More things than are dreamt of in your biology: Information processing in biologically-inspired robots

Aaron Sloman and Ron Chrisley

School of Computer Science,
The University of Birmingham, UK
http://www.cs.bham.ac.uk/˜axs/
http://www.cs.bham.ac.uk/˜rlc/

THIS PRESENTATION IS ON-LINE IN POSTSCRIPT AND PDF AS TALK 16 AT

http://www.cs.bham.ac.uk/˜axs/misc/talks/

(Last changed August 21, 2002)

Extended version of presentation at

EPSRC/BBSRC International Workshop on
Biologically-Inspired Robotics:
The Legacy of W. Grey Walter
14-16 August 2002, HP Bristol Labs, UK

http://www.ecs.soton.ac.uk/˜rid/wgw02/home.html

The printed paper (in the proceedings) is available here in three formats:

http://www.cs.bham.ac.uk/research/cogaff/sloman-chrisley-wgw02.ps

http://www.cs.bham.ac.uk/research/cogaff/sloman-chrisley-wgw02.pdf

http://www.cs.bham.ac.uk/research/cogaff/sloman-chrisley-wgw02-html

Acknowledgements

We are grateful for help from

Anonymous reviewers, and
Luc Beaudoin, Catriona Kennedy,
Brian Logan, Matthias Scheutz, Ian Wright,

and other past and present members of the
Birmingham Cognition and Affect Group

and many great thinkers in other places

Related papers and slide presentations can be found at
http://www.cs.bham.ac.uk/research/cogaff/
http://www.cs.bham.ac.uk/˜axs/misc/talks/

This work is funded by a grant from the Leverhulme Trust for work on
Evolvable virtual information processing architectures for human-like minds.

Abstract

- Work in biologically-inspired robotics (and in several other research fields) has suffered from ontological blindness, an inability to see, or think about, certain abstract entities and properties.
- In particular, there is a blindness regarding informational entities and processes present in biological systems.
- We counter this
  - by showing how biological systems can be usefully understood as information-processing machines.
  - By drawing attention to important varieties of information processing in biological systems sometimes not noticed even by people who accept that organisms are information processors.
  - By presenting an “architecture schema” that can usefully drive such research, e.g. by helping to reduce ontological blindness.
- It is hoped that this will better enable both the construction of biologically-inspired robots and the development of explanatory theories and models in biology and psychology.

Overview

- **What is information and what is information processing (IP)?**
- **Ontological blindness:**
  - **What it is,**
  - **How you get it,**
  - **How to get rid of it,**
  - **What the benefits of curing it are**
- **The CogAff IP architecture schema**
  - **How it helps us see some new tasks and problems**
  - **How it helps us think about possible solutions**
- **Links between CogAff and:**
  - **Empirical research in biology;**
  - **Design in AI and robotics.**

THIS TALK GIVES ONLY A BRIEF SKETCH: SEE OUR WEBSITE FOR DETAILS:
http://www.cs.bham.ac.uk/research/cogaff/

Two ways of being biologically inspired

Direct: Replication of –

- structure, including physical mechanisms,
- behaviour (external only, or both external and internal – e.g. manipulation of data-structures, patterns of neural activation)

Indirect: Conceptually-mediated design –

- Use the concepts that are required to understand biological systems in order to design artificial ones that may have no counterpart in nature; (e.g., Raibert's 3D hopper, or flying machines).
- Not just a matter of determining how organisms do what they do; but
- Also determining what are they doing: finding a set of concepts to characterise behaviour/internal processing

NOTES:
Most cases of "inspiration" probably involve both forms.
Normally in an iterated way.

Both approaches may be impeded by ontological blindness,
e.g. blindness to some relevant functions or to some appropriate description of behaviour.

David Marr on two ways:

. . . "mechanism-based approaches are genuinely dangerous. The problem is that the goal of such studies is mimicry rather than true understanding . . . "

"If we believe that the aim of information-processing studies is to formulate and understand particular information-processing problems, then the structure of those problems is central, not the mechanisms through which their solutions are implemented."

Both on page 347 of D.Marr *Vision*, 1982

NOTE: He misleadingly used the word "computational" for the problem-oriented rather than mechanism-oriented descriptions.

What is information processing?

- In a sense we all know - so no answer is needed: it's what we refer to when we talk about
  - having information
  - acquiring information
  - needing information
  - using information, ... etc.

- But we have only an intuitive understanding and there is a need to make this explicit, eventually: there's no time for a full analysis today.

  For more on this see talks 4 and 6 available here
  http://www.cs.bham.ac.uk/˜axs/misc/talks/

- Clarifying "information processing" depends on clarifying the notion of "information", and that has several interpretations.

- For now we use only the lay notion of "information", linked to "meaning", "content", "reference", "inference", etc.

  (Not the technical mathematical notion of information, which has little to do with meaning.)

What information processing involves

So information processing involves: acquiring, storing, deriving, manipulating, inferring, analysing, interpreting, and above all using information in that "lay" sense.

- **This has no commitment as to**
  - **whether the processing is in machines or in organisms,**
  - **whether it is digital(discrete) or analog(continuous),**
  - **whether it is encoded explicitly and locally or implicitly in distributed form,**
  - **whether it is encoded physically (writing, pictures, ...) or in virtual machines (e.g. in abstract data-structures, rules, axioms, networks, graphs)**
  - **Whether it is encoded within the information user or external to it, in the environment.**
- **All these are notions we need to analyse but don't have time for today. (Though every software engineer understands and uses them.)**

See also:

Brian C Smith, 'The foundations of computing' in M.Scheutz, ed., *Computationalism: New Directions*, MIT press, 2002.

The papers on varieties of representation here http://www.cs.bham.ac.uk/research/cogaff/

What is information? 1

- We use "information" to mean:
  - Something like the ordinary notions of "content" and "meaning"
  - Not the Shannon-Weaver notion of information since:
    * information can be false.
    * items of information can stand in relations like consequence, contradiction and relevance
    * items of information can understood or misunderstood.
    * it has nothing to do with unexpectedness: information content is sometimes completely predictable

- Information is non-physical (albeit physically realised)
  - But this does not make it unsuitable for use in biology: compare "niche", "gene", etc.
  - It does, however, mean that specialised methodologies are required for identifying, and explaining, information processing. These differ from the methods of the physical sciences.
  - Compare the differences between identifying and fixing a faulty circuit and identifying and fixing a software bug – which may manifest itself in many physical implementations.

What is information? 2

The full answer is quite complex.

Partial answers can be found in talk 4 and talk 6 here:
http://www.cs.bham.ac.uk/˜axs/misc/talks/.

It's probably a mistake to seek an explicit definition of information

- Rather, "information", like "energy", will be implicitly defined by the role it plays in theories.
- Roughly, when you know:
  - the forms that information can take,
  - the variety of contents it can have,
  - the various ways it can be
    - acquired,
    - manipulated, analysed, interpreted,
    - stored, transmitted, tested,
    - and, above all, used,
- Then you know (to a first approximation) what information is.
- That knowledge evolves over time, like our knowledge of what energy is.

Information processing: only artificial?

Some take information processing to be something that only artificial machines can do: e.g., Steven Rose:

"Pinker goes on at great length about how what minds do is deal with information. I want to insist that what minds and brains exactly are not, are computers. That is, computers deal with information, you pull out a file in your computer, it's a bit of dead information, you manipulate it, you put it back and it stays there until you want it again. What brains and minds do, is deal with meaning." **(http://www.abc.net.au/science/descent/trans2b.htm)**

- **The mistake here is to think that information has to be "dead"**
- **This is to confuse information with the vehicle of information (e.g. bit patterns or data-structures.)**
- **Rather, one can see "living" meaning as arising out of a complex of information processing activities and capabilities.**

Information processing: only human?

Others take information processing to be something that only humans can do: e.g., Maturana and Varela. They believe (as summarised by Boden, 2000):

- **Information processing requires communicating meanings with a language.**
- **"Function" applies in a non-metaphorical way only to human design – not biology: "*the neuron cannot be considered as a functional unit of the nervous system*".**
- **Attempts to understand, e.g., DNA as an informational code are fundamentally misleading.**
- **Organisms do not have input or output, only perturbations: the organism itself cannot distinguish between internal vs. external perturbations.**
- **Talk of representation may be "*metaphorically useful, but inadequate and misleading* [in revealing] *the organisation of an autopoetic system*"**

Information processing: natural and artificial

We disagree with both Rose on one hand:

(he claims only artificial systems do information processing)

and Maturana and Varela on the other:

(they claim only humans using a language for communication do information processing)

Information processing as the term is normally understood is something which occurs in both:

- **Natural biological systems in general (not just humans), and**
- **Artificial systems.**

But that doesn't trivialise the notion: not everything is usefully understood as an information processor.

Natural information processing: The very idea

Why might one think information processing occurs in biological systems at all?

- An essential feature of life is energy budgeting
- An organism which is better at selecting when and on what to spend energy is more successful

  (e.g. compared with one that is merely physically driven by ambient energy sources, e.g. pulled down by gravity.)
- Good selections require information about: generalities, contingencies in the world, oneself, the current context

  (whether the information is represented and manipulated explicitly in formal structures or implicitly in the operation of selection mechanisms)
- So information processing is an essential part of being alive, being biological.

However, evolution has produced myriad forms of IP and we have barely begun to understand the diversity of such forms

(not to be confused with the closely related diversity of physical forms and physical behaviours)

Attacking the causes of blindness 1

An approach to curing ontological blindness is to remove the reasons why people cannot see information processing (IP) concepts in biology.

Sometimes we don’t see X because we:

- Are unwilling or even afraid to see X (for whatever reason)
- Are unable to see X (e.g., because we lack the appropriate concepts)
- Don’t think seeing X is possible (e.g. because our favoured theories imply that it is impossible), or
- Think X is an illusion
- Lack the educational background, e.g. being ignorant of how software engineers regularly work with virtual machines of many types.

Attacking the causes of blindness 2

People might be blind to the idea of IP in biology because:

- They misunderstand the nature of IP –
  (e.g. Rose, and Maturana and Varela, above; IP $\neq$ GOFAI)
- They think "information isn't physical, so it isn't real"
  (but consider niche, gene, etc. – are poverty and ignorance real?)
- They fear that biology-as-IP will diminish the "wonder of life"
- They believe that "sight" isn't necessary for design: Just evolve it! However:
  - Unlikely (certainly not guaranteed) to achieve results
    (e.g. producing human-like or chimp-like intelligence).
  - All the disadvantages of theft over honest toil: little understanding is gained to support the next round of design
- Wrongly believe that causal powers of virtual information processing systems require causal gaps in physics.
  (They don't understand how software architectures running on physical hardware have causal powers – a topic discussed at length on our web site.)

Implications of “the IP stance”

Regarding something as an information processor has important consequences.

This stance leads to very different research questions from a stance that views organisms as physical systems in a physical environment, to be studied using the methods of the physical sciences.

E.g. it provokes questions:

- What kinds of information does the system use?
- Where does the information come from?
- How is it acquired?
- In what forms is it expressed or encoded? (e.g. what syntax is used?)
- What is it used for?
- How is it stored, transformed, interpreted, used?
- What sort of architecture, using what sorts of mechanisms enable this?

These questions don’t make much sense if asked about a rock or a thundercloud – unless stretched a lot.

It's not easy

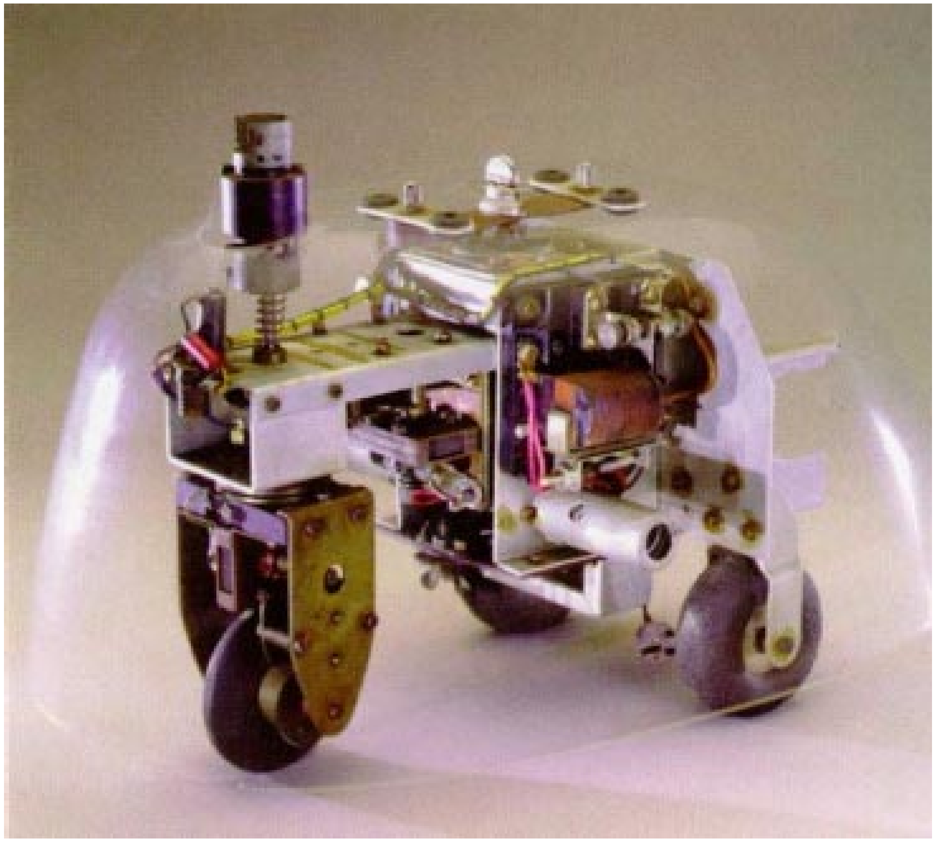
You can't tell what information an organism is using just by looking at

- the environment,
- the organism's behaviour
- its physiology, including neural circuitry

Compare trying to work out what operating system, compilers, algorithms, data-structures, inference methods, a computing system is using by empirical observation.

For a start: you need to have an appropriate set of concepts to describe an information processing system, and its environment – its "niche".

What Walter's tortoise taught us



Working out what this is doing may be relatively easy.

Betty Crow the hook-maker

Working out what she is doing will be much harder.

See the video here: http://news.bbc.co.uk/1/hi/sci/tech/2178920.stm

Betty and Abel learnt to use bent wire to fish a bucket of food out of the vertical tube (as in the picture).

After Abel flew off with the hook (why??) Betty tried using a straight piece of wire for a while, and failed. She then wedged it somehow and bent it with her beak, making a hook, which she used successfully thereafter.
Reported in Nature and shown on BBC TV (August 2002).

Do we know what sort of ontology is relevant to describing and explaining what Betty did?

Did some part of Betty's mind have information in advance about what the consequences of bending the wire would be, and drive a search for a suitable means to do it?

If so we can ask all the questions from our previous list of IP questions about Betty,

People who have first hand experience of designing and implementing various kinds of reactive and deliberative plan-using sytems will be able to ask more specific questions.

People with no such experience may be blind to the problems.

The dangers of over-ascribing

- **We must be careful, of course, not to attribute unnecessary mechanisms to biological systems.**
- **We must not let the way we see (and how we would generate) a behaviour blind us to the infomation processing that nature has stumbled upon to achieve that behaviour.**
- **Remember Lloyd Morgan's Canon (1894):**

  "In no case may we interpret an action as the outcome of the exercise of a higher psychical faculty, if it can be interpreted as the outcome of one which stands lower in the psychological scale."

The dangers of under-ascribing

- **But Lloyd Morgan later published a rider to his Canon (1900):**

  "The canon by no means excludes the interpretation of a particular act as the outcome of the higher mental processes if we already have independent evidence of their occurrence in the agent."

- **Or, as Einstein put it:**

  "Everything should be made as simple as possible, but not simpler."

  For example, we must not let the fact that all sophisticated information processing must ultimately be realised in simple reactive mechanisms blind us to the presence of the larger explanatory pattern.

Is blindness holding us back?

Ontological Blindness

Ontological blindness is the inability to be sensitive to or think about the existence of a certain class of entities or properties.

- **We suspect that many people (including ourselves!) suffer from various degrees and kinds of ontological blindness when observing, thinking about and explaining biological systems.**
- **Even people who are not blind to the general presence of information processing (IP) entities and properties in biology may be blind to some of the specific, very unobvious, varieties of information and information processing, e.g.**
  - **perception of affordances,**
  - **some kinds of reasoning (e.g. required in nest-building birds)**
  - **some kinds of learning and problem-solving (e.g. why are people so surprised at Betty's achievement?),**
  - **some kinds of communication (e.g. between whales?).**
  - **Features of niches for creatures of interest.**

Some questions for the ontologically sighted

- What information enabled Betty the crow to bend the wire?
- Did she understand why the straight wire did not work? Did she grasp the consequnces of bending?
- Where did she get the information?
- How did she absorb it?
- In what form was it stored?
- What enabled her to determine its relevance?
- In what way did the stored information drive her behaviour?
- While she was bending it, was the goal of lifting the bucket somehow stored?
- In what form was it stored?
- How was that goal re-activated after the goal of bending the wire had been achieved?

Work in AI has already investigated various forms of answers to these questions. Perhaps there are more types of answers to be found, or alternative questions to be posed about Betty as an information processor.

Such questions may have similar answers for different species or different individuals — even if the physical details of their brains are all different.

Other examples from the WGW02 workshop

- **Mandyam Srinivasan described research on bees which implied that there were four processes of information transformation**

  1. Scouting bees find where food is and somehow store distance and direction information in a changed state.
  2. On returning to the hive the stored information is transformed into a bee dance
  3. Other bees perceiving the dance transform the information into a changed state
  4. That state then drives process of flying out in the approriate direction for the appropriate distance.

  All the transformations work as required only under some physical conditions: The environment shares some of the information burden.

- **Michael Arbib reported a sequence of increasingly sophisticated information processing capabilities found in animals (up to proto-deliberation), and described information processing architectures that might explain them.**

  That illustrates the process of analysing possible evolutionary trajectories and exploring neighbourhoods in design space, recommended in CogAff papers.

For details see workshop proceedings.

Manifestations of ontological blindness to IP

Blindness to information processing phenomena may either be global or specific:

- **Blind to information processing (IP) as a whole**
- **Blind to alternative information processing (IP) characterisations E.g.**
  - **Using only 'low level' physical descriptions of the environment, as opposed to descriptions relevant to an organism's niche.**
  - **Describing information-processing mechanisms in brains in terms of very low level information processing: e.g. like describing the data-structures used in a compiler or operating system without saying what they refer to.**

More on causes of ontological blindness to IP

- **One understandable reason why people may be blind to alternative IP characterisations of biological systems:**

  Due to its non-physical nature, information and the processing of it are much harder to detect than lower-order, physical properties of systems, and resistant to numerical forms of representation

- **Hard to think in terms of/recognise a conceptual/representational scheme different from one's own**

  E.g. many who know nothing about software engineering find the notion of a "virtual machine" hard to grasp, and doubt that virtual machine events can be causes.

- **Conceptual scheme of system being studied (e.g. insects, bats, apes, infants) may be incommensurable with/in conflict with that of the theorist/designer**

  It's always risky to use human language to describe information processing in non-human animals – or even in human infants!

- **Ignorance of one's own conceptual scheme (and its limits).**

  (Very common in the history of science.)

Precedents in the history of science

Many advances in science have involved

- **not new laws but**
- **new views of what is possible**
  – what exists and can exist.

(Compare the analysis of science as the study of what is possible,
in Chapter 2 of The Computer Revolution in Philosophy (1978),
online here http://www.cs.bham.ac.uk/research/cogaff/crp/

Also M.Boden *The Creative Mind: Myths and Mechanisms*, (1990).)

Curing by doing

- The best way to cure ontological blindness might be to get people to engage in certain kinds of activity involving the entities to which they are blind:
  - designing,
  - explaining,
  - constructing,
  - debugging,
  - interacting (including using), etc.
- In the case of information processing (IP) blindness, such activities would be facilitated by having a general schema for different possible IP architectures, along with their trade-offs: their capabilities and limitations, strengths and weaknesses, costs and benefits.

Curing ontological blindness: Summary

At least four methods for curing ontological blindness:

1. **Interaction with (models of and actual) IP systems**
   Including designing, implementing and debugging.
2. **Conceptual arguments**
3. **Empirical results**
4. **Attacking the causes of blindness**

... A systematic schema for possible information processing architectures can be of assistance in at least 1, 3 and 4.

To this end, we offer the CogAff architecture schema as an example of a framework which helps systematize the variety of information processing systems one might encounter in biological organisms (including humans).

The CogAff Architecture schema: Key Ideas 1

The CogAff schema is the result of many years of research combining approaches from AI, philosophy, psychology, neuroscience, linguistics and biology.

It involves:

- **Dividing up functionality, assuming rough architectural divisions,**

  E.g. (1) perception, (2) internal state changes and storage, (3) action

- **Distinguishing varieties of types of information processing**

  E.g. Varying in levels of sophistication, in forms of representation used, in types of semantic content, etc. For instance reactive, deliberative and meta-management processing.

- **Investigating mechanisms and forms of representation suited to the various kinds of functionality**
- **Investigating relations of the above to possible niches (sets of requirements)**
- **Investigating developmental and evolutionary trajectories**

  See talk 15 here http://www.cs.bham.ac.uk/~axs/misc/talks/

Our analysis is based on a study of many examples from humans and other organisms.

The CogAff Architecture schema: Key Ideas 2

Divide up functionality, assuming rough global architectural divisions:

1. **Perceiving things in the environment**
   including perceiving other agents and their mental states
2. **Internal information manipulation and storage**
   including adaptation, learning, goal or drive activation, selection, etc.
3. **Acting on the environment**
   (including other agents)

Compare Nilsson's "triple-tower" architecture, in his
***Artificial Intelligence: A New Synthesis* 1998.**

- **All that seems trivial – until you try to make it fit a wide range of animal competences.**
- **Doing that reveals great diversity in the implementation of the functional divisions and shows how they can be more blurred in some organisms than in others.**

PILLARS + LAYERS = GRID

Combining the two distinctions

- the functionality (represented as three vertical pillars)
- the sophistication (represented by horizontal slices)
- gives a 3x3 grid of types of components of an information processing organism or robot.

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

PILLARS + LAYERS = GRID

Combining the two distinctions

- the **functionality** (represented as three vertical pillars)
- the **sophistication** (represented by horizontal slices)
- gives a **3x3 grid of types of components** of an information processing organism or robot.

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

Concurrently:

- Different perceptual layers will process information relevant to different central layers: **Perceptual processing at different levels of abstraction**

  e.g. simultaneously seeing a face as approaching and as angry, seeing something as having a certain size and as being graspable.

- Different action layers perform acts selected and/or directed by different central layers. **Actions occur at different levels of abstraction, sequentially or concurrently**

  E.g. talking and breathing, or moving a hand and signalling to someone.

Elaborating the grid

THIS IS A GENERIC ARCHITECTURE-SCHEMA NOT AN ARCHITECTURE

- **Not all the components, need be present in all species of natural or artificial architecture.**

  E.g. subsumption systems, and probably all insects use only the bottom layer (which can include layered components)

- **Each box may itself include a variety of types of components, e.g. within each level there could be**
  - Neural nets
  - Rule-based systems
  - Chemical information processors
  - Other mechanisms more specialised for the level
- **It does NOT specify control flow, or dominance of control: many options left open.**
- **There are many possible routes for flow of information (including control information)**
  - between the "boxes"
  - between components within abox

This is a very general schema, with many special cases.

Subsumption: a special case of CogAff
E.g. in insects?

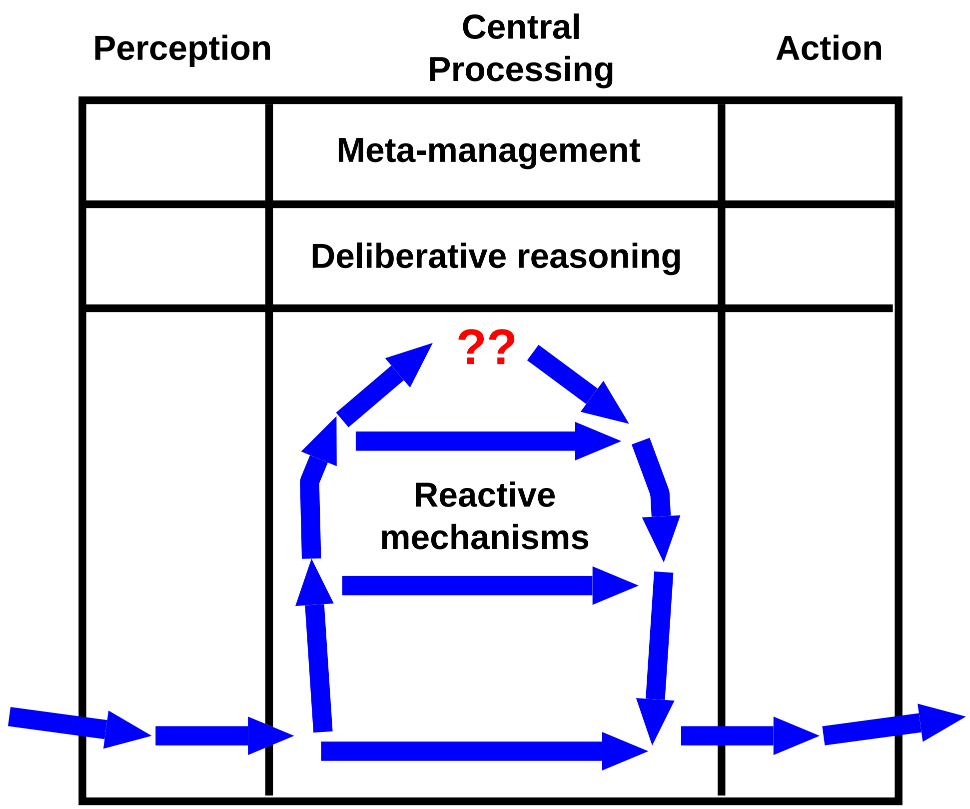
Subsumption architectures

Brooks and others allow several layers of control but only within the reactive sub-space.

(These layers form a sort of dominance hierarchy – unlike the CogAff layers.)

Some proponents of (reactive) behaviour based systems deny that animals (even humans) need or use deliberative mechanisms.

How do they get to overseas conferences?

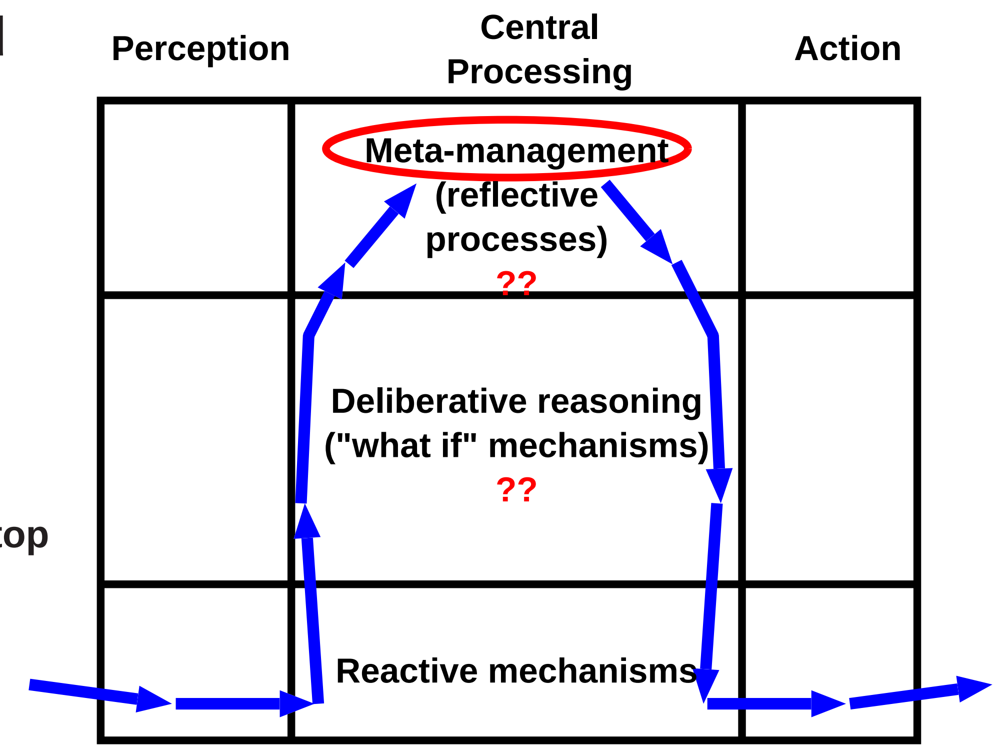


The “Omega” model of information flow

Another popular type of layered architecture is the “Omega” architecture, with Ω-shaped information pathways.

Examples:

- The “contention scheduling” model of (Cooper and Shallice 2000).
- The model proposed by Albus in his 1981 book, where he postulates a “will” at the top



Omega architectures (implicitly) reject

- layered concurrent perceptual systems
- layered concurrent action systems
- layered concurrent central processing using many different forms of representation.

What we are NOT claiming

NOTE:

- **We are NOT saying that observed capabilities of any particular type of organism absolutely require the deliberative or meta-management layer:**

  whatever more sophisticated layers do can always, in principle, be "compiled" into a sufficiently large collection of reactive behaviours along with with a memory mechanism to produce the effects of learning: indeed this seems to be how deliberative processes are often replaced by fast and fluent skilled actions.

- **I.e. purely reactive architectures can, in principle, meet requirements for human-like capabilities, or any other finitely observable behaviours**
- **Whether that's a plausible scientific theory will depend on**
  - **whether the evolutionary history of the species and individual learning by individuals actually provide opportunities for the evolution of the relevant reactive behaviours**
  - **whether the brain mechanisms provide adequate storage capabilities, etc.**

    (The full game tree for chess could not fit into our physical universe: biology has to find ways to defuse combinatorial explosions: one answer is to use deliberative mechanisms.)

Architecture wars?

Subsumption, like the Omega architecture, and many other architectures, uses only a subset of the mechanisms allowed in the CogAff schema.

- **There's nothing wrong with investigating what can be achieved using a limited set of forms of information processing.**
- **But there's no justification for claiming that only one's favoured forms are worthy of scientific investigation, as often happens.**

We should avoid all dogmatism and ideology, and allow people to investigate

- **which subsets are useful for which organisms or machines,**
- **how the various types might be implemented in biological mechanisms**
- **how they might have evolved – i.e. possible evolutionary trajectories constrained by what we know from biology, geology, paleontology, etc.**

That way we'll learn instead of fruitlessly debating restrictive precepts: Ideological battles do not make good science.

Some possible objections

There are many points that need clarification.

E.g.

What is the difference between processes in the perceptual column and processes in the central column?

Possible partial answer:

- Multi-level (multi-window) perception uses dedicated concurrent parsing and interpretation of sensory arrays, e.g. building new data-structures in registration with sensory arrays.
- There can be parallel data-paths from the same low level sensory transducers through different sensory analysers and interpreters.
- Contrast "peephole" perception.

Likewise multi-window action vs peephole action.

(See Albus (1981) and research by Norman and others on skilled actions such as typing, driving, social actions)

These and other problems for the CogAff project, along with conjectures regarding evolutionary and developmental trajectories in design space, are discussed in papers in the CogAff directory and the talks directory

http://www.cs.bham.ac.uk/research/cogaff/

http://www.cs.bham.ac.uk/ axs/misc/talks/

CogAff's perceptual layers: Compare Marr

David Marr's work on vision was outstanding, but even he apparently suffered from ontological blindness: ignoring visual affordances, despite having read Gibson.

Page 36 of Marr's 1982 book *Vision* describes the function of vision as essentially concerned with information about spatial structure and relationships.

> **'quintessential fact of human vision – that it tells about shape and space and spatial arrangement.'**

He admits that

> **'it also tells about the illumination and about the reflectances of the surfaces that make the shapes – their brightnesses and colours and visual textures – and about their motion.'**

But he regards these things as secondary

> **'... they could be hung off a theory in which the main job of vision was to derive a representation of shape'.**

This viewpoint has led to a huge amount of research on vision in AI, but we claim that it ignores important biological functions of vision, for organisms with diverse needs and capabilities.

What's vision really for?

Well, it will not be used in the same way by chickens, chimps, and architectural designers.

QUESTION: for an organism with complex needs and capabilities,

- **is shape information the only or the most important information required from perception?**
- **what about abstract classifications of objects, actions, situations that can play a role in learnt associations?**
  - **apples are good to eat**
  - **chased rabbits often escape down holes**
  - **scorpions have a nasty sting**
  - **a piece of thin wire can be bent to form a hook that's usable for lifting**
  - **a thin piece of grass cannot be used that way**
- **what about Gibson's affordances: the opportunities, obstructions, threats, functional possibilities afforded by things in the environment?**
- **relative to information really required by many organisms (e.g. graspability, bendability, obstruction, support, accessibility, edibility), shape and other physical properties form just a subset of a multi-faceted ontology.**

For more on this see papers on vision at http://www.cs.bham.ac.uk/research/cogaff/

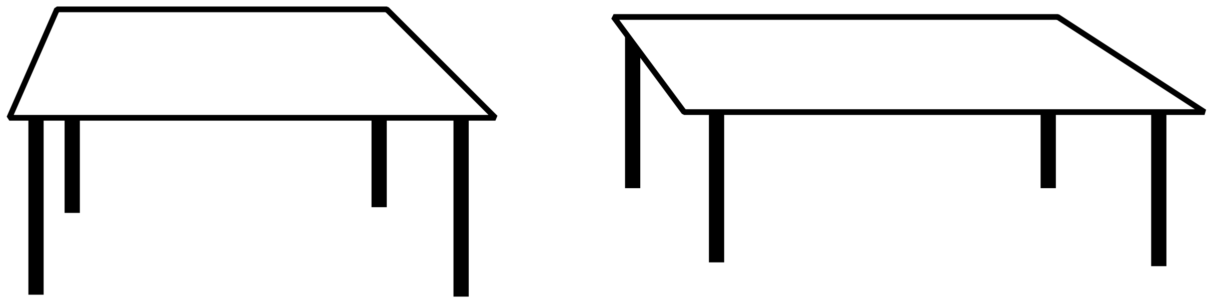
A biological niche includes a collection of affordances

- Two animals in the same physical location have different biological niches (eg. a chicken and a chimp).
- An animal's niche is not simply defined by or deducible from the physical properties of the environment.
- It depends on a complex collection of abstract relationships between what can and cannot occur or be done in the environment and the organism's capabilities and needs.
- Identifying the organism's niche, and therefore what kinds of information the perceptual system may need to provide is a task that benefits from an engineer's stance, informed by an ethologist's:

  What would I have to do in order to design something that can achieve what that sort of organism does. (Compare: Marc Hauser *Wild Minds*, Penguin 2001.)
- In the case of humans we have other sources of information: the many things in textbooks of vision and common experiences.

Removing ontological blinkers in vision research

Studying vision: Look for varieties of affordances (Gibson) as clues to architectural requirements in visual systems.



Marr's question:

How do animals get from 2-D patterns of illumination on retinas to percepts of a 3-D world (seeing useful structural invariants).

A biological (Gibsonian) question:

How does a visual system acquire and represent information about affordances: e.g. functional properties and relationships such as:

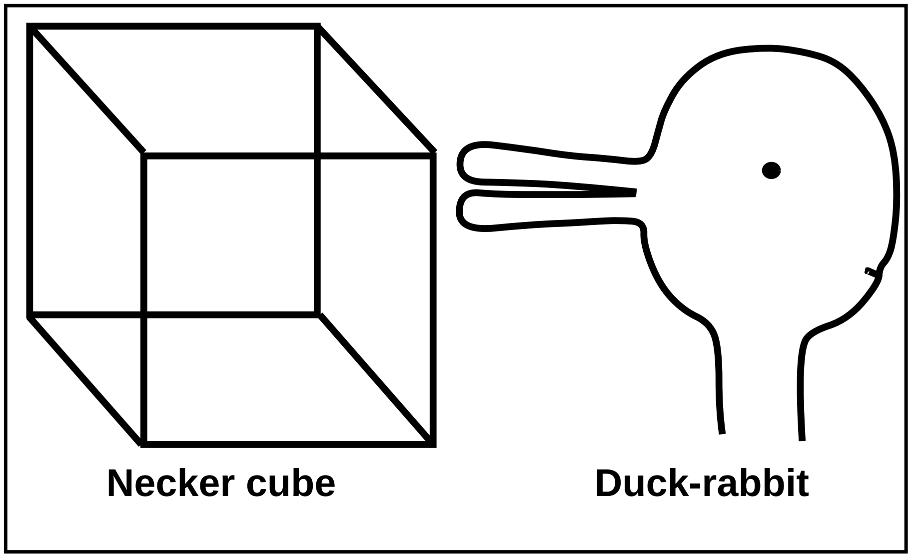
obstruction, passability, support, graspability, bendability, an opportunity to bend a wire gripped only at one point?

NB: We don't yet know any good way to represent most affordances.

Modal logics are unlikely candidates, for most animals.

Another biological question:

How can we see the same 2-D visual input in different ways?



What's the difference between

- **geometric ambiguity (e.g. which side of the cube is nearer?)**
- **non-geometric ambiguity (e.g. which way is the animal looking?)**

Does a visual system use different sorts of ontologies for these? How?

Contrast ontologies required for "online" reactive/continuous visuo-motor processing, e.g. posture control, or controlling grasping vs seeing graspability. Different kinds of brain damage can affect these differentially.

Animals need an information processing ontology

For prey, or for a predator, information about which way something is looking may not just be physical information, but rather information about another information processing entity's access to information.

- **For instance, knowing which way X is looking may be relevant to deciding**
  - **whether X can see you if you try to escape from X**
  - **whether X can see you if you try to get near enough to pounce on X.**
- **This is information about information and information processing: i.e. meta-level information.**
- **The same ontology, and therefore some of the same forms of representation and mechanisms for manipulating those representations may be useful both for internal uses and for external uses.**
- **So the meta-management layer in the CogAff schema is not concerned only with self-observation, self-categorisation in information processing terms.**

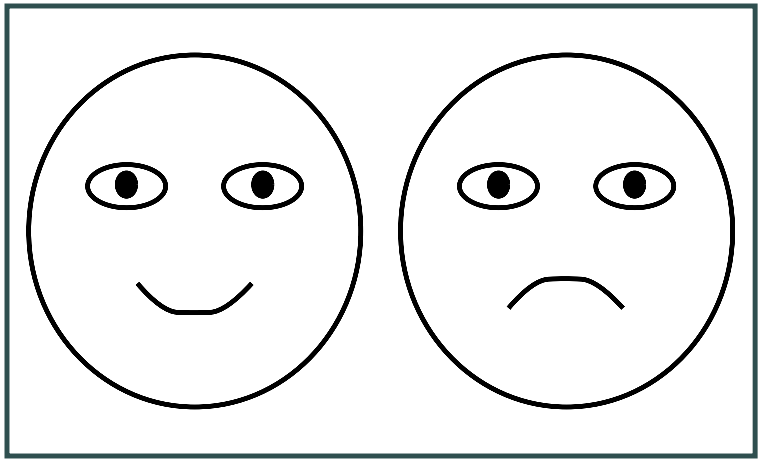
NOTE:
Some of the same behavioural capabilities may be achieved far more simply if they can be compiled into simple reactive pattern detection mechanisms with associated attack or escape responses. But that is feasible only if the variety of usefully detectable mental states of others is small.

Conjecture: dedicated mentalistic visual feature detection

Our conjecture is that

because of the biological importance of meta-level information about other agents, a subset of animals, including humans, and some other mammals, evolved perceptual abilities dedicated to acquiring it.

This required extensions to the visual architecture as well as to central mechanisms able to represent and use information about mental states and processes.



What does it mean to perceive something as having mental states, e.g. having intentions, moods, emotions, knowledge, and performing deliberate actions, ...?

How wide-spread are these various abilities in different organisms?

How diverse are the forms of representation and the ontologies used for this purpose?

Conjecture:

The ability to use concepts in a self-directed meta-management system co-evolved with the ability to use such concepts in characterising others, aided by the evolution of dedicated perceptual mechanisms.

(I.e. evolution solved the ‘other minds’ problem long before human philosophers discovered it.)

NOTE

The cultural diversity of characterisations of mind indicates that in addition to whatever mechanisms, representations and ontologies evolved, for these purposes prior to the development of communicative language was considerably enriched after language became available.

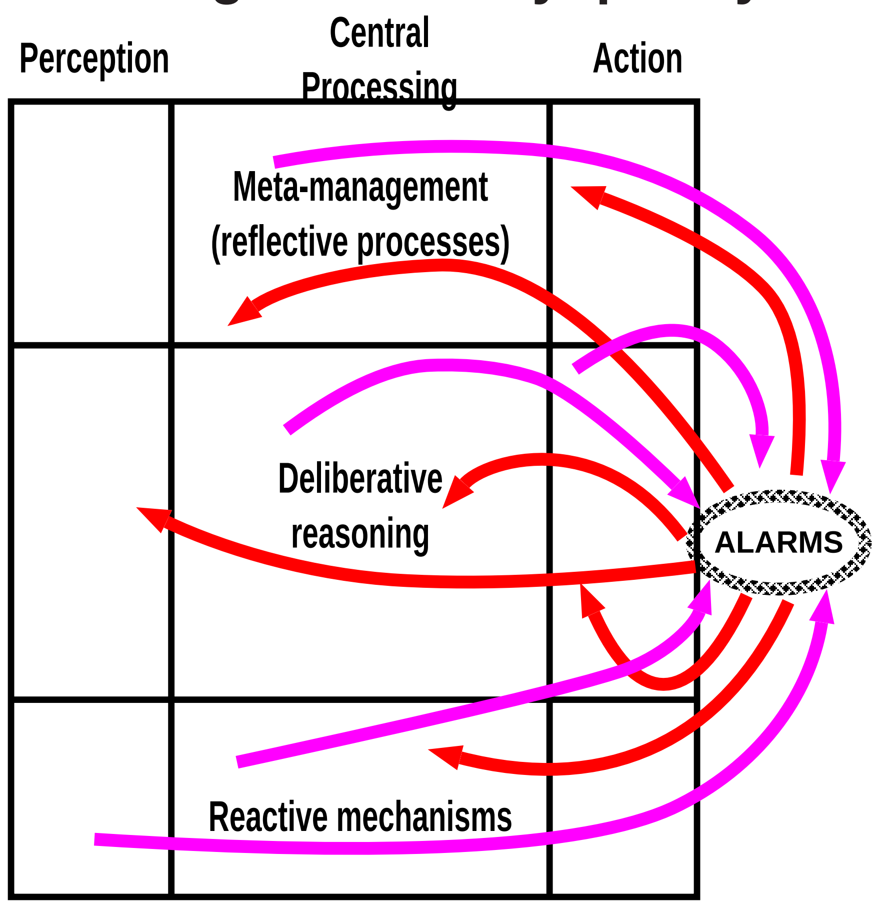
Robots able to think and communicate like humans may need similar mechanisms for absorbing a culture, through linguistic and other means.

The need for “alarm” mechanisms

As processing grows more sophisticated, so it can be come slower, to the point of danger – if some other things move very quickly.

A possible remedy is to use one or more fast, powerful, “(relatively) global alarm systems” able to interrupt and redirect “normal” processing.

An alarm mechanism is just part of the reactive sub-system. Drawing it separately merely serves the pedagogic function of indicating its role. Various kinds of more or less global, more or less rapid, re-direction or re-organisation of processing.



Alarm mechanisms can sometimes support “the five Fs”: feeding, fighting, fleeing, freezing, and reproduction

Many sorts of alarms

- **Alarms allow rapid redirection of the whole system or specific parts of the system required for a particular task (e.g. blinking to protect eyes.)**
- **The alarms can include specialised learnt responses: switching modes of thinking after noticing a potential problem.**
- **E.g. doing mathematics, you suddenly notice a new opportunity and switch direction. Maybe this uses an evolved version of a very old alarm mechanism.**
- **The need for (POSSIBLY RAPID) pattern-directed re-direction by meta-management is often confused with the need for emotions e.g. by Damasio, et. al.**
- **Towards a science of affect:**

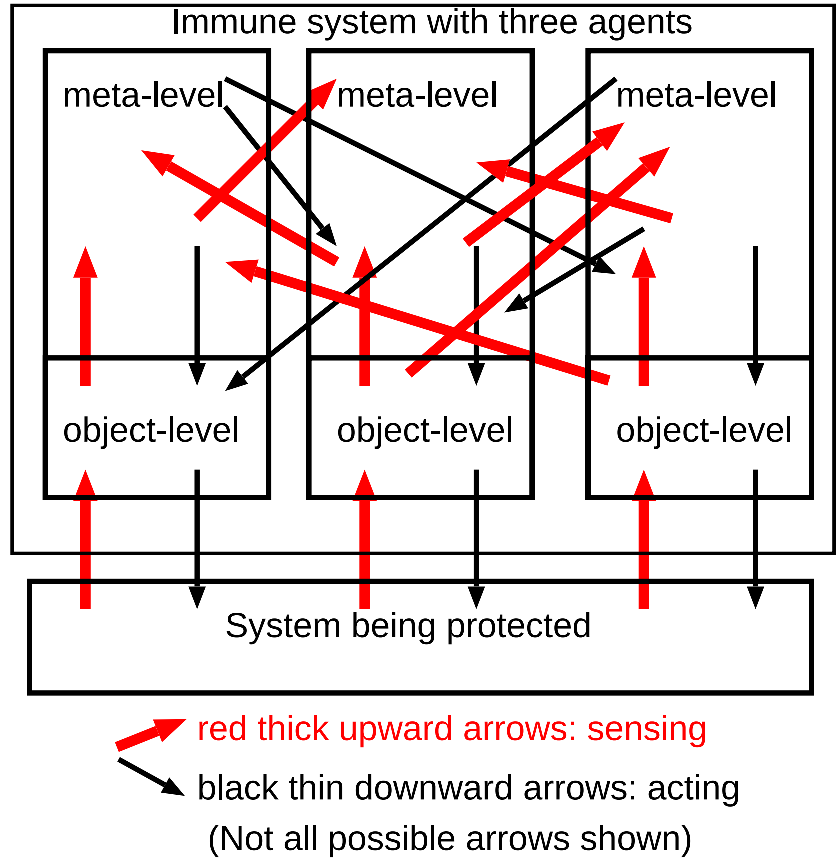
  Not just alarms – many sorts of control mechanisms, evaluators, modulators, mood controllers, personality selectors, etc.

A mutual meta-management system

Catriona Kennedy has been working on extending these ideas in the design of a robust system for detecting and repairing code damaged by hostile intruders.

To avoid the fragility of having only one monitor, Kennedy proposes a collection of them each observing not only the system being protected but also one another's observations, and, if appropriate, taking "corrective" action, e.g. repairing damaged code.

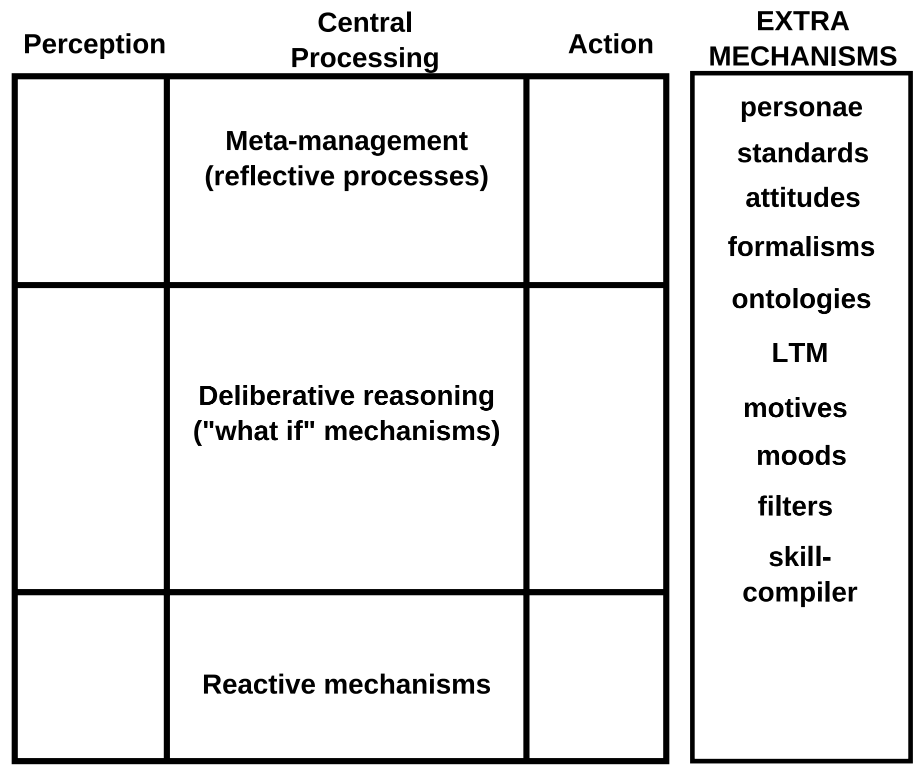
The "object level" components monitor and act on the system being protected. The meta-level components monitor and act on the object- and meta-level components (which may be reactive, deliberative or a mixture).



Some of Kennedy's papers outlining the theoretical ideas and describing a prototype implementation can be found here:

http://www.cs.bham.ac.uk/research/cogaff/0-INDEX00-05.html

Additional components are needed



A partial list is on the right:

Many profound implications regarding varieties of possible architectures, possible types of learning and development, possible effects of brain damage, varieties of affective control states.

Example:

Different sorts of learning can occur within individual sub-systems and also different sorts links between sub-systems can be learnt.

(Not only links shown so far.)

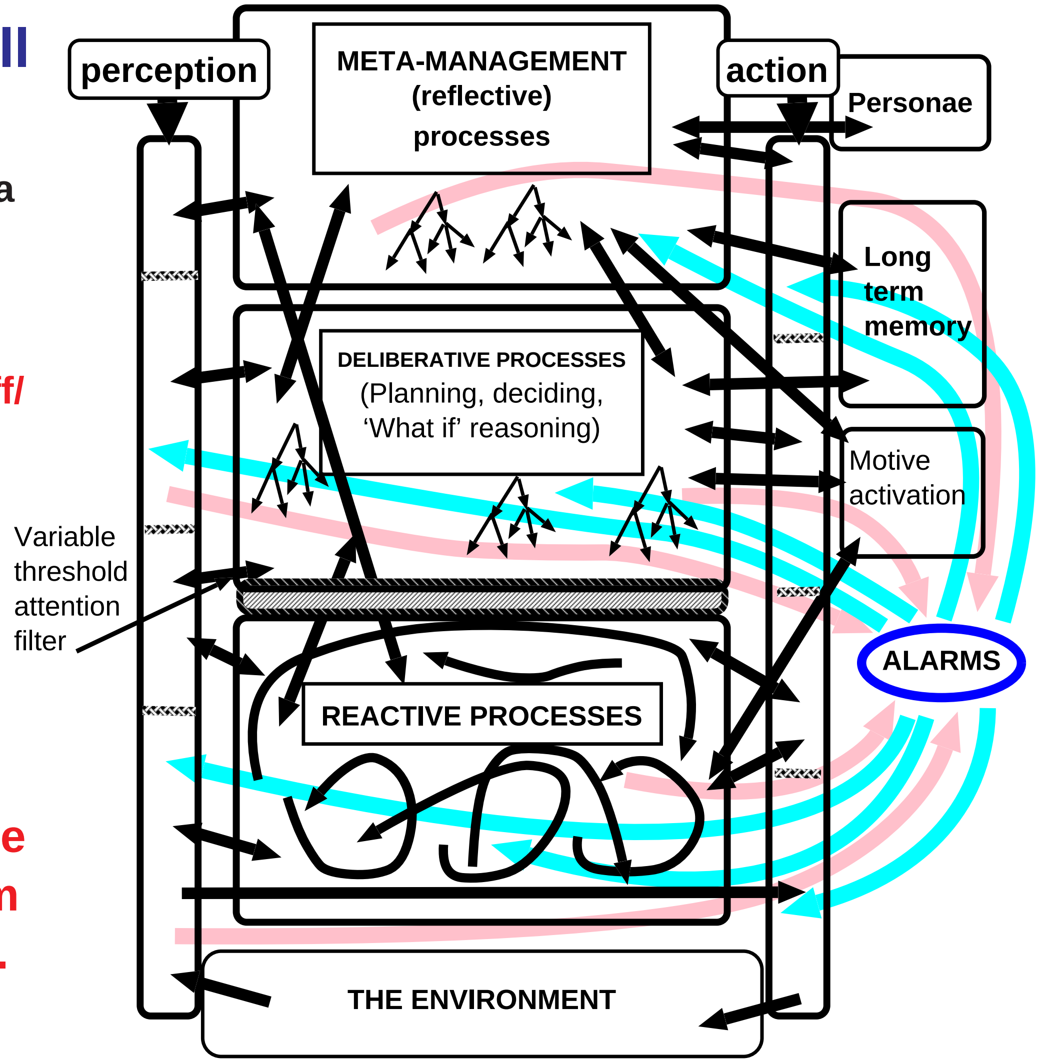
Some forms of development may ‘grow’ new subsystems, e.g. learning to talk? Learning mathematics? Learning to play violin? New forms of self-control?

H-COGAFF: A human-like architecture.

An instance of CogAff using all the components.

The diagram is very impressionistic, not a precise "blue-print".

Described in more detail in papers in the Cogaff directory:
http://www.cs.bham.ac.uk/research/cogaff/



Dynamic attention filter partially protects resource-limited deliberative and meta-management systems from relatively unimportant interrupts e.g. from alarm mechanisms.

But no filter is perfect: hence some kinds of emotions.

Links between CogAff and empirical research 1

How CogAff fits in with what we already know about biological information processing

- Observed diversity of competences among different species and different stages of development in individuals requires a uniform explanatory framework
- Evolutionary trajectories (much still conjectural) and apparent functional layering in brains (e.g. “triune brain”)
- Successful AI modelling of different naturally occurring capabilities using different IP mechanisms
  - reactive behaviours
  - neural nets
  - deliberative planners, problem solvers, language processors
  - some simple examples of meta-management
- Schematic CogAff-based explanations of a vast array of empirical facts about humans, including several sorts of emotions, varieties of attention control, varieties of learning and reasoning
- Schematic CogAff-based explanations of effects of various kinds of brain damage (e.g. blind-sight, different visual pathways, damage to proprioceptive systems used in balance control, compensated by cognitive mechanisms)
- Invention of similar ideas by neuropsychiatrist R. Barkley (book on ADHD)

Links between CogAff and empirical research 2

To be of use, however, the CogAff schema should inform, not just be informed by, empirical research.

There are several ways CogAff can do this:

- **Exploring design space**
- **Provide an organising framework for comparison**
- **Help uncover problems by designing solutions**
- **Assist in asking questions about an information processing system**
  - **Attempting to design something performing a task leads to a much better idea of what the task requirements are: you can't tell just by looking.**
  - **Failures of our design attempts often point to inadequate task analyses** (**e.g. a game playing program defeated by combinatorial explosions indicates a need for a knowledge of more than just the rules of the game: being able to detect heuristic patterns helps – but patterns and reactions alone don't suffice, look-ahead needed too.**)
- **Extending our design ontology: creative ideas from software engineering and AI suggest new things to look for in nature: e.g. new links between boxes, new forms of learning.**
- **Enriching our conceptual frameworks:**
  e.g. CogAff led us to a new taxonomy of affective states, including emotions..

Cogaff, Emotions and Attention

Papers in the Cognition and Affect web directory attempt to show how the CogAff framework provides a more comprehensive framework than other approaches for investigating emotions and other affective states and processes in humans and other animals.

See http://www.cs.bham.ac.uk/research/cogaff/

In particular, we elaborate a first-draft distinction between

- **Primary emotions**
- **Secondary emotions**
- **Tertiary emotions**

related to the three architectural layers in CogAff.

E.g. see the 1996 paper on grief by by Wright, Sloman and Beaudoin.

More empirical connections

- The CogAff framework, and in particular the postulated H-Cogaff architecture for human-like minds, indicates a sub-division between types of higher level functioning normally conflated by psychologists under the label “executive function”.
- It turns out that R.Barkley’s 1997 book *ADHD and the nature of self-control* shows how empirical evidence concerning attentional disorders led him to reach a theory that seems closely related to ours.

  There is still work to be done exploring the links.
- We expect that much of the empirical material on animal information processing in books like M.D. Hauser *Wild Minds: What Animals Really Think* (2001) can be usefully illuminated from the CogAff standpoint. But that is also work still to be done.
- A number of empirical discoveries relating to alternative visual pathways are predicted by the CogAff approach, though we offer a different description – not “what” vs “where”, but “reactive” vs “deliberative”, and many more.

The arguments supporting a particular architecture are never conclusive

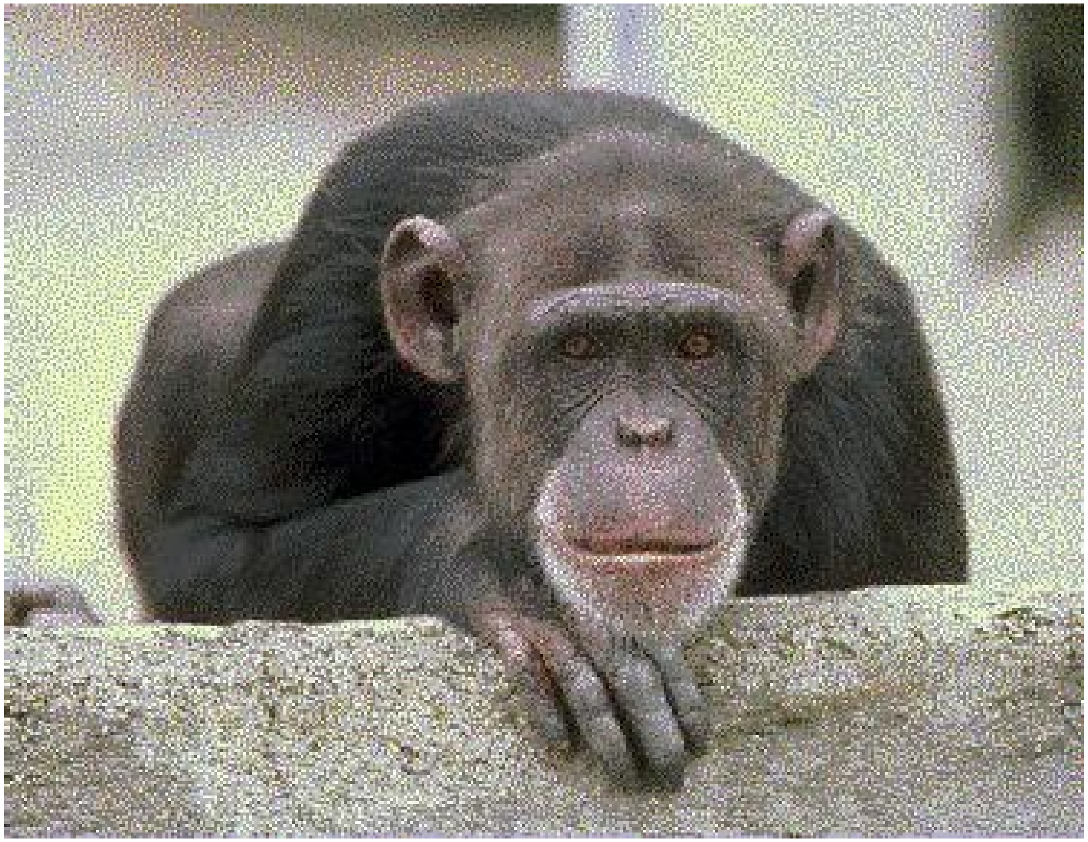
Instead use a mixture of

- Analysis of evolutionary constraints restricting plausible hypotheses about e.g. purely reactive explanations
- Analysis of memory requirements for systems with adequate flexibility, variety of responses.
- Analysis of representational and processing requirements for particular tasks, e.g. picking berries in a thorn bush, building nests, catching large moving prey – though task specifications are conjectural and defeasible.
- Analysis of forms of processing that might in principle be supported by biological mechanisms –
  We don't yet know what sorts of things brains can and cannot do, e.g. what sorts of virtual machines they can and cannot support.
  (We don't yet know what computers can and cannot do either.)
- Demonstrations of practical feasibility in working models

Conclusions

- The ability of organisms to perceive and reason about the world means that we can and should allow for this possibility when thinking about how to design biologically-inspired robots.
- In particular, the abstraction of a virtual information-processing machine, itself understandable on various layers of abstraction, is one which seems to be required to explain many biological behaviours.
- Some biologically-inspired robots should include not just reactive but also deliberative and meta-management layers, not just because humans have them, but for the more substantial reason that merely reactive architectures do not even adequately characterise many simpler organisms.
- It's useful to start now with a schema of possible architectures that encompasses as many forms of biological cognition as possible, rather than limiting ourselves (either out of choice, or due to "ontological blindness") to a design methodology which permits only low-level physical phenomena, or only reactive behaviours, or only currently understood models of neural computation, or ...

What sort of ontology fits this guy?



What sort of architecture makes it possible to be bored?
Or to wonder what someone else is doing?
Or to find something funny?
Or to wonder whether humans have minds?

Courtesy of http://www.zoosociety.com/animals/chimp-pics.htm

Join us

There's obviously a lot more work to be done (300 years? 3000?)

THE PROJECT NEEDS A LOT MORE RESEARCHERS, FROM MANY DIFFERENT DISCIPLINES.

PLEASE JOIN IN.

**We have some nice tools for exploring architectures in simulated environments – all available at our web site.
Good for teaching and research.**

Ask for demos if interested

More Acknowledgements

There is considerable overlap with ideas about architectures in the work of Marvin Minsky, e.g. in *The Society of Mind* and in his draft book *The Emotion Machine* available on his web site: http://web.media.mit.edu/~minsky/

There is also overlap with John McCarthy's papers http://www-formal.stanford.edu/jmc/

Compare Dan Dennett's book *Kinds of minds*

SEI at CMU

Related information and discussion, and a list of definitions of "software architecture" can be found here: http://www.sei.cmu.edu/ata/ata_init.html

Joanna Bryson's recent PhD also attempts a comparative, architecture-based, analysis of design options http://www.ai.mit.edu/people/joanna/publications.html

THERE ARE MANY MORE RELEVANT PUBLICATIONS.

The Birmingham Cognition and Affect Project

PAPERS (mostly postscript and PDF):

http://www.cs.bham.ac.uk/research/cogaff/

(References to other work can be found in papers in this directory)

TOOLS:

http://www.cs.bham.ac.uk/research/poplog/freepoplog.html

(Including the SIM_AGENT toolkit)

SLIDES FOR TALKS (Including IJCAI01 philosophy of AI tutorial with Matthias Scheutz):

http://www.cs.bham.ac.uk/~axs/misc/talks/

Free online book: The Computer Revolution in Philosophy (1978)

http://www.cs.bham.ac.uk/research/cogaff/crp/

(With some recently added notes and comments.)