in the environment as made of different kinds of stuff, of which metals are a subset, take several years to develop in humans, requiring development of several layers of dynamical system in Figure 2.

It is also possible for an information-bearing structure $S$ to express different information, $X$, $X'$, $X''$, for the same user $U$ in different contexts, e.g. because $S$ includes an explicit indexical element (e.g. 'this', 'here', 'you', 'now', or non-local variables in a computer program). Another factor that makes it possible for $U$ to take a structure $S$ to express different meanings in different contexts can be that $S$ includes a component whose semantic role is to express a higher order function which generates semantic content from the context.

The fact that the information content of $S$ depends on a user $U$ may suggest that information is something subjective. However the deeper fact is that $S$ potentially expresses information $X$ for a *type* of user, even if no user of that type happens to exists. That potentiality is an objective fact about $S$.

# 16 Conclusion

Biological evolution produced systems that vary enormously in complexity. They all acquire and use information, but they use different forms of information representation (e.g. scalar, structured, continuous, discrete, static, dynamic) different kinds of information contents (e.g. physical quantities, representations of spatial structures and spatial processes including both geometric and topological structures and changes, causal and functional relations, meta-semantic information, predictive theories, explanatory theories, and many kinds of historical and geographical information about the individual's world). They acquire information in different ways (including use of very varied physical sensors), process it in different ways, store it over different time scales, and apply it for different purposes, including controlling different interfaces to the environment and, in some cases, constructing or extending their own information-processing capabilities, often triggered by failures of prediction or control.

The differences between the simplest organisms and the most sophisticated ones are enormous (though the latter subsume the former). The intermediate cases do not form a continuum: there are many significant discontinuities both in evolution and in individual development, and in any case biological differences are all ultimately based on molecular differences and therefore discrete.

There is no point trying to introduce a binary division between those organisms that do and those that do not use information. Rather we need to distinguish different kinds of information, different kinds of manipulation of information, different uses of information and different kinds of information-user: exploring the space of designs and the space of sets of requirements (niches), the trade-offs between them and the ways they can change, in evolution, in individuals, in cultures and in ecosystems.

This analysis of the space of designs and possible sets of requirements can cover both natural and artificial systems. Although much work has already been done, it is clear that there are some kinds of information-processing (e.g. human visual information-processing) that evolution produced that we do not understand at present, and that, in their domain are vastly superior to anything we know how to do with computers.

So there is much more work to be done. It also seems likely that instead of being done mainly in computer science departments, more and more of the scientific study of information-processing, including developing theories about types of information-processing and analysing their properties, will be done within other disciplines, though they will need to use, and probably extend some of the key ideas developed in computer science, such as notions of complexity and the idea of a virtual machine.

It it is possible that computer science departments will be left in a backwater, dealing only with the residual subset of problems related to artificial bit-manipulation machines and their applications. But the future doesn't have to be like that.

## References


[Brooks 1990] R. A. Brooks. Elephants Don't Play Chess. *Robotics and Autonomous Systems*, 6:3–15, 1990. http://people.csail.mit.edu/brooks/papers/elephants.pdf.

[Brooks 1991] R. A. Brooks. Intelligence without representation. *Artificial Intelligence*, 47:139–159, 1991.

[Dawkins 1982] R. Dawkins. *The Extended Phenotype: The long reach of the gene*. Oxford University Press, Oxford, New York, 1982.

[Jablonka and Lamb 2005] Eva Jablonka and Marion J. Lamb. *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*. MIT Press, Cambridge MA, 2005.

[Minsky 2006] M. L. Minsky. *The Emotion Machine*. Pantheon, New York, 2006.

[Sloman and Chappell 2007] Aaron Sloman and Jackie Chappell. Computational Cognitive Epigenetics (Commentary on (Jablonka and Lamb 2005)). *Behavioral and Brain Sciences*, 30(4):375–6, 2007. http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0703.

[Sloman 1979] A. Sloman. The primacy of non-communicative language. In M. MacCafferty and K. Gray, editors, *The analysis of Meaning: Informatics 5 Proceedings ASLIB/BCS Conference, Oxford, March 1979*, pages 1–15, London, 1979. Aslib. http://www.cs.bham.ac.uk/research/projects/cogaff/81-95.html#43.

[Sloman 1996] A. Sloman. Beyond turing equivalence. In P.J.R. Millican and A. Clark, editors, *Machines and Thought: The Legacy of Alan Turing (vol I)*, pages 179–219. The Clarendon Press, Oxford, 1996. (Presented at Turing90 Colloquium, Sussex University, April 1990. Also Cognitive Science technical report: CSRP-95-7).

[Sloman 2003] A. Sloman. The Cognition and Affect Project: Architectures, Architecture-Schemas, And The New Science of Mind. Technical report, School of Computer Science, University of Birmingham, 2003. http://www.cs.bham.ac.uk/research/projects/cogaff/03.html#200307 (Revised August 2008).

[Sloman 2008a] A. Sloman. The Well-Designed Young Mathematician. *Artificial Intelligence*, 172(18):2015–2034, 2008. http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0807.

[Sloman 2008b] Aaron Sloman. Kantian Philosophy of Mathematics and Young Robots. In S. Autexier, J. Campbell, J. Rubio, V. Sorge, M. Suzuki, and F. Wiedijk, editors, *Intelligent Computer Mathematics*, LLNCS no 5144, pages 558–573, Berlin/Heidelberg, July 2008. Springer. http://www.cs.bham.ac.uk/research/projects/cosy/papers#tr0802.

[Sloman 2008c] Aaron Sloman. Varieties of Meta-cognition in Natural and Artificial Systems. In M. T. Cox and A. Raja, editors, *Workshop on Metareasoning, AAAI'08 Conference*, pages 12–20, Menlo Park, CA, 2008. AAAI Press. http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0803.