Notes for Workshop on "Thinking Architecturally"

28-9 November 2012, Computer Lab, Cambridge University

Background Notes (HTML):
http://tinyurl.com/CogMisc/archthink-sloman.html

Related Presentations (PDF):
http://tinyurl.com/BhamCog/talks/

Related discussions (mostly html):
http://tinyurl.com/CogMisc

Aaron Sloman
http://www.cs.bham.ac.uk/~axs/
School of Computer Science, University of Birmingham.
(Philosopher in a Computer Science department)

These slides will later be added to my “talks” directory:
http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#archthink

A longer set of slides is being prepared for an extended tutorial
on the Meta-Morphogenesis project, summarised here:
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/agi-2012-tut-sloman.html

Architecture? Internet?

I don't work on networks or the internet.

Digression: I first used the internet about 1974 at Sussex University, connected (via UCL) from a 10cps paper teletype to a MAXC computer at Xerox Palo Alto Research Centre, trying out their LOGO system. That convinced me that the Edinburgh POP-2 language would be much better for teaching AI to beginners than Logo. We developed a teaching-enhanced subset of POP-2, called Pop-11, running on a DEC PDP11/40.

I have been thinking about architectures for minds since about 1972.

NOTE:

Arguing about definitions – e,g, what does "architecture" mean? – is (nearly) always a complete waste of time.

Instead:

Talk about problems, evidence, theories, explanations, predictions, uses, useful formalisms, modes of reasoning/inference, and how they relate to different architectures, their uses, limitations, etc. ....

I have some notes (not definitions) related to architectures, designs and requirements (all viewed as potentially dynamic) in my answers to questions circulated before the workshop, here:
`http://www.cs.bham.ac.uk/research/projects/cogaff/misc/archthink-sloman.html`

Rejoinder to Karl Marx

Engineers try to change the world.

Rejoinder to Karl Marx

Engineers try to change the world.

But that requires understanding it.

Rejoinder to Karl Marx

Engineers try to change the world.

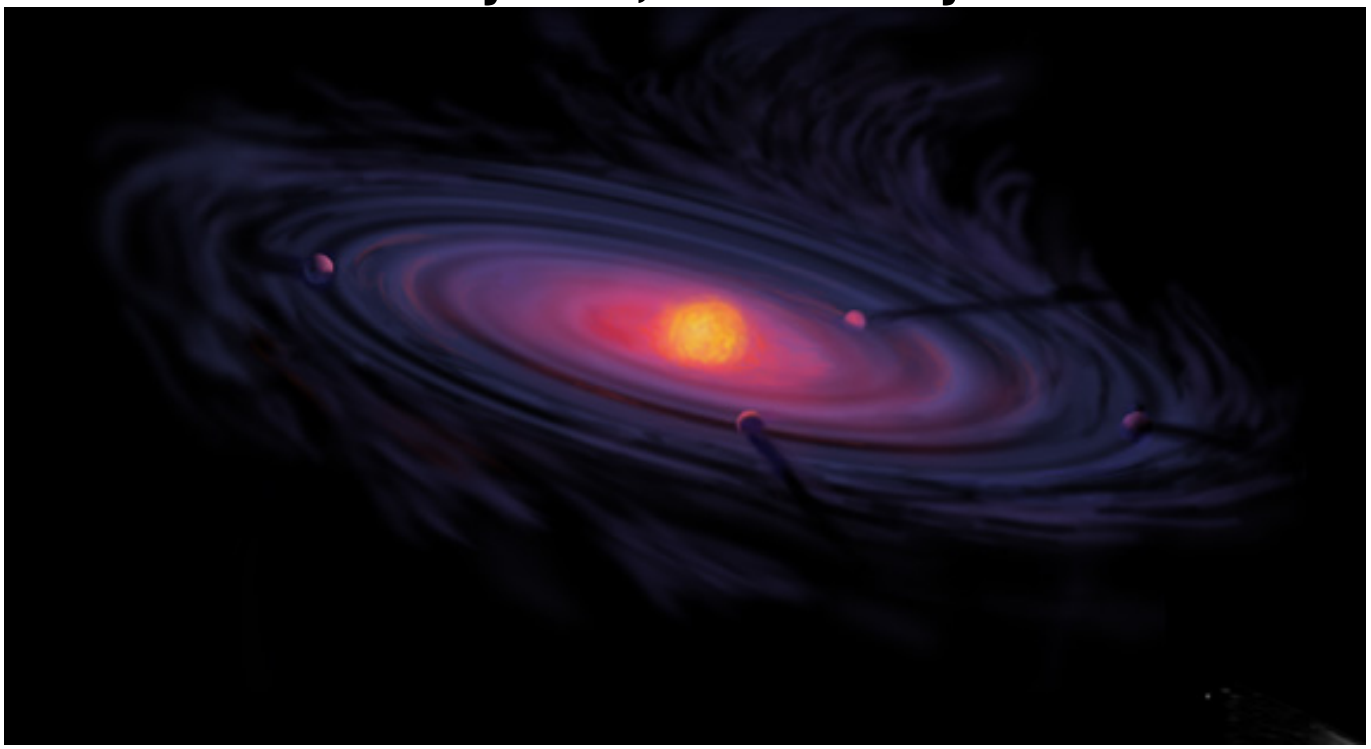
But that requires understanding it.

Which, in turn, requires certain sorts of information processing mechanisms.

As evolution “discovered”, when producing more complex, more competent organisms.

The context of my thinking about architectures

Wayback, when dusty



Artist's concept of a protoplanetary disk (NASA)

How can a cloud of dust containing only physical/chemical structures produce musicians, mathematicians, metaphysicians, megalomaniacs, monkeys, mice, microbes, music, mayhem, murder, munificence, marmite and other wondrous products?

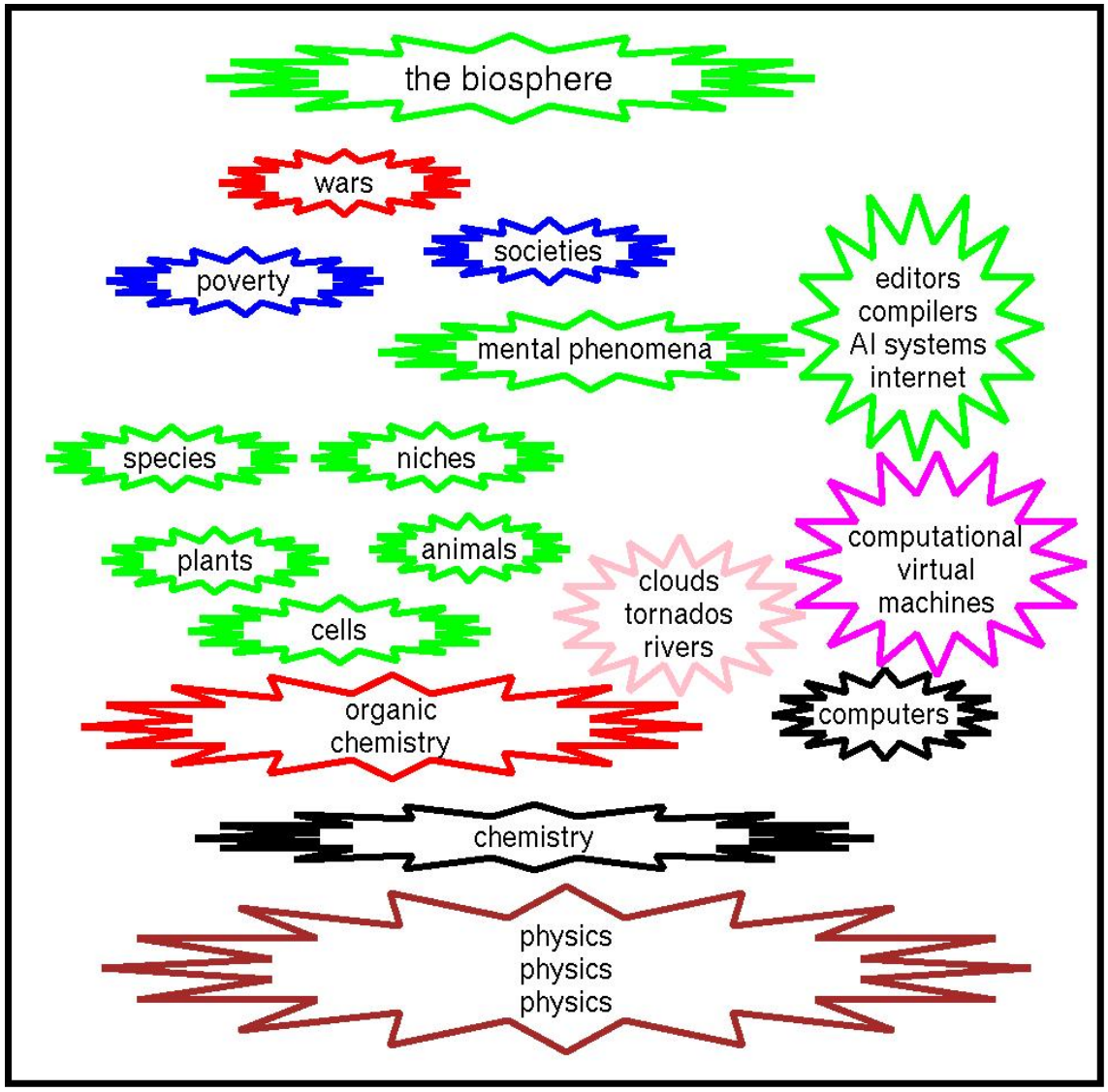
The question driving the Meta-Morphogenesis project

(Partly inspired by Turing's work on morphogenesis)

`http://tinyurl.com/CogMisc/misc/meta-morphogenesis.html`

`http://tinyurl.com/CogMisc/misc/evolution-info-transitions.html`

The biosphere – a layered(?) architecture



Notice the different routes “upwards” from physics, and how they merge.

This is not intended to be a complete specification of the biosphere!

Termite cathedrals and human cathedrals both have architectures

They are each instances of (multiple) architecture types (designs).

Everything instantiates multiple types (designs) at different levels of abstraction.

NOTE: A design does not require a designer

See:

`http://www.cs.bham.ac.uk/research/projects/cogaff/misc/archthink-sloman.html`

Main topics

- How can complex life forms and their mental processes come out of a cloud of dust?
  (Where does the information required come from? Discussed in (Davies, 1999))
- What sorts of transitions in information-processing can occur ?
  New kinds of: uses of information, information contents, information sources, forms of representation, modes of reasoning, types of communication, mechanisms, architectures, sensors, motors, types of motivation, sources of motivation, ways of managing multiple motives, ...
  Especially motives (and mechanisms) related to understanding the world.
  Contrast transitions in:
  – morphology
  – behaviours
  – environments
- What sorts of underlying mechanisms are required to support such processes?
  – In what sorts of worlds could they occur?
  – Could a Turing machine support them?
  – What's special about chemistry?
- There are many products of the process that we still cannot replicate in AI/robots
  – Perception and use of spatial affordances (Broom pushing video, parrot video)
  – Geometric reasoning
  – Interplay of development and learning
  – many, many, more! (Including enjoying a joke, or music.)
- Is this a worthwhile (massive) collaborative, multi-disciplinary project or just foolish hand-waving, or...?

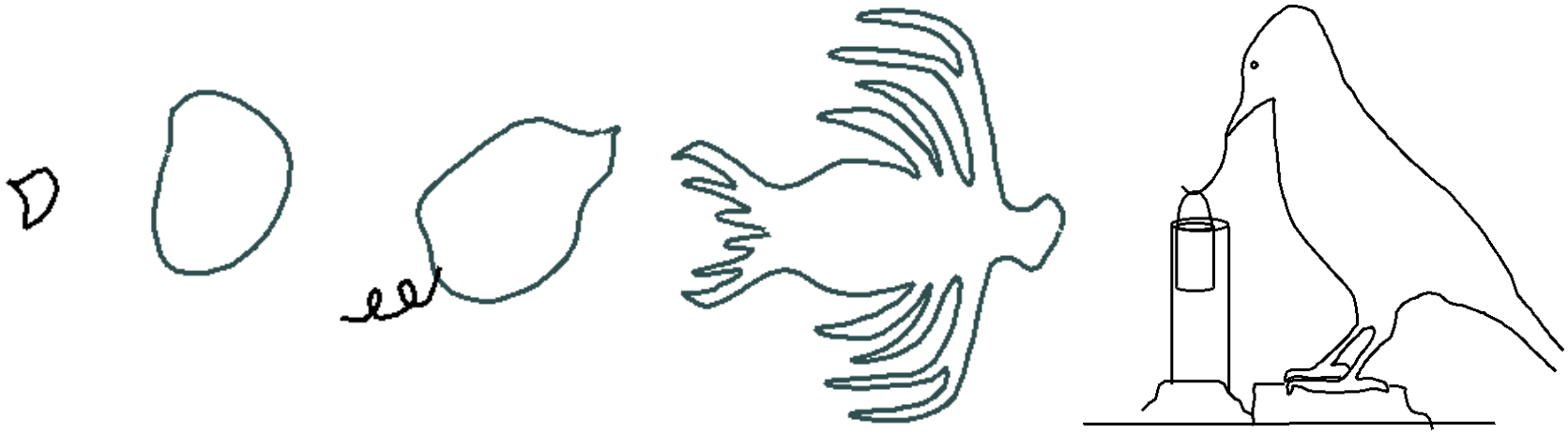
What sorts of transitions can occur?

A beginning here:

`http://tinyurl.com/CogMisc/evolution-info-transitions.html`

"A DRAFT list of types of transitions in biological information-processing
or
Varieties of Evolved Computation"

All organisms, including microbes, are information-processors but the information to be processed, the uses of the information, and the means of processing vary enormously



Things that change include

– Environments/Requirements (niches)
– Morphology (including types of sensors and motors)
– Types of information processed
– Types of processing
– Architectures
– Kinds of motivation
– ... (lots more)

Some of the products:

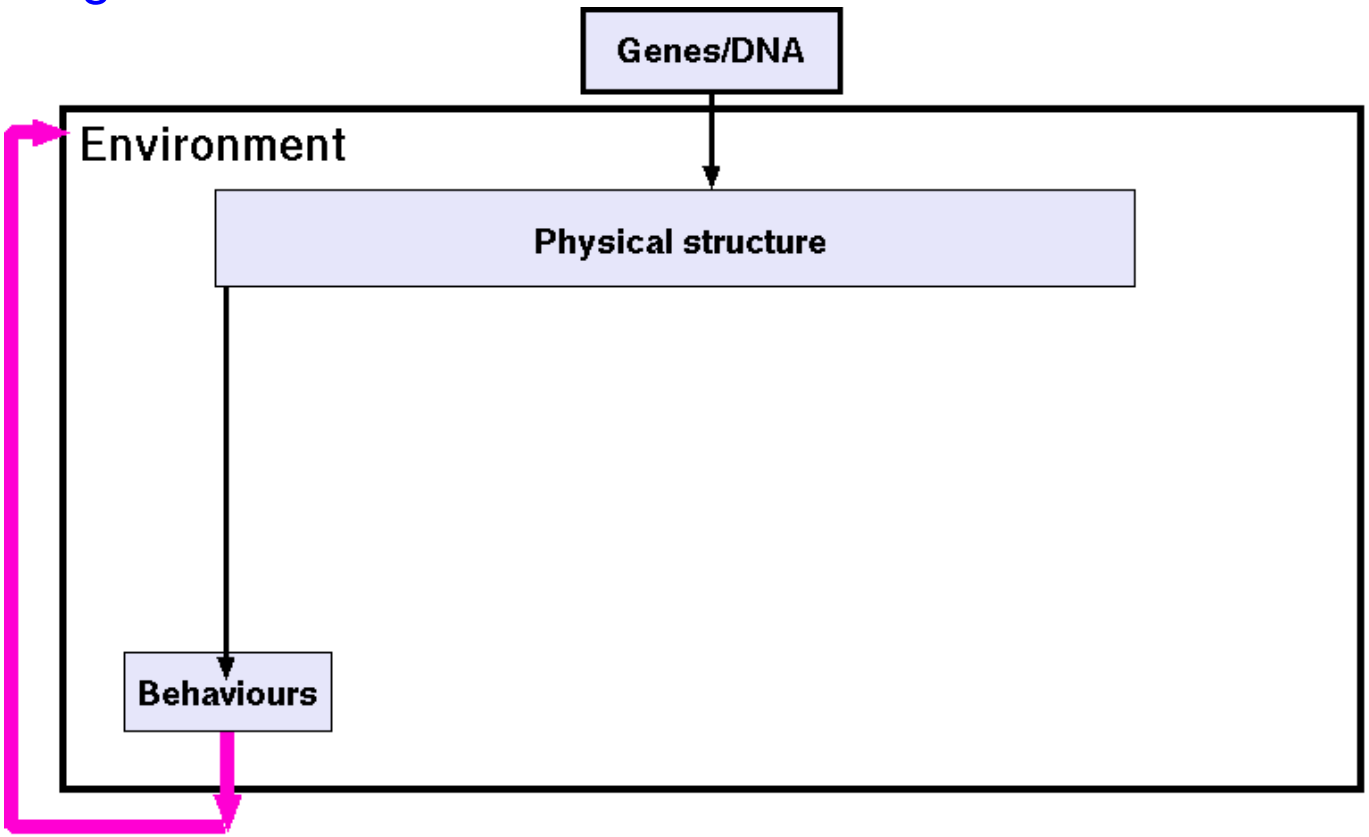
Meta-Morphogenesis of information-processing systems (M-M).

- Evolution produces (sometimes with the aid of learning, or learning and teaching) not only instances of various types but also specifications of instances.
  - E.g. a typical reproductive process does not directly create the body and functions of a new individual.
  - Rather, a fertilised egg, or seed, or spore, is produced, which includes something that controls chemical and other processes to produce (eventually) a new adult ready to reproduce.
- But the specifications are often schematic allowing different instances to be produced on the basis of additional information acquired from the environment.
  E.g. language learning.
- and in some cases
  - specifications of specifications
  - specifications of specifications of specifications
  - ....

Compare software development using class hierarchies and inheritance – including multiple-inheritance.

Individual developmental trajectories

Routes from genome to behaviour : the direct model.



The genome fairly directly specifies the physical mechanisms and architecture of the individual, and those mechanisms fairly directly specify behaviours, including responses to stimulation (innate reflexes).

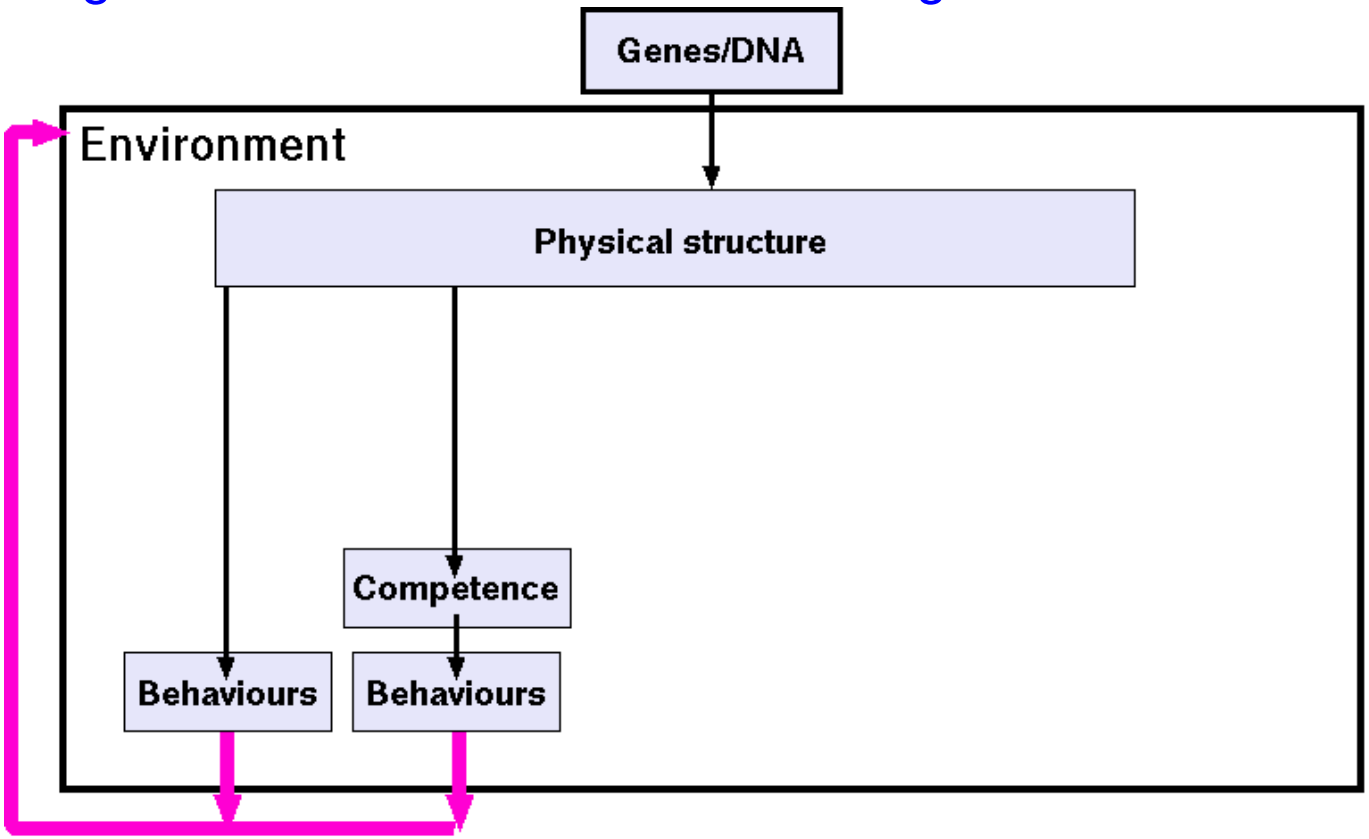
The vast majority of organisms (including micro-organisms) are like this.

Many don't live long enough to learn much – they have to make do with innate reflexes. Other organisms have more "inside the box".

Diagrams adapted from (Chappell & Sloman, 2007), using suggestions from Chris Miall.

Individual developmental trajectories

Routes from genome to behaviour : the two-stage model.

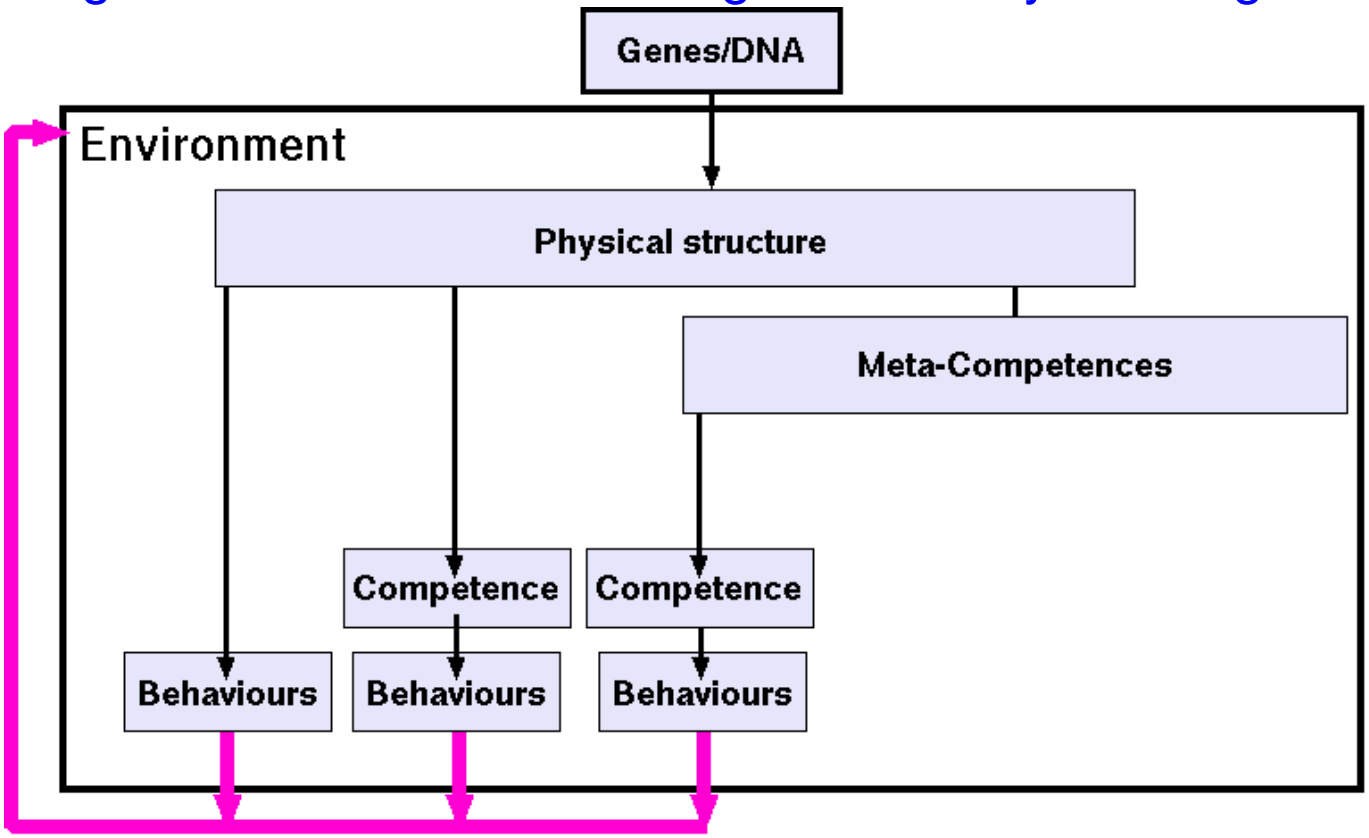


The genome also specifies mechanisms that can generate responses that are triggered in a more flexible way by combinations of circumstances, e.g. internal states as well as external triggers.

Some more complex organisms, instead of having only rigid (reflex) behaviours, also have competences that allow them to respond in fairly flexible ways to the environment: adapting behaviours to needs and contexts.

Individual developmental trajectories

Routes from genome to behaviour : stages added by learning.

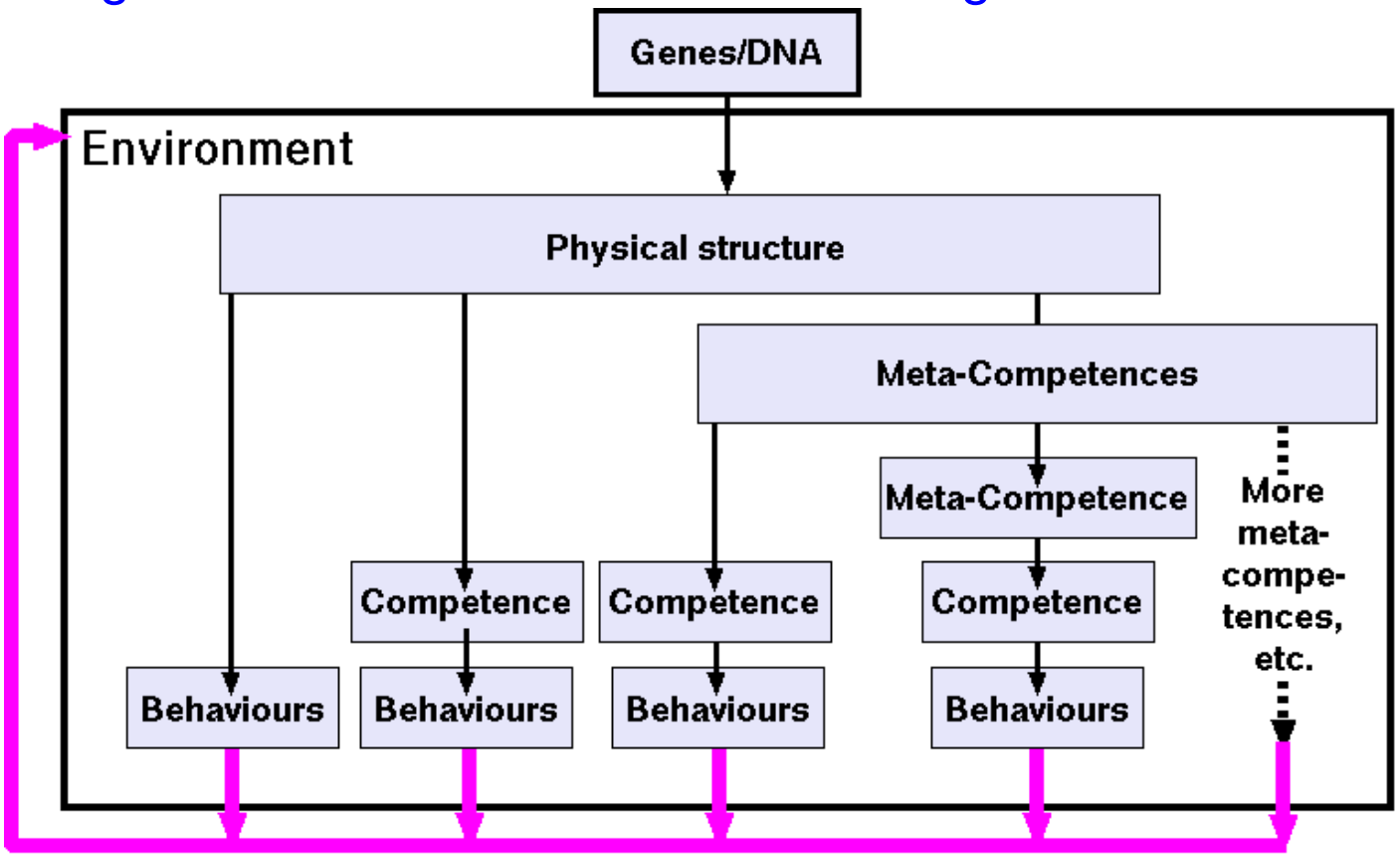


The genome also specifies mechanisms (meta-competences) that can observe the outcomes of the previous sorts of mechanisms in various environments, and, possibly after some delay, on the basis of what has been learnt produces new, more or less flexible, competences. In some cases an evolved competence may be replaced or supplemented with a more abstract and version more widely useful.

Genetically determined meta-competences allow individuals to respond to the environment by producing new types of competence, increasing flexibility and generality.

Individual developmental trajectories

Routes from genome to behaviour : the multi-stage model.



Genome-specified meta-meta competences may use the results of all the previously described processes, combined with genetic developmental mechanisms, to produce higher level meta-competences, again adding flexibility and generality, partly tailored to the individual's environment.

Some can also develop new meta-competences, on the basis of meta-meta competences.

Humans seem to go on developing meta-meta-...competences till late in life. (Sloman, 2008)

Meta-competences and abstraction

The ability to produce meta-competences that can be instantiated in many different ways can be compared with some of the developments in programming languages and programming strategies in the last 60 years.

In particular, it looks as if evolution was often able to discover that something evolved was a special case of something more general that could be instantiated in different ways.

Evolving the more general mechanism requires something like mathematical abstraction.

Are there information-processing mechanisms that "naturally" support this process?

(A sort of "reverse Baldwin" effect.)

This is also a recurring feature of mathematical discovery.

Compare Annette Karmiloff-Smith on "Representational redescription". (Karmiloff-Smith, 1992)

Warning

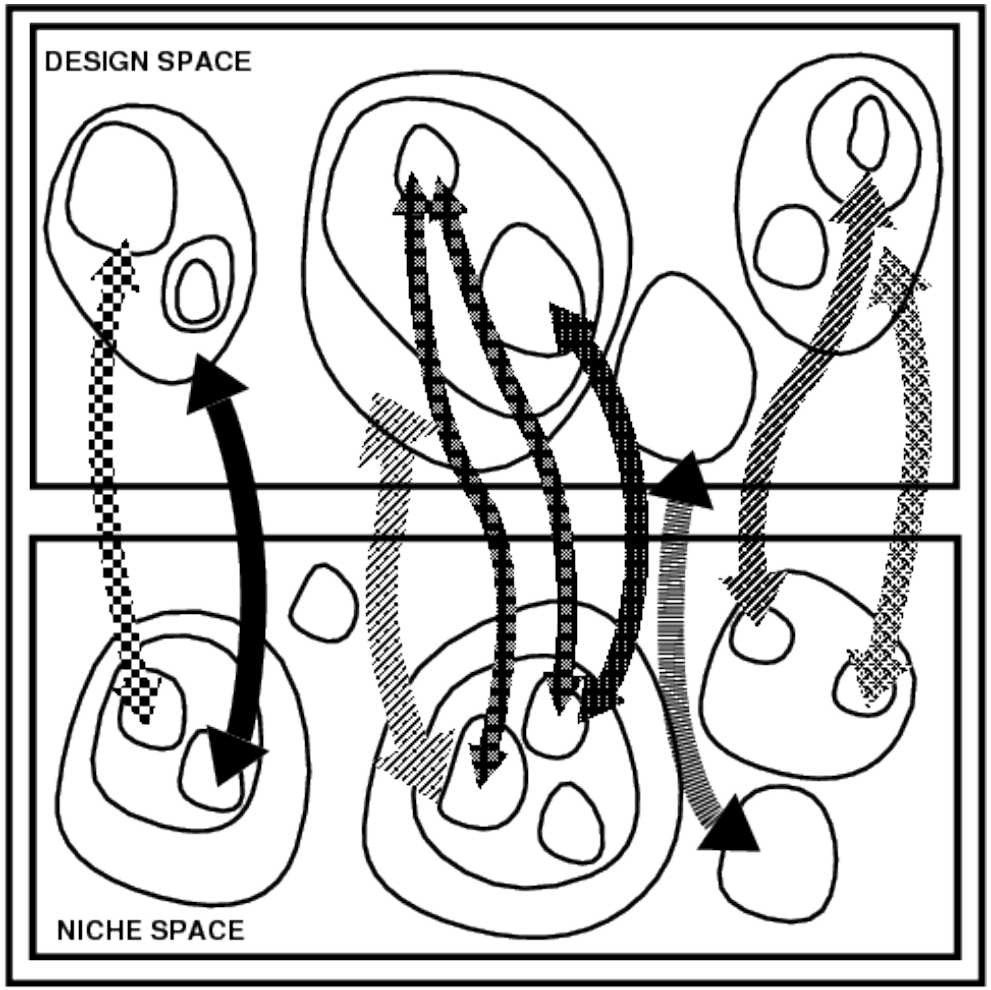
Those diagrams are not to be taken as expressing a precise theory.

They are merely intended to suggest varieties of "multi-layered" interactions between the genome and the environment (which may or may not contain other intelligent systems) during development and learning.

It looks as if one of the evolutionary transitions required for such mechanisms to work was delayed development of some brain mechanisms, e.g. meta-cognitive mechanisms whose operation depends on inspecting results of substantial prior learning and then developing new forms – e.g. replacing collections of empirical summaries with a generative system.

Compare the transition in language development from use of a large collection of linguistic patterns to use of a generative grammar, which, at first, leads to errors because exceptions to the rules are not accommodated in the new architecture.

Varieties of architecture and varieties of requirements



A niche is a set of requirements – e.g. produced by the environment.

Relationships between designs and niches can vary in many ways:

it's not just a scalar fitness function. (But that's a long story).

(X) Layered learning and layered evolution

Can those ideas about varieties of learning and development supported in parallel by the environment and increasingly sophisticated genomes be transferred to ways in which the environment produces increasingly sophisticated ways of producing new genomes?

Obvious examples include

– forms of symbiosis in which two species influence each other's evolution

– uses of individual cognition in influencing mate selection, selective feeding and caring for young (natural eugenics??), uses of selective breeding of kinds of food, or pets, or work-animals

– uses of social pressures in controlling types and rates of reproduction and/or nurturing.

Some closely related ideas have been expressed by Annette Karmiloff-Smith in

(Karmiloff-Smith, 1992)

See her ideas about "representational redescription" occurring in a variety of different domains, in both humans and other species: contrasted with Piaget's ideas about global stages of development.

I think there are strong connections between her ideas concerning types of Representational Redescription and Immanuel Kant's ideas (Kant, 1781) about kinds of ways of acquiring new knowledge, e.g. empirical knowledge and mathematical knowledge.

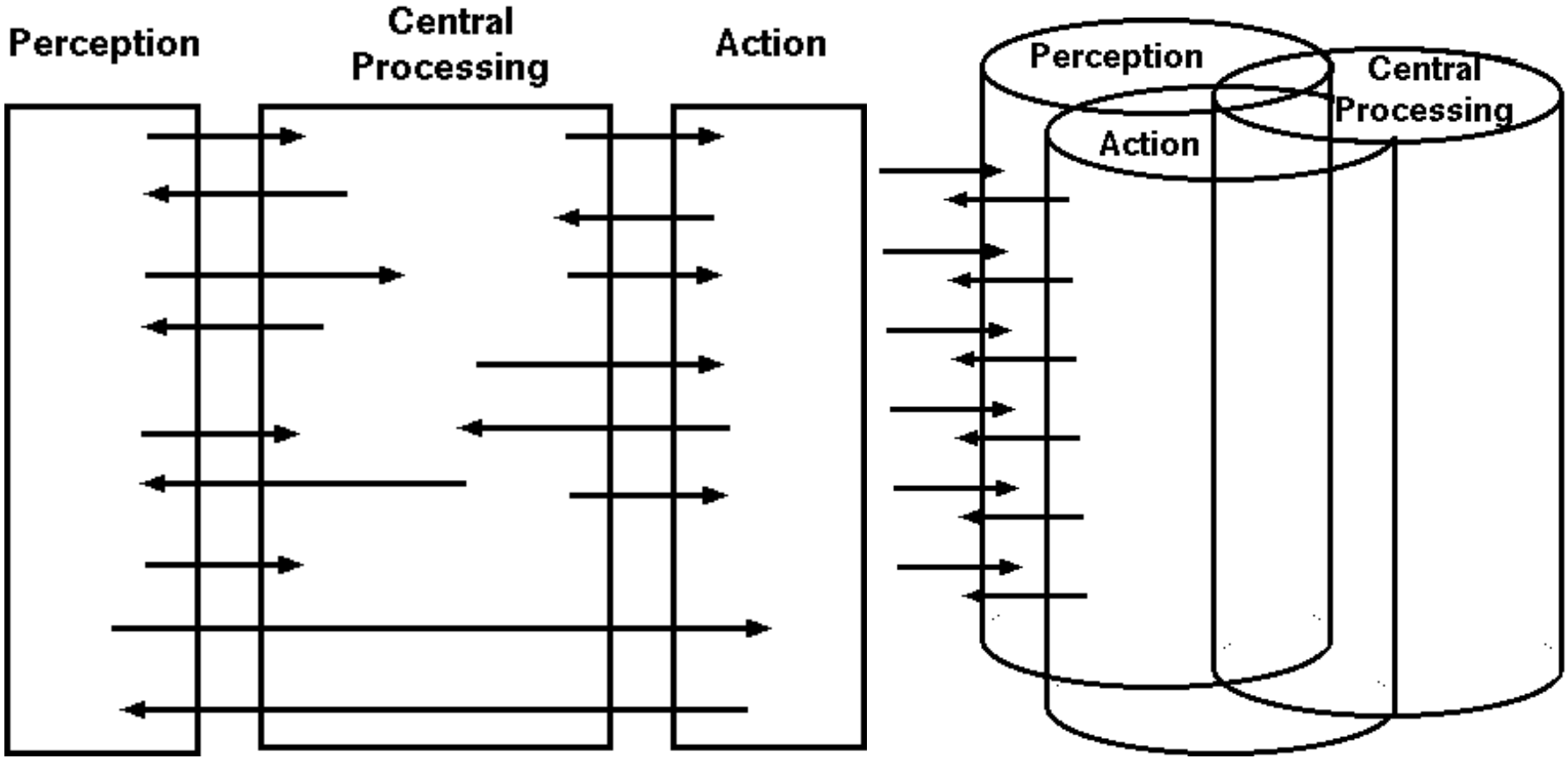
Discussed in `http://www.cs.bham.ac.uk/research/projects/cogaff/misc/beyond-modularity.html`

Skip for now

Ways of slicing information-processing architectures

Thanks to Nils Nilsson (Nilsson, 1998): Towers and layers

Triple Towers (two ways of thinking about them):

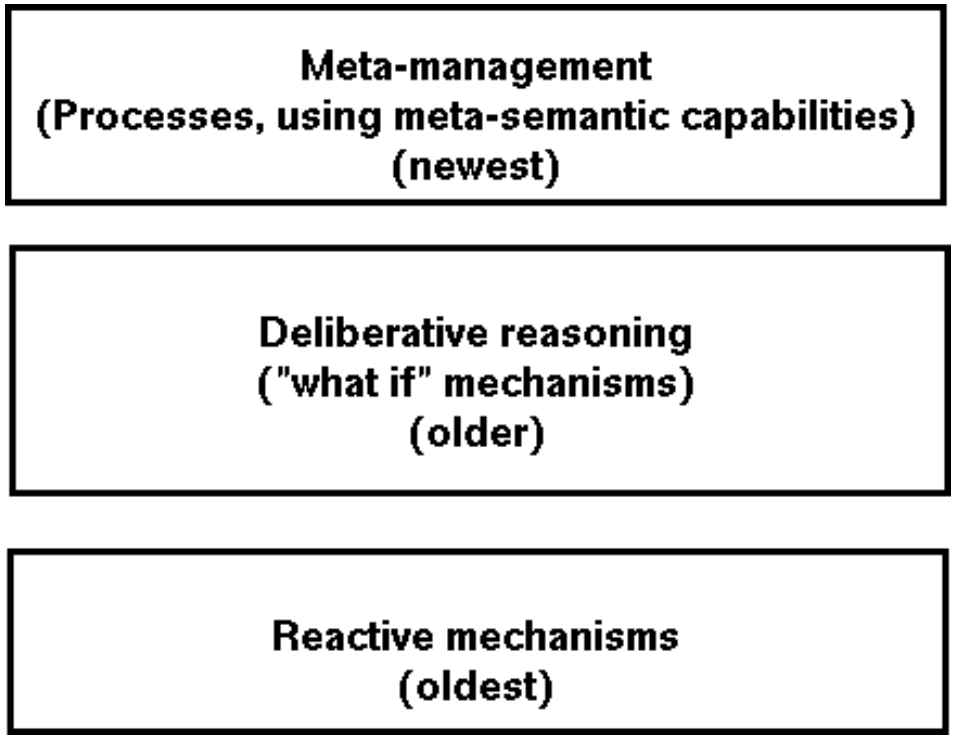


Patterns of information flow may be different in different cases.

J.J. Gibson and others have emphasised the need for the right hand view of perception: perception is generally active, not passive reception.

Ways of slicing information-processing architectures

Triple Layers (oldest and simplest(?) layers at the bottom):



There are probably many intermediate cases.

(Compare Minsky's 6 layers – (Minsky, 2006))

Many researchers in AI/Cognitive science focus on only one layer, or possibly two, often differentiated not by function but by mechanism, e.g. symbolic vs neural.

It is important to make full use of analyses of functions (uses, requirements) for comparing design alternatives.

There are usually trade-offs: no best solution.

Superimpose layers and towers: The CogAff Schema

By dividing sub-mechanisms into “Perception”, “Central Processing” and “Actions”, and also according to whether they are “Purely reactive” or include some “Deliberative capabilities” (thinking about past, remote entities, and possible futures) or including some “Meta-management” (self-monitoring, or self-modulating) mechanisms, we get a 3x3 grid of possible components.

Some boxes, or rows, or columns, may be empty in some designs:

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

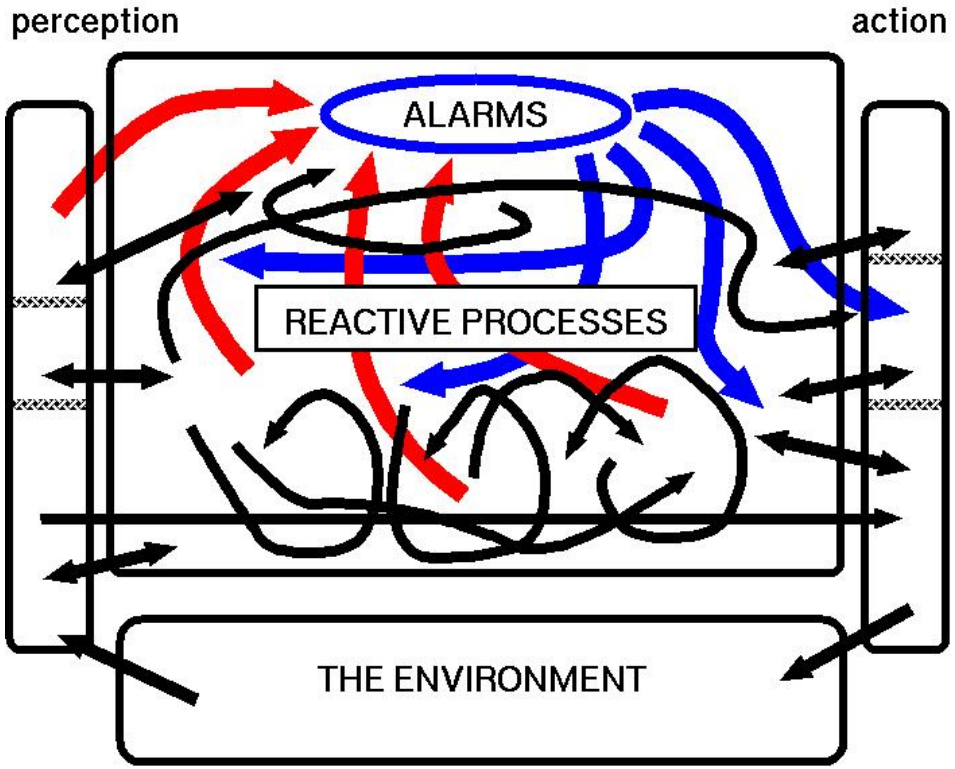
For more on this see the Cognition and Affect project papers:

http://www.cs.bham.ac.uk/research/projects/cogaff/#overview

Special case of the CogAff Schema: Purely reactive

A purely reactive architecture lacks the deliberative and meta-management layers, and can still be quite complex and highly functional – e.g. many insects and other invertebrates?

It seems likely that all microbes and insects (and most other invertebrates?) are like this:



"Alarm" subsystems may detect abnormalities, risks, or opportunities and very quickly globally redirect actions.

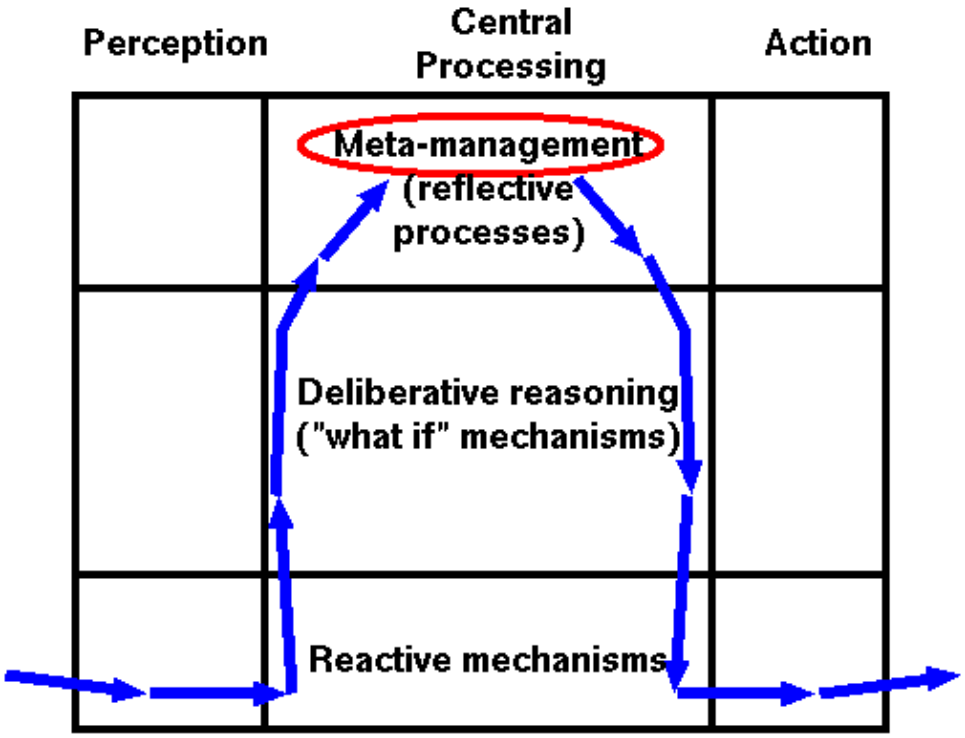
The Five Fs: feeding, fighting, fleeing, freezing and reproduction.

(Most people leave out freezing.)

Another special case of the CogAff Schema

The Omega architecture

E.g. “contention scheduling” systems. (Norman, Shallice, Cooper).

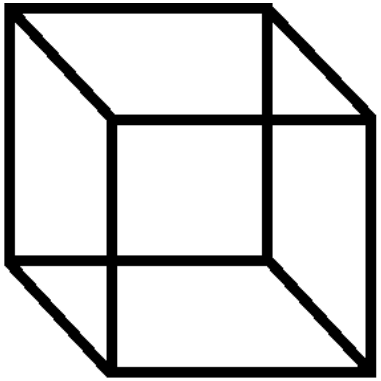


The “Omega” architecture, like many others, uses “peephole” perception and action, restricting the processes to very low level sensing and acting signals.

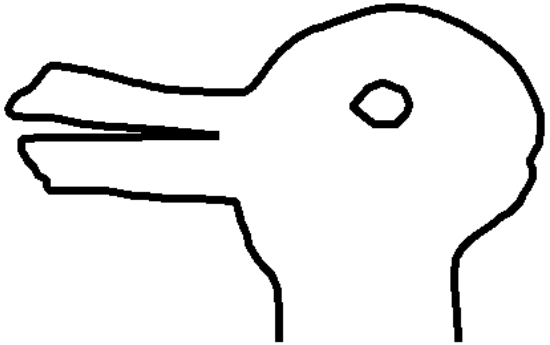
We can get clues as so what is missing by studying the changing perceptual experiences caused by ambiguous figures.

Need for layered perception

These figures are both ambiguous, but when they 'flip', the change in content of what is seen is **only geometric** in the case of the Necker cube on the left, but there's **no geometric change** in the duck-rabbit: only different biological interpretations of portions of the image, and the whole image, including the direction of "facing".



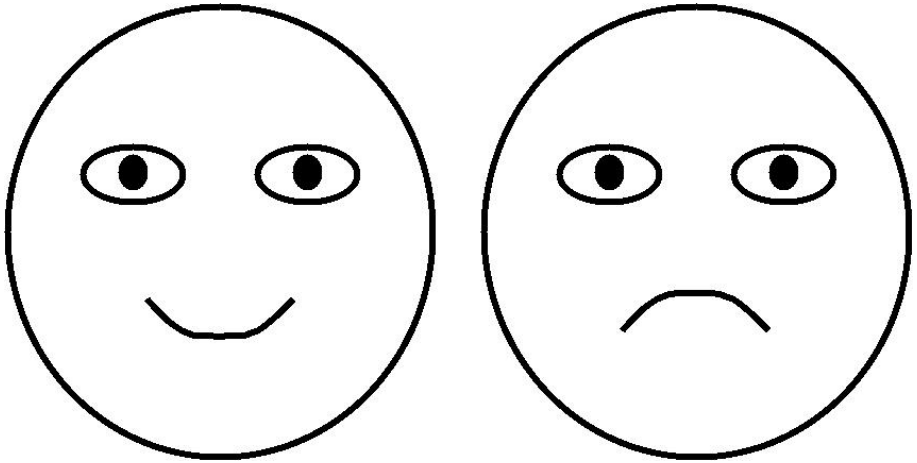
Necker Cube



Duck-rabbit

This is evidence for the visual system adding non-geometric layers of interpretation.

Likewise the "hallucinated" difference between the eyes in the two lower pictures – a hallucinated difference in eyes triggered in some perceivers by the different mouths.



This seems to be evidence of ways in which components of the meta-management architectural layer (with meta-semantic competences) can be shared with perceptual mechanisms.

Why should a system have only one architecture?

Current computing systems have multiple architectural layers: mostly virtual machinery.

Perhaps biological evolution also produced layers of virtual machinery?

They don't all seem to be there at birth.

Should there be innate specifications for the architectures grown?

Or should it be a combined genetic and epigenetic process?

New layers and new mechanisms within layers may be grown as solutions to the problems posed and opportunities provided by the environment.

Computer scientists and engineers who define the architecture of a system as something fixed from the start deprive themselves of an important collection of ideas.

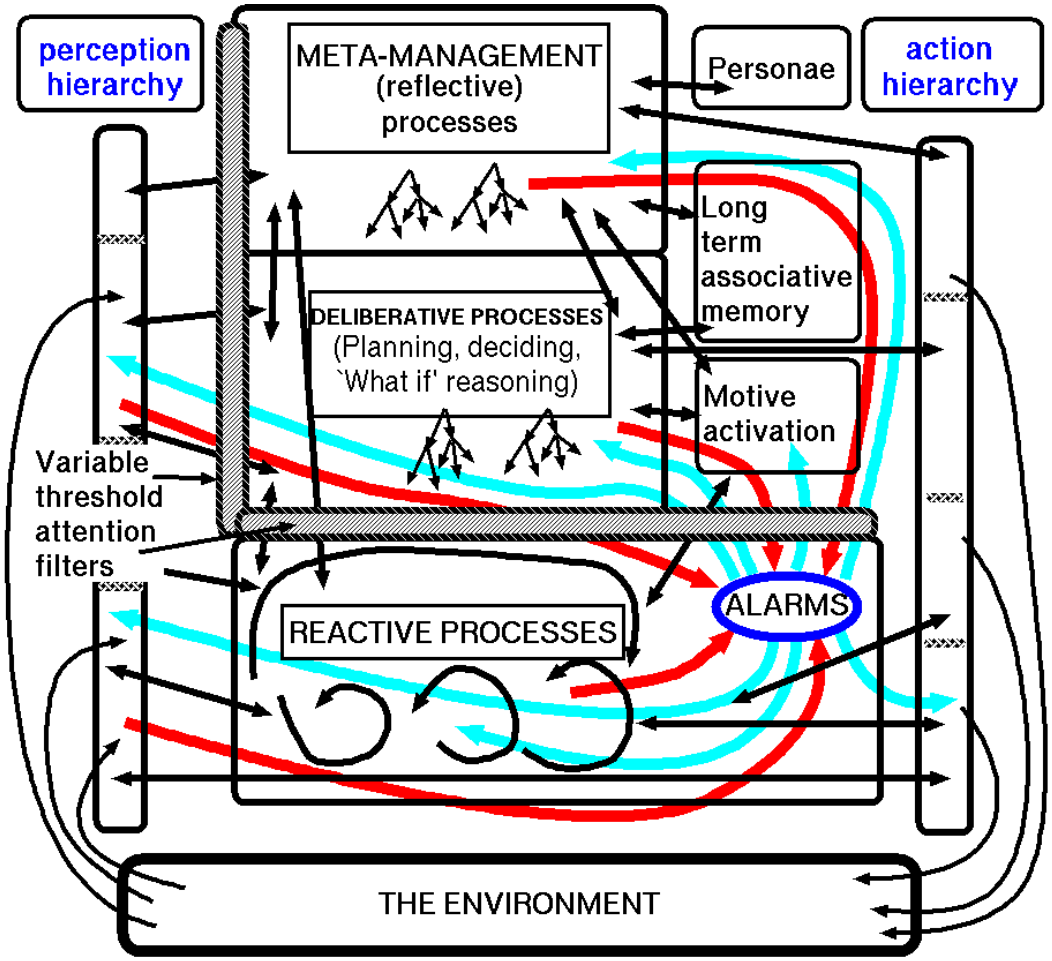
NOTE:

At one time virtual machines were thought of as functionally equivalent replacements for physical machines – so that a virtual machine could, in principle could be replaced by the "real", i.e. physical, equivalent.

Increasingly, however, the use of virtual machinery has led to new types of application, new types of functionality (including new types of self-monitoring and self-control) which could not have been achieved using physical machinery: e.g. rapid re-design while running. (Examples needed.)

H-Cogaff: multi-layered human-like (?) architecture-schema

A special subset of the CogAff architectural schema
Is something like this a framework for adult human architectures ?
Not all instances of the framework need be identical.



The conjectured H-Cogaff (Human-Cogaff) architecture

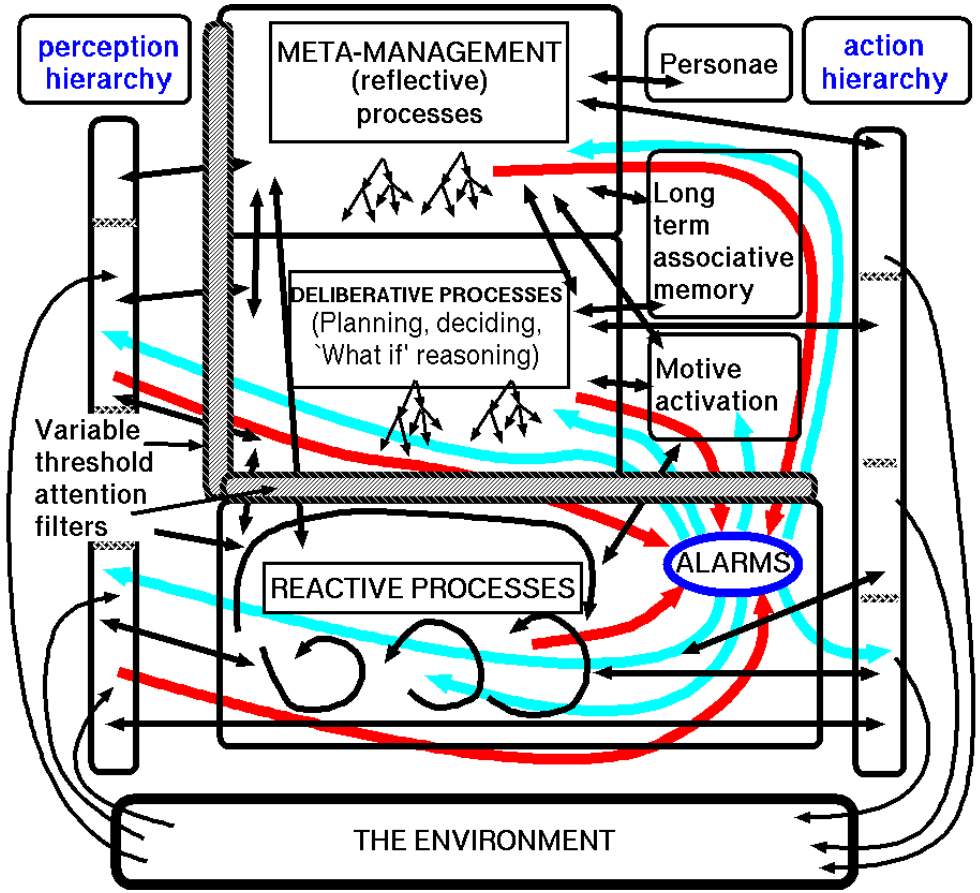
See the web site: `http://www.cs.bham.ac.uk/research/cogaff/#overview`

H-Cogaff: multi-layered architecture
(much more work on this sort of thing is needed)

Such an architecture

- has to somehow be specified in the genome, at least partly/ schematically
- has to grow itself, possibly over many years in some species
- is mostly made up of components that are not physical mechanisms, but instead use virtual machinery, and therefore cannot be detected by physical observations, or physical measuring devices (e.g. brain scanners)

New multi-disciplinary approaches are needed for understanding the requirements, constructing explanatory designs, testing theories, working out practical implications, e.g. for psychology, for psychotherapy, for education, for social policy, etc.



The conjectured H-Cogaff (Human-Cogaff) architecture

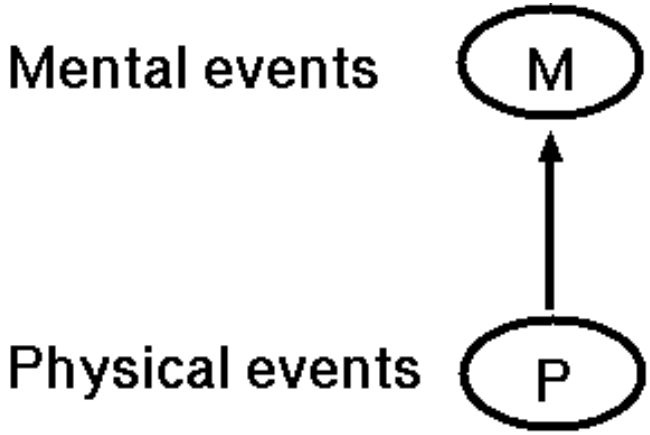
See the web site: `http://www.cs.bham.ac.uk/research/cogaff/`

Will intelligent artificial architectures also have to grow themselves?

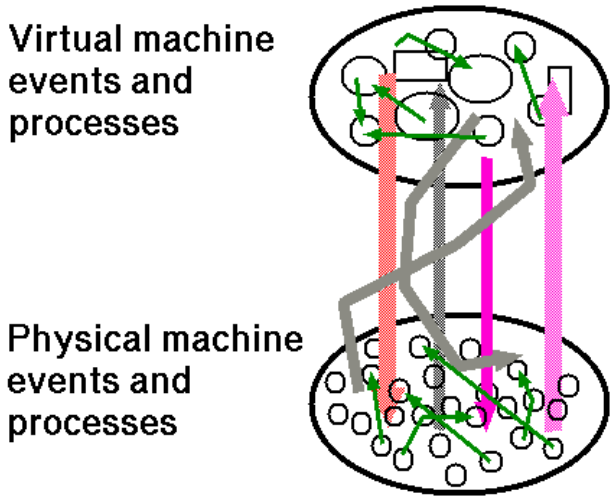
The 20th C Philosophical breakthrough: Virtual machinery

Brief introduction to the philosophical significance of the technology of virtual machinery (not virtual reality) developed over the last six decades: Processes and events in running virtual machines can be causes and effects, despite being implemented in deterministic physical mechanisms.

Bad model:



Good model:



The erroneous picture on the left implies that there is only one-way causation from physical to mental (epiphenomenalism) as in the Mormann & Koch diagram, above.

As the picture on right indicates, we need to think about running virtual machinery that co-exists with, and influences, underlying physical machinery, which it helps to control, even though the virtual machinery is all fully implemented in the physical machinery.

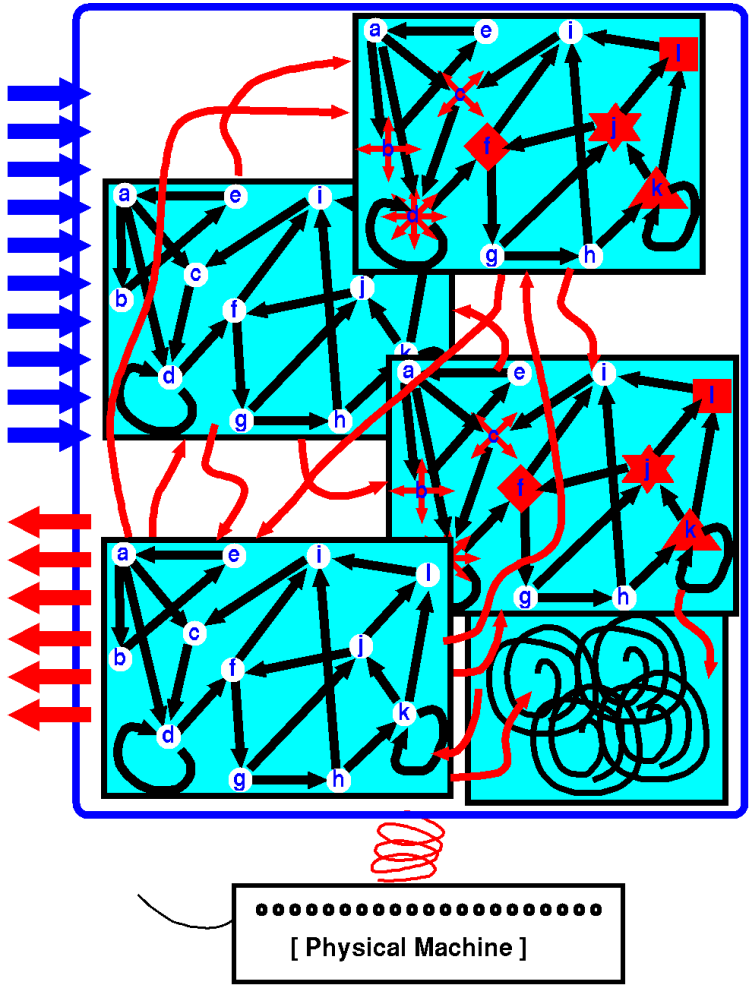
I.e. running software can cause changes in physical hardware, just as increases in poverty can cause increases in crimes involving movement of stolen objects.

Granularity differs at different levels

coarse-grained, relatively simple, VMs may be needed in self-monitoring, self-modulating, self-debugging, self-extending machines. (E.g. minds!)

Running VMs of varying complexity

As crudely indicated here we need to allow multiple concurrent inputs (blue) and outputs (red), multiple interacting subsystems, some discrete some continuous, with the ability to spawn new subsystems/processes as needed.



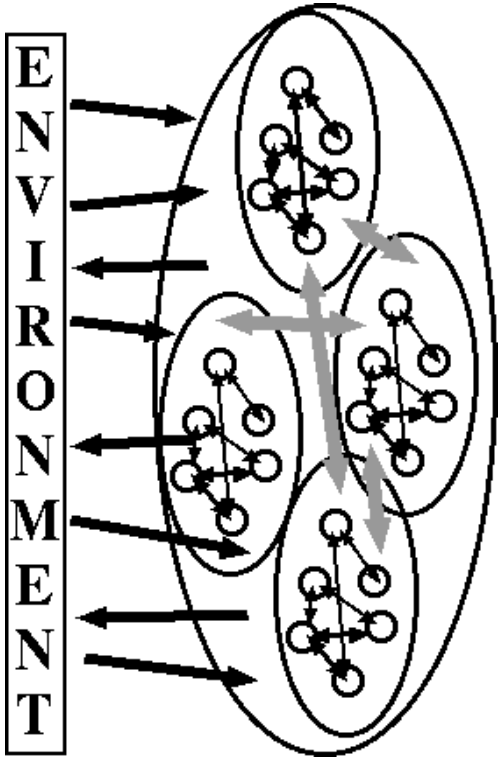
There are multiple feedback loops involving the environment.

Dynamical systems I ("Online intelligence")

Many researchers who emphasise the importance of embodiment, also emphasise dynamical systems —

Especially dynamical systems closely coupled with the environment –

Where the nature of the coupling depends partly on the agent's morphology and sensory motor systems.

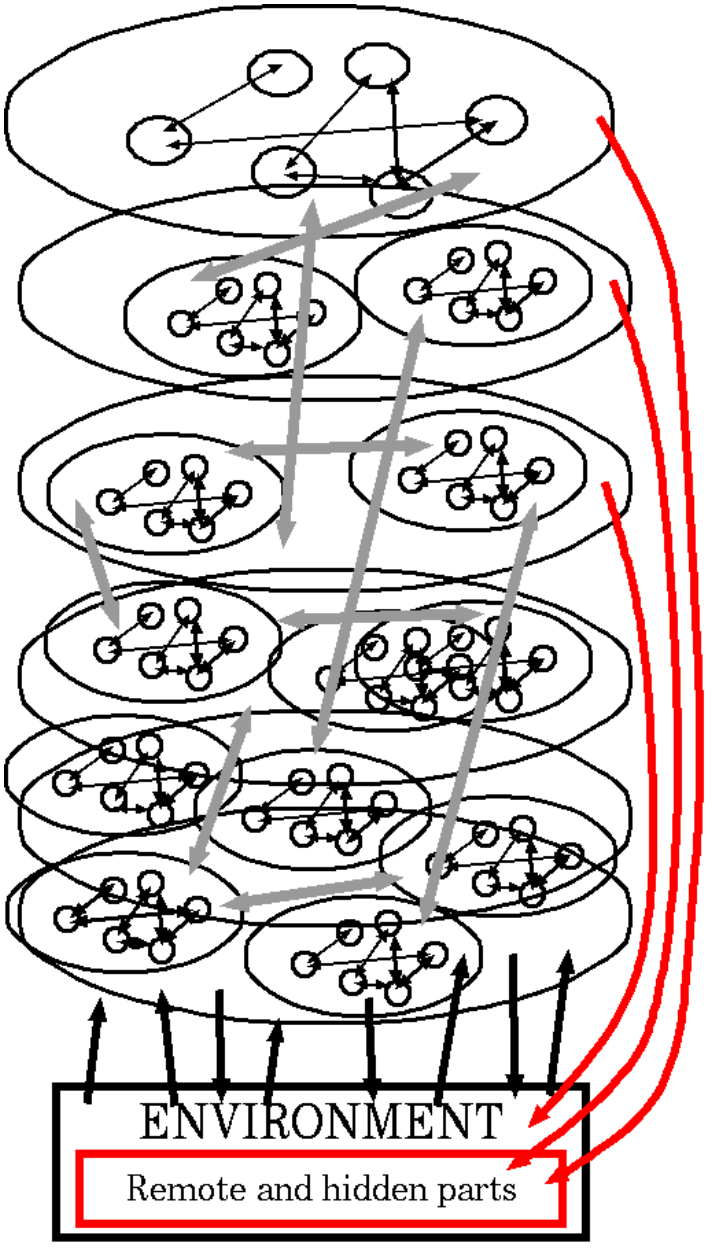


But there are many more types of dynamical system with very different properties – some supporting understanding of the world.

Dynamical systems II ("Online and offline intelligence")

A more complex dynamical system may involve large numbers of sub-systems many of which are dormant most of the time but can be re-activated as needed.

E.g. Perceiving, adaptation, learning, acting, self-monitoring can all involve information processed at multiple levels of abstraction.
Hypothetical reasoning: Science, mathematics, philosophy...
Some of the more abstract processes may be totally decoupled from the environment – but may produce results that are stored in case they are useful...
Which species can do what? – What are intermediate stages:
– in evolution?
– in individual development?



Do not believe symbol-grounding theory: use **theory-tethering** instead.

`http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#models`
(Label proposed by Jackie Chappell).

Dynamical systems III

A point that is obvious to mathematicians, but so far ignored by many roboticists

At a later stage of evolution, the forms of reasoning and representation used began to challenge the brain's working memory capabilities.

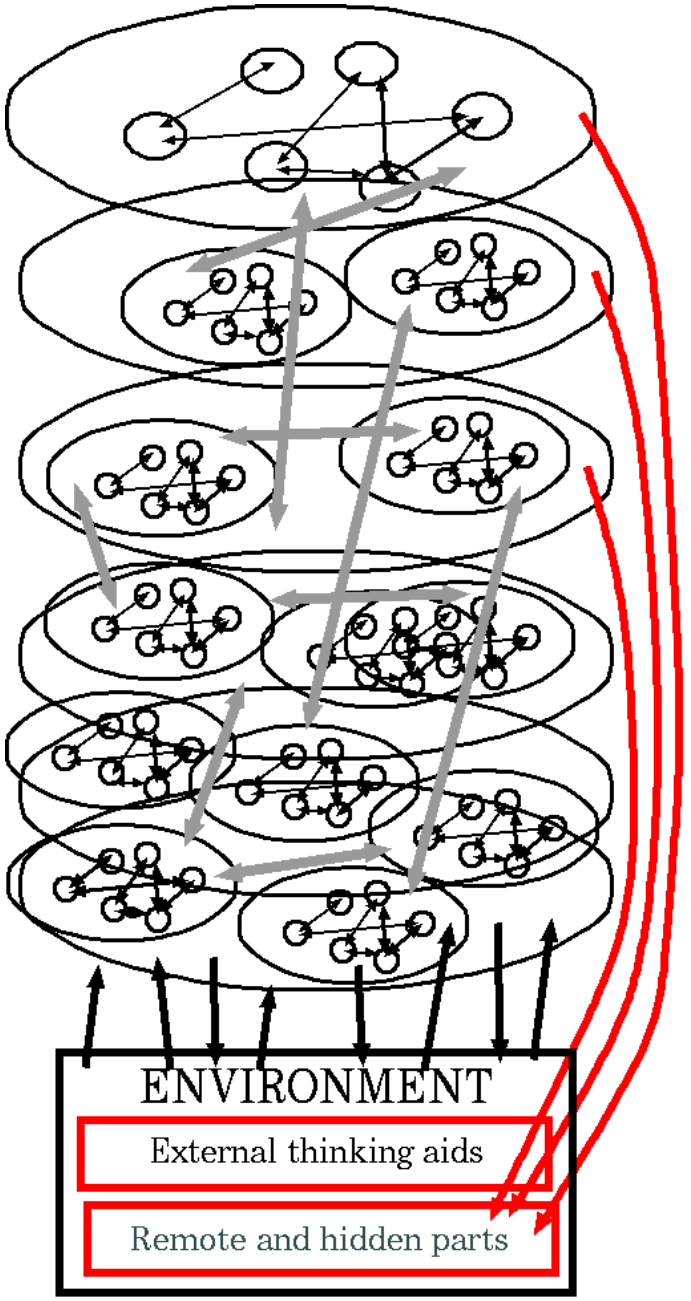
As a result some portions of the external environment were used as extensions of internal forms of representation.

E.g. using diagrams or calculations on paper, or using an abacus when motivated.
(Sloman, 1971, 1978)

[To be revised and extended.
Comments welcome]

Some examples of the use of diagrams in thinking about spatial structures and processes are here:

`http://tinyurl.com/CogMisc/triangle-theorem.html`



Bibliography

References

Chappell, J., & Sloman, A. (2007). Natural and artificial meta-configured altricial information-processing systems. *International Journal of Unconventional Computing*, *3*(3), 211–239. (http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0609)

Davies, P. (1999). *The Fifth Miracle: The Search for the Origin and Meaning of Life*. Simon & Schuster.

Kant, I. (1781). *Critique of pure reason*. London: Macmillan. (Translated (1929) by Norman Kemp Smith)

Karmiloff-Smith, A. (1992). *Beyond Modularity: A Developmental Perspective on Cognitive Science*. Cambridge, MA: MIT Press.

Minsky, M. L. (2006). *The Emotion Machine*. New York: Pantheon.

Nilsson, N. (1998). *Artificial intelligence: A new synthesis*. San Francisco: Morgan Kaufmann.

Sloman, A. (1971). Interactions between philosophy and AI: The role of intuition and non-logical reasoning in intelligence. In *Proc 2nd ijcai* (pp. 209–226). London: William Kaufmann. (http://www.cs.bham.ac.uk/research/cogaff/04.html#200407)

Sloman, A. (1978). *The computer revolution in philosophy*. Hassocks, Sussex: Harvester Press (and Humanities Press). Available from `http://www.cs.bham.ac.uk/research/cogaff/crp`

Sloman, A. (2008). Varieties of Meta-cognition in Natural and Artificial Systems. In M. T. Cox & A. Raja (Eds.), *Workshop on Metareasoning, AAAI'08 Conference* (pp. 12–20). Menlo Park, CA: AAAI Press. Available from `http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0803`

Needs to be extended.

See `http://tinyurl.com/CogMisc/meta-morphogenesis.html`

`http://tinyurl.com/CogMisc/evolution-info-transitions.html`

`http://tinyurl.com/CogMisc/toddler-theorems.html`

`http://tinyurl.com/CogMisc/beyond-modularity.html`