Requirements for a Human-like Information Processing Architecture that Builds Itself by Interacting with a Rich Environment

Aaron Sloman
http://www.cs.bham.ac.uk/~axs
School of Computer Science
University of Birmingham

Last updated: 9 Jun 2008 (For Complexity CRN Birmingham University).
A closely related paper, on visions of computing, is here.

Origins

This was originally an extended abstract for an invited talk at The 8th Understanding Complex Systems Symposium,
at UIUC, Illinois, May 12-15 2008.

As indicated on the Symposium Schedule the talk was presented on Tuesday 13th May.

The web file was expanded for presentation at the Symposium because I had trouble interfacing my laptop with the projector.

Later the material was slightly expanded for a presentation on "Modelling Multilevel Dynamical Systems", at the University of Birmingham on Monday 9th June 2008, at a Workshop on Complexity and Critical Infrastructures - Environment focus, organised by the Informatics Collaborative Research Network (CRN).

```
The location of this file is
   http://www.cs.bham.ac.uk/research/projects/cogaff/misc/ucs2008.html

A pdf version of this file is
   http://www.cs.bham.ac.uk/research/projects/cogaff/misc/ucs2008.pdf
```

To illustrate the theoretical ideas presented below, the talk was intended to start with some videos showing infants and toddlers of different ages, apparently performing different tasks, at several levels of abstraction, concurrently, and Betty the New Caledonian crow making a hook.

For more information about Betty, including pictures and videos see the Oxford University Behavioural Ecology lab web page:
http://users.ox.ac.uk/~kgroup/

(I showed the movie of trial 7, but all the others are worth watching, as they raise varied questions about the representations, information-processing architectures, and mechanisms used).

Some videos of children related to the topic of this talk were assembled for another talk in September 2007
http://www.cs.bham.ac.uk/research/projects/cosy/conferences/mofm-paris-07/sloman/vid

For example, watch the short video of a 15 month old child, maintaining his balance, pushing and steering a broom in front of him, and apparently noticing opportunities to turn, and obstacles to avoid well enough in advance to prepare for them in the way he steers the broom. There is a very short, easily missed pair of episodes near the beginning where he works out (a) how to extract the back of the broom handle from between two vertical rails, and (b) how to cope with the fact that he has pushed the broom into the wall instead of towards the corridor.

Felix Warneken's web site, with videos of chimps and children displaying spontaneous cooperation is here
http://email.eva.mpg.de/~warneken/

Warneken's research is mainly focused on questions about altruism in very young (pre-verbal) children and chimpanzees, whereas my interest is concerned with what representations and information-processing architectures are required for an individual (especially an individual without human language) to perceive that another individual has a goal, to work out what the goal is, to detect that the goal is not being achieved, to decide to help, and then devise a plan to achieve the goal and carry out the plan.

Original Abstract

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 capabilities 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.

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.

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.

Such competences are usually studied separately, in different disciplines.

Likewise people who attempt to 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, seeing explanations, and interacting socially, provides very demanding requirements to be met by:

- human-like robots that develop through interacting with a rich and complex 3-D world
- an explanatory theory of how humans (and similar animals) do what they do.

Both tasks 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.

That manipulation includes assembling and disassembling structures of varying complexity, and varying modes of composition -- including food!

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 the products of evolution.

Consideration of a space of niches and a space of designs for different sorts of animal and different sorts of machine reveals nature/nurture tradeoffs, and indicates hard problems that AI researchers, psychologists and neuroscientists have not addressed, e.g. why a robot or animal that learns through play and exploration in a complex, changing 3-D environment needs competences (e.g. the ability to perceive and reason about both action affordances and epistemic affordances) that also seem to underlie human abilities to do mathematics (including geometry and topology).

Viewing mathematical competence as a side-effect 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.

I shall talk briefly about 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.

The hypothesized required architecture consists of a very complex dynamical system
- composed of a network of dynamical systems of different sorts
- that grows itself,
- which can be contrasted with the very much simpler kinds of dynamical system that have so far been investigated in biologically inspired robotics.

The two diagrams below illustrate, in a sketchy fashion, both the simple systems and the kind of complexity that we require, and which I suggest exists in virtual machines that run on human brains and the brains of some other animals, and will need to be replicated in human-like robots.

TWO KINDS OF DYNAMICAL SYSTEM
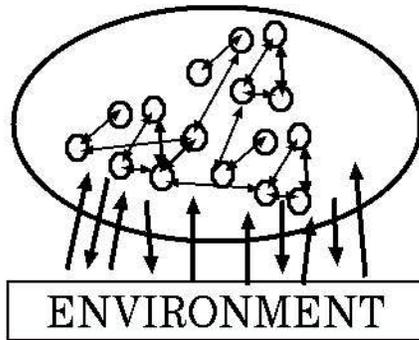
One kind, on the top left, is closely coupled with the environment through sensors and effectors.

The other kind, below, has many sub-systems, with different levels of abstraction and decoupling from the environment, including subsystems concerned with theorising, reasoning, planning, remembering, predicting and explaining, or even free-wheeling daydreaming, wondering, etc.

Some subsystems change continuously, especially those closely coupled with the environment, while others change discretely.

At any time many subsystems are dormant, but able to be activated very rapidly by constraint-propagation mechanisms.

Evolution produced both kinds and many intermediate kinds.

The more complex ones grow themselves after birth, under the influence of interactions with the environment (e.g. a human information-processing architecture).
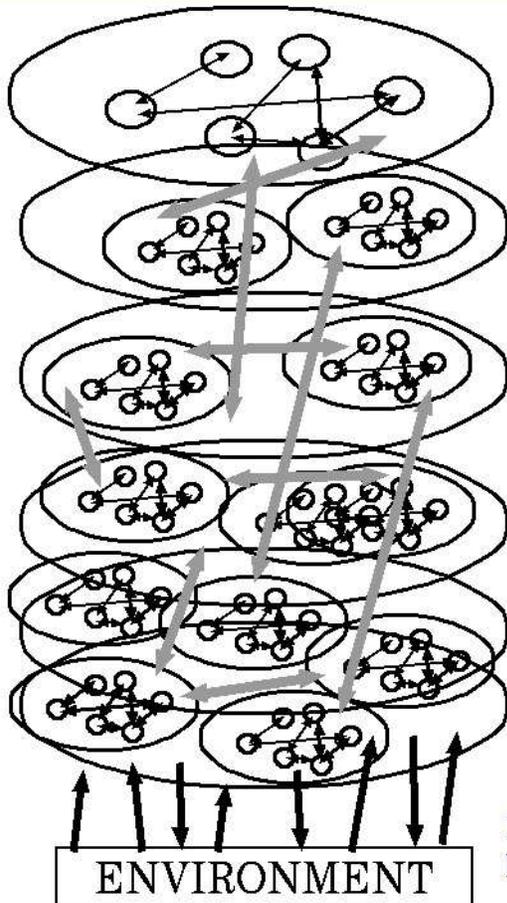
For more on this see these talks
http://www.cs.bham.ac.uk/~axs/talks.html

I assume that most of these dynamical systems are *virtual machines* implemented on physical machines in brains and bodies.

Confused intellectual fashions

Many people who have grown disillusioned with symbolic AI mechanisms fail to realise that something like those mechanisms are needed for the more abstract, more loosely coupled dynamical systems, for example the ones that enable you to do algebra in your head, discuss philosophy, make plans for conference travel, or read these notes.

Likewise, 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 while pushing a broom, drawing a picture with a pencil, and playing a violin. Many such activities require both sorts of mechanism operating concurrently, along with additional monitoring, evaluating, and learning (e.g. debugging) mechanisms.

Sometimes the fact that humans, other animals, and robots are *embodied* can lead thinkers to make the mistake of assuming that only the first type of dynamical system is required because they forget that some embodied systems (including themselves) can think about past, future, distant places, games of chess, transfinite ordinals and how to design intelligent systems.

They also forget that humans are capable of overcoming many physical obstacles produced by genetic deformity, illness or accidents causing serious damage (blindness, deafness, loss of one or more limbs, etc.). There is no unique route from birth to the mind of an adult human. Instead there are many different possible competences that are only partially ordered. The fact that serious physical deformity does not prevent development of normal human vision and other aspects of normal adult human cognition, probably depends on the fact that the individual's brain has many structures that evolved to meet the needs of ancestors with the full range of modes of interaction with the environment.

A paper criticising over-emphasis, or misplaced emphasis, on the importance of embodiment is here:
http://www.cs.bham.ac.uk/research/projects/cogaff/09.html#912 (PDF)
Some Requirements for Human-like Robots: Why the recent emphasis on embodiment has held up progress.

Multiple dynamic systems with different cognitive and affective functions

The different roles of dynamical systems operating concurrently at different levels of abstraction need to be accommodated in any theory of motivation and affect, since motives, evaluations, preferences, and the like can exist at many levels.

For a critique of some shallow theories of emotion see
http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#cafe04
"Do machines, natural or artificial, really need emotions?" (PDF presentation).

and
http://www.cs.bham.ac.uk/research/projects/cogaff/04.html#200403
"What are emotion theories about?" (PDF -- paper for AAAI Spring Symposium 2004).

An alternative type of model

Many researchers think that the way to design an intelligent robot, is to build it on the basis of a cycle of operations

SENSE --> THINK --> ACT --> SENSE --> THINK --> ACT ....

Some would require 'THINK' to be decomposed into something like this:

THINK = (INTERPRET --> PLAN --> DECIDE)

The SENSE and ACT steps could also be decomposed, of course, especially where perception requires multiple layers of interpretation, and action requires multiple layers of control.

There are some who don't believe that, but think that all AI researchers believe it -- possibly with different views as to the complexity of the three phases.

What's wrong with it?
There are two main things wrong.

1. It assumes that all the internal information-processing is made up of discrete steps, ruling out the possibility of continuous sensing, continuous control and continuous forms of reasoning or plan formation.

2. It assumes that the architecture is inherently sequential, doing one thing at a time, whereas biological organisms and robots will need to be able to do more than one thing at a time, possibly very many things at a time, possibly concerned with very different tasks,

   e.g. walking to the door while you talk to someone, or plan what you are going to say to the person in the next room.

Types of Architecture

We can crudely decompose the variety of sub-processes that occur in biological organisms in two dimensions
- Whether perceptual/sensory, or central or concerned with effectors/actions.
- Whether based on evolutionarily old reactive mechanisms, or deliberative mechanisms or meta-management mechanisms (concerned with self-monitoring or control, or using meta-semantic competences in relation to other agents).

That produces a 3x3 grid of types of sub-system as illustrated below.
(Note that 'reactive' here does not imply closely coupled with the environment).

The CogAff Architecture Schema

Perception	Central Processing	Action
	Meta-management (reflective processes) (newest)	
	Deliberative reasoning ("what if" mechanisms) (older)	
	Reactive mechanisms (oldest)	

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).

N.B. 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 development.

The Complexity and Speed of Human Vision

Humans seem to be able when confronted with a new scene or picture, whose contents are not predictable from what happened previously, to produce several levels of interpretation at very high speed.

An informal demonstration is here:
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/multipic-challenge.pdf

The fact that information is gained from perception at different levels of abstraction (not just in a purely bottom-up way, even if the process is *initiated* bottom-up) indicates the need for perceptual systems to be *layered*.

For similar reasons, action systems need to be layered: control operates at different levels of abstraction, e.g. walking while you control direction, speed and perhaps try to adopt a style that suggests nonchalance.

The H-CogAff (Human-CogAff) Architecture Sketch (for an Adult human)
The architecture must grow itself, in humans more slowly than in other animals.
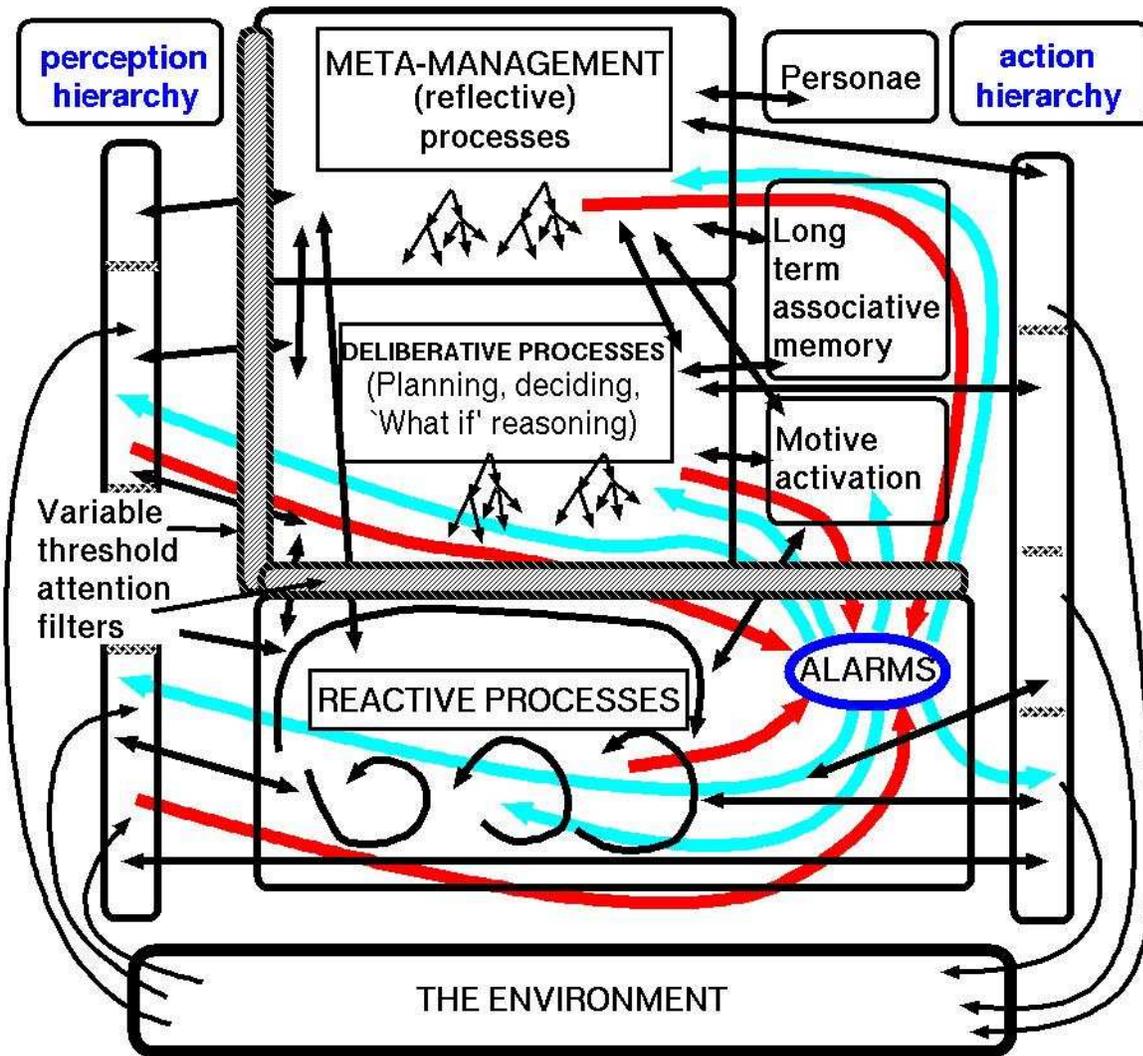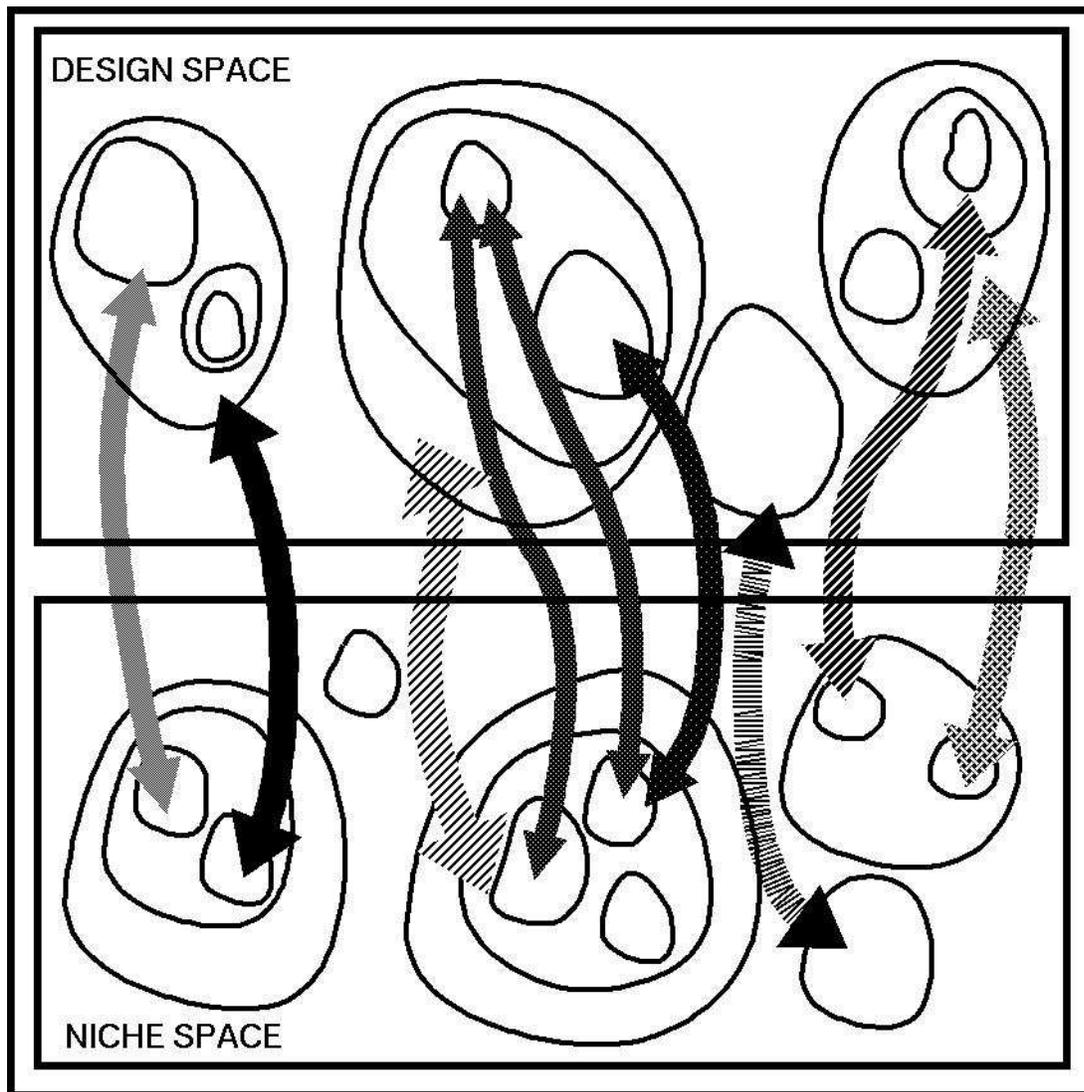
There will not be time to explain all of this, but there is lots more here:

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.

It is not helpful to think of the relationship between a design and a niche as numerical fitness value, nor even a *vector* of fitness values in different dimensions.

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 programmer think about how to improve the program. Merely being given numbers produced by quality measures would be *useless* to a software engineer responsible for maintaining a complex piece of software.

Design-niche interactions

Since the niche of a particular organism depends in part on the design features of a set of *other* organisms (helping to determine their physical characteristics and their behaviours, learning abilities, etc.) we can say that the niche inhabited by instances of a particular design is partly produced by other coexisting designs.

That niche will tend to produce evolutionary pressures on the designs instantiated in it. So designs will change, and there are consequently evolutionary trajectories through design space.
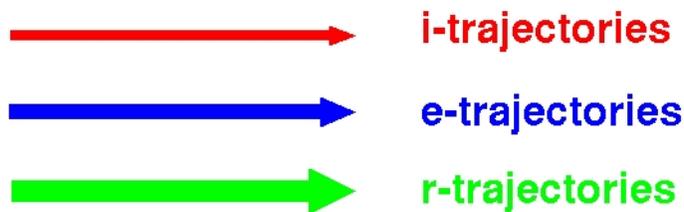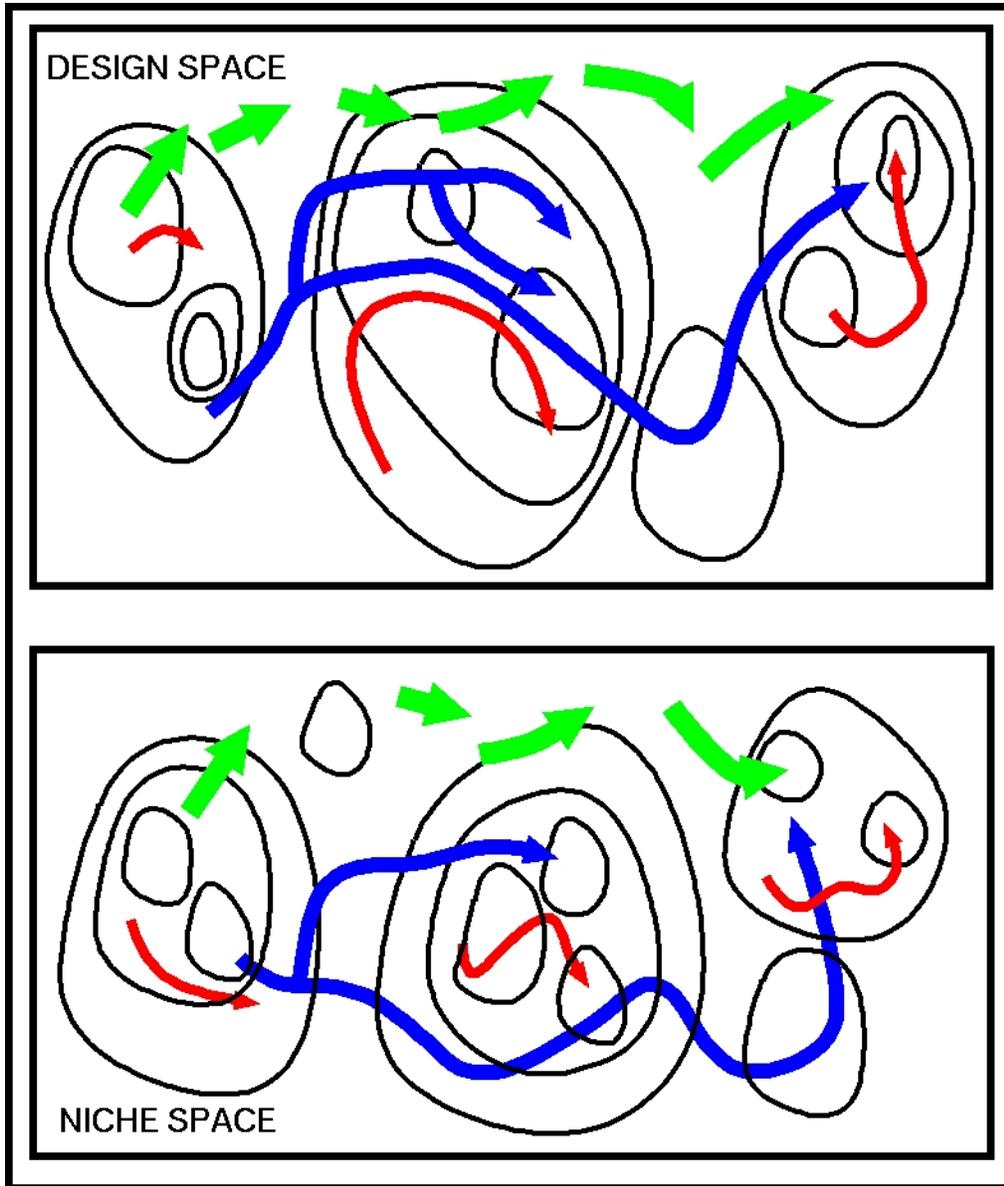
However, since changes in the designs will lead to changes in the niches they produce there will also be evolutionary trajectories through niche space (e-trajectories).

The individuals that are instances of evolving designs will undergo learning and development, producing individual trajectories (i-trajectories).

At every stage the designs and individuals must be biologically viable. Contrast that with the case of a human 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) will be discontinuous, unlike i-trajectories and e-trajectories.

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

The concurrent evolution of both multiple designs and multiple (often co-located) niches forms an enormously complex dynamical system, which can be thought of as exploring several spaces of possibilities concurrently.

High level view of an I-trajectory for a human-like system
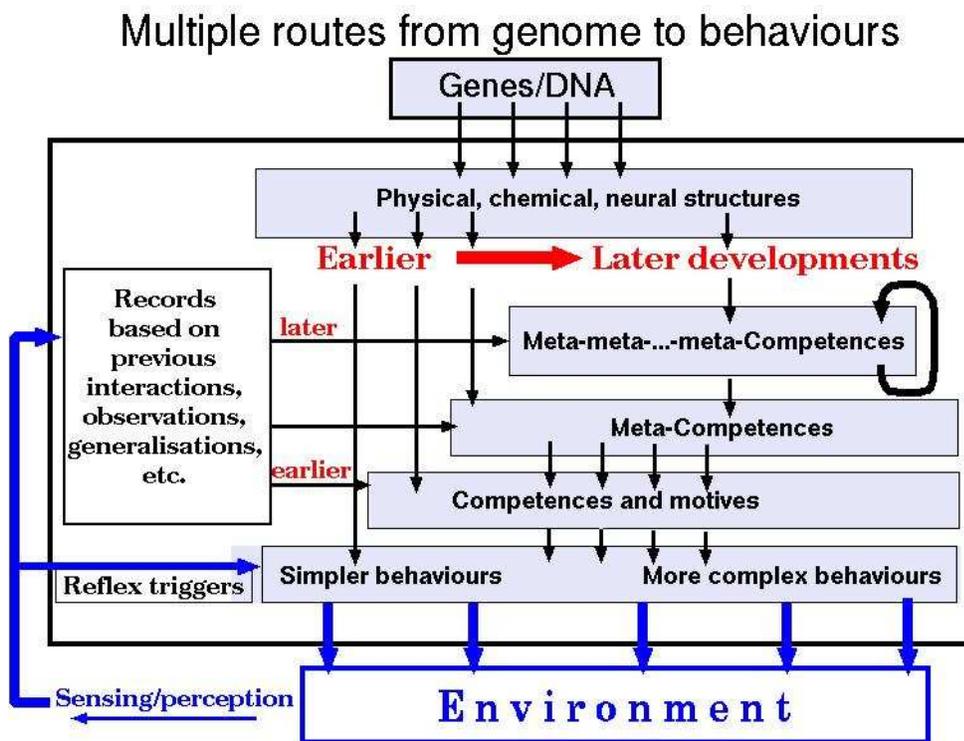
A learner can build its own layers of competence tailored to features of the environment, at increasing layers of abstraction.

Most organisms generate behaviours only via routes on the left of the diagram, i.e. largely controlled by the genome.

In part that's because mechanisms to operate towards the right of the chart are expensive, so that species using them have to be near the peak of a food pyramid.

See

Jackie Chappell, Aaron Sloman, 2007,
Natural and artificial meta-configured altricial information-processing systems,
In *International Journal of Unconventional Computing,* 3, 3, pp. 211--239,
http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0609,

Multiple virtual machines everywhere

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 a sentence than a force or voltage).

The 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 interested on the basis of physical principles as we currently understand them.

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

NOTE

A presentation that enlarges on the connection between biological spatial competences and the ability to become a mathematician is here
http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#math-robot

Some recent papers, closely related

- http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0801
  Architectural and representational requirements for seeing processes and affordances.
- http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0802
  Kantian Philosophy of Mathematics and Young Robots
- http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0803
  Varieties of Meta-cognition in Natural and Artificial Systems