

**A GRAND CHALLENGE**  
**Architecture of Brain and Mind**  
**Integrating High Level Cognitive Processes**  
**with Brain Mechanisms and Functions**

**Aaron Sloman**

<http://www.cs.bham.ac.uk/~axs>  
School of Computer Science, The University of Birmingham

**With much help from Mike Denham**  
and others involved in Panel D  
at the UKCRC Grand Challenges Workshop  
(Edinburgh November 2002)

See: [http://umbriel.dcs.gla.ac.uk/NeSC/general/esi/events/Grand\\_Challenges/](http://umbriel.dcs.gla.ac.uk/NeSC/general/esi/events/Grand_Challenges/)

These slides will be made available online  
<http://www.cs.bham.ac.uk/research/cogaff/gc/>

# THE PROBLEM

---

How can we understand and model brains and minds of humans and other animals?

Premises:

- **Understanding natural intelligence involves investigation at different levels of abstraction**
  - **Brain:**  
The physical machine, with physical, chemical, physiological and functional levels performing many different types of tasks in parallel including information-processing tasks and others (e.g. supplying energy).
  - **Mind:**  
The “virtual machine” (or collection of interacting virtual machines) performing many different types of information-processing tasks in parallel.
- **An enormously difficult long-term task, requiring cooperation between many disciplines, e.g.**
  - Neuroscience
  - Psychology
  - Computer science
  - Software engineering
  - Artificial Intelligence
  - Linguistics
  - Social sciences
  - Biology
  - Ethology
  - Anthropology
  - Philosophy
  - Mathematics / Logic

**(Organisms are information-processors. Therefore biology studies computers.)**

# MAIN FEATURES OF THE CHALLENGE

---

For the original proposal see

[http://umbriel.dcs.gla.ac.uk/NeSC/general/esi/events/  
Grand\\_Challenges/proposals/ArchitectureOfBrainAndMind.pdf](http://umbriel.dcs.gla.ac.uk/NeSC/general/esi/events/Grand_Challenges/proposals/ArchitectureOfBrainAndMind.pdf)

## PUTTING THE PIECES BACK TOGETHER

**‘Divide and conquer’ is sometimes a useful research strategy but it seems that the study of various aspects of natural and artificial intelligence has become so fragmented that the risk of dead ends is very high: development of ‘solutions’ that will not scale up, or generalise to new contexts, or combine effectively with other parts of an intelligent system.**

Part of the problem is the need to understand the relationships between the virtual and physical machine levels. That requires a mixture of science, engineering, and philosophy.

That topic is discussed in

<http://www.cs.bham.ac.uk/research/cogaff/talks/#talk23>

# A deep cultural and educational problem

**Most scientists and engineers are not trained to think about problems from multiple viewpoints, neither are they knowledgeable about the facts and constraints investigated in different disciplines.**

**(There's too much pressure to publish and get grants: no time to explore and think.)**

- Most people in AI know little or nothing about psychology, linguistics, neuroscience or philosophy.
- Most people who study vision never think about how language or problem-solving works.
- Most who study language never think about how vision and planning work.
- Many who study planning or problem solving don't think about how language or vision work.
- Psychologists are not normally trained to design anything that works, so most of them lack understanding of mechanisms, representations, architectures, etc.
- People who study human capabilities often restrict their investigations to adult humans and do not think about
  - Other animals
  - Human infants
  - Brain damaged humans

# Architecture of Brain and Mind

---

## Putting the pieces back together:

- We aim to understand and model brains/minds as **integrated** systems functioning at different levels of abstraction, including
  - **Physiological properties of brain mechanisms, e.g. cortical microcircuits.**
  - **Neural information processing functions**  
(possibly requiring a new ontology of functions)
  - **Higher level cognitive and affective functions of many sorts**
  - **Behaviours of complete agents (including social behaviours).**
- This requires us to understand how the different levels, and the different components at each level, combine to form an integrated functioning system
  - **some levels implementing others,**
  - **some sub-systems cooperating with or competing with others**
- However, we aim to abstract **principles** of operation rather than always merely trying to mimic biology in great detail.
- That includes finding good characterisations of **requirements** for architectures, mechanisms, formalisms, ...

**A good way to study integration of multiple kinds of functionality is to attempt to design a biologically-inspired robot with as many different capabilities as possible.**

# Examples of sub-tasks: Brain mechanism

---

Because the primary function of the nervous system is to gather, represent, interpret, use, store and transmit **information**, neuroscience is inherently a computational discipline, and, as for other forms of computation, different levels and types of functioning will need to be investigated and modelled.

The project will progressively deepen the following activities:

- Review what is known about brain architecture in humans and other animals.
- Review important research results on anatomy and functioning of brain mechanisms, including relations between cortical and sub-cortical mechanisms, discoveries concerning neural micro-circuits and the ability of neural connectivity to vary very rapidly, supporting new forms of computation.
- Attempt to develop an understanding of the fundamental principles involved and the mechanisms of self-organisation and adaption in brains.
- Build working models to demonstrate what can be done by such mechanisms within such architectures.
- Demonstrate how large scale implementations can support some of the higher level functional requirements.
- Implement working models that approximate to the functionality of the biological mechanisms at levels of abstraction that enables us to test their use in working robots combining different capabilities (e.g. perception, motivation, planning, learning, controlling actions, introspection, etc.)

# Examples of sub-tasks: virtual machines

---

## The project will:

- Review what is known about natural intelligent systems, and organise a coherent subset in a manner that can guide design work.
- Identify important types of virtual machines found in natural intelligent systems (including both symbolic and sub-symbolic machines).
- Identify important functional decompositions within virtual machines
  - **This may be harder than it seems:**  
e.g. it is not at all easy to identify all the functions of vision, of language, of learning...
- Investigate the types of mechanisms, formalisms, ontologies relevant to various kinds of sub-tasks, and their trade-offs. (E.g. speed/flexibility, precision/generalizability.)
- Explore generic mechanisms and architectures for self-organisation, adaptability, creativity in addition to powerful domain specific capabilities.
- Attempt implementation-neutral specifications so that we can explore alternative implementations, then compare brain-inspired implementations with more conventional computational implementations.
- Formalise ideas about requirements, architectures, representations, mechanisms, both informally and in mathematical form.
- Develop working models, including both simulations and physical robots.
- In particular, work towards building one or more robots combining many different kinds of capabilities in a coherent fashion.

**Specifying requirements for such robots will be a major part of the project.**

# What sort of robot? There are many possibilities.

**One proposal discussed in some detail, with pros and cons:**

**Aim towards: both simulated and physical robots with many of the generic capabilities of a typical 3-5 year old child.**

**Why not model an infant?**

**or an adult, e.g. a clerical assistant?**

**Because**

- **Infants are too inscrutable: they are building an architecture, but we can't tell what sort, or how they are building it, except perhaps by understanding later stages and working backwards.**

**(Compare COG (MIT) and Lucy (Cyberlife) projects.)**

- **Adults are too much a product of culture and individual learning building on generic capabilities of a child in unknown, often idiosyncratic, ways.**

**(Compare discussions in the DARPA Cognitive Systems project.)**

- **A good working model of generic child-like intelligence, including early forms of self-understanding, could lead to important new explanations of both earlier and later stages of development, and could be the basis of many different sorts of demanding practical applications.**

# Counter-proposals for target robots

---

- Some discussants thought that mention of anything like a robot child would produce strong negative public reactions (e.g. because of the AI movie?)
- Others thought the opposite – there has always been a deep interest in the possibility of human-like machines (since the Golem story and earlier)!  
(The problem with the movie AI seems to have been execution rather than conception.)
- It is very easy to come up with proposals for specific robot projects: and arguments against them (too difficult, too easy, too boring, not principled enough, etc.)
- At present there are many robot projects, including international robot competitions, though most are specified by some practical target (e.g. winning at soccer, or helping search and rescue teams) rather than human-inspired design principles.
- So different research groups will explore different types – they should be encouraged to combine increasingly varied types of capabilities, and the different groups of researchers should communicate and compare goals, theories, methods and results.
- Deciding whether the robot is to be child-like, and if child-like which age range is chosen, is less important than trying to be precise about interesting and demanding sets of human and animal capabilities to be combined in a robot.

## Towards an integrated robot: How can a machine:

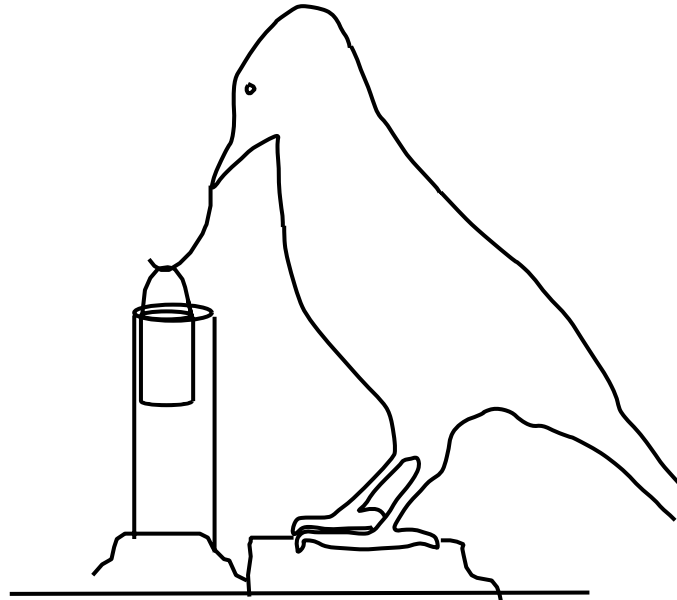
- See structured but changing (rigid and flexible, inanimate and animate) objects.
- Build plasticine, lego, tinker toy and meccano models – and want to do so.
- Dress and undress dolls, or itself.
- Learn its way round a room, a house, a garden, a village.
- Climb up a ladder to fetch a fragile object off a shelf.
- Learn to count, then later use that ability in many tasks.
- Learn to think about numbers, then later about infinite sets.
- Learn to read and write, learn to read music and sing or play it.
- Communicate with other intelligent systems
  - Asking questions and giving answers
  - Requesting or providing explanations, and using them
  - Reporting what it did last week.
- Explain how a pendulum clock works.
- Explain what a clock is for.
- Tie shoelaces.
- Reason about geometrical relations.
- Feel fear, pity, shame, pride, jealousy, ... enjoy dancing, music, painting, poetry.
- Care about how another feels.
- Become more skilled at catching a ball, dressing and undressing itself, and judging when it is safe to cross the road.
- Help a less able person with everyday tasks?
  - E.g. a younger child, or someone physically infirm, or blind, or deaf, or ...

# **We could also study other animals.**

---

## **Betty Crow: Cognitive Agent and Hook-maker**

Two crows, Betty and Abel, learnt to use bent wire to fish a bucket of food out of the vertical tube (as in the picture). Then Abel flew off with the hook.



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

To find more, give google: **betty crow hook**

- Betty tried using a straight piece of wire for a while, and failed.
- She then pushed one end of the wire into the tape holding the tube and moved the other round with her beak, making a hook, which she used to lift the bucket.
- She did this 9 times out of 10. **Reported in Nature and shown on BBC TV (August 2002).**

## **HOW CAN A ROBOT REPLICATE BETTY'S MENTAL PROCESSES?**

# Vision and affordances

---

Vision and understanding of space and motion are crucial.

Vision is not just about

- Object recognition
- Perception of geometrical and physical structure and motion
- Building spatial 'maps' for route-planning

There's something deeper, not yet properly characterised, which can be called **perception of affordances**.

- Affordances are not "objective" properties intrinsic to physical configurations.
- They are **relational** features dependent on the perceiver's
  - Common or likely goals and needs
  - Capabilities for action (physical design + software)
  - Constraints and preferences (avoid stress, injury)
- Perceiving an affordance involves seeing **what does not yet exist but might**.

How should affordances be perceived, represented, used, explained to others?

**Affordances in a complex scene can be construed as**

- **sets of sets** of counterfactual conditionals,
- **spatially indexed**: different sets attached to different parts of objects.

Different representations and mechanisms handle affordances in different architectural layers - e.g. skilled behaviour is mostly reactively controlled.

**What affordances did Betty need to see?**

**What sort of robot child could see them, and use them to solve similar problems?**

# Impressive robots made by Honda and Sony



THE STATE OF THE ART IN 2002



(c) Sony Corp.

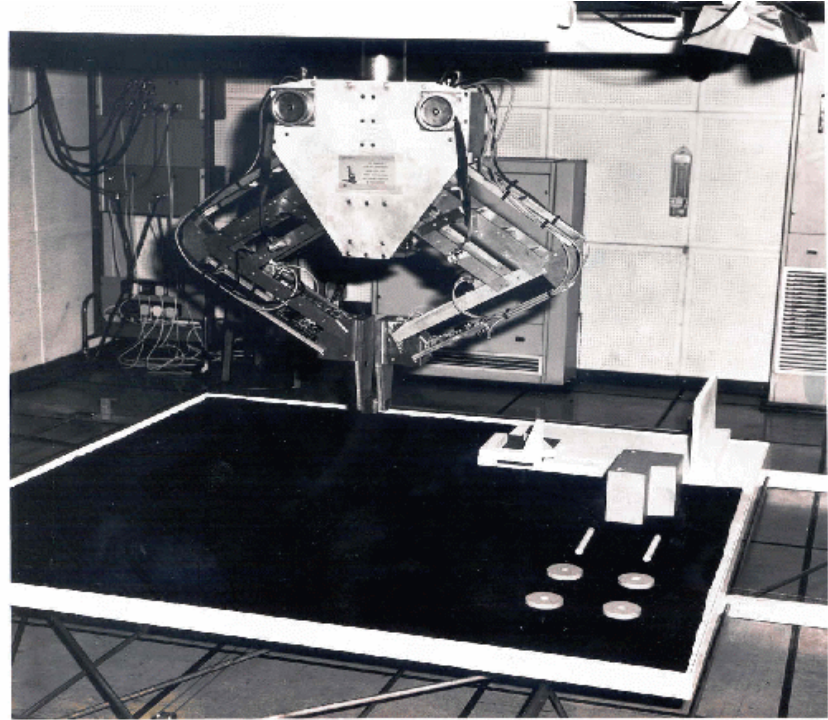
<http://www.aibo.com/>

<http://world.honda.com/news/2002/c021205.html>

In both cases the engineering is very impressive. But present day robots look incompetent if given a task that is even slightly different from what they have been programmed to do – unlike a child or chimp or squirrel. Mostly they have purely reactive behaviours, lacking the deliberative ability to think or wonder ‘what would happen if...’. They also have very little self-knowledge or self-understanding, e.g. about their limitations.

# Compare Freddy the 1973 Edinburgh Robot

Some people might say that apart from the wondrous advances in mechanical and electronic engineering there has been little increase in sophistication since the time of Freddy, the 'scottish' Robot, built in Edinburgh around 1972-3. Freddy could assemble a toy car from the components (body, two axles, two wheels) shown. They did not need to be laid out neatly as in the picture. However, Freddy had many limitations arising out of the technology of the time.



E.g. Freddy could not simultaneously see and act.

There is more information on Freddy here

<http://www.ipab.informatics.ed.ac.uk/IAS.html>

<http://www-robotics.cs.umass.edu/pop/VAP.html>

**In order to understand the limitations of robots built so far, we need to understand much better exactly what animals do: we have to look at animals with the eyes of (software) engineers.**

# How to make progress

---

There are various practical and theoretical prerequisites for progress

- We'll need new standards and powerful new tools for integrating diverse hardware and software resources in a complete intelligent system, in particular tools supporting rapid prototyping, essential for exploratory research on very hard problems.

Often it is only by doing preliminary implementations and incrementally interacting with and extending them in different ways that you can understand the problems.

**An outline Basic Technology proposal for a rapid prototyping integration platform:**

<http://www.cs.bham.ac.uk/~axs/basictech/>

(NB: open source is essential for effective international collaborative exploration and development.)

- We shall need to collect many examples of tasks involving a wide variety of capabilities, including perception, motor-control, planning, reasoning, learning, communicating.
- Some example (first draft, partial) specifications are here  
<http://www.cs.bham.ac.uk/research/cogaff/gc/>  
<http://www.cs.bham.ac.uk/research/cogaff/manip/>
- We shall have to design several hundred (or several thousand) scenarios (e.g. in the form of small film-scripts), defining various types of performances to aim for.  
**Samples:** <http://www.cs.bham.ac.uk/research/cogaff/gc/targets.html>

## **Example scenario fragment: one of many**

---

1. Robbie wants to get box from high shelf. Ladder is in place. Robbie climbs ladder, grasps box then climbs down.
2. As for 1 except that Robbie climbs ladder, finds the box is too far to one side to reach, so climbs down, moves the ladder sideways then as 1.
3. The ladder is lying on the floor at the far end of the room. Robbie drags it across the room lifts it against the wall, then as 1.
4. As for 1, except that if asked “Why are you climbing the ladder?” Robbie answers: something like “To get the box” (not “To increase my altitude”).
5. As 2 and 3, but when moving the ladder Robbie can be asked: “Why are you moving the ladder?” And gives a sensible reply. (What isn’t sensible?)
6. If asked: would it be safe to climb if the foot of the ladder is right up against the wall, Robbie answers No, and if asked why not, gives a sensible answer (such as?)
7. Robbie can answer questions about **‘what would happen if’**:
  - Q: What would happen if foot of ladder were further from the wall?
  - A: Top of ladder will be lower, and may be too low
  - Q: Why?
  - Etc., (Robbie may not always be able to give sensible answers. How many humans could?)What is required for Robbie to understand “ladder will be unsafe to climb”?

There are many other scenarios, involving, games, telling or understanding stories, exploring, fighting, helping, etc.

# We need to understand the space of architectures

A sort of **generative grammar** for a class of architectures for integrated agents, perceiving and acting on: a complex and changing environment.

Different architectures include mechanisms in different subsets of the boxes, and different possible information links, different possible control relationships.

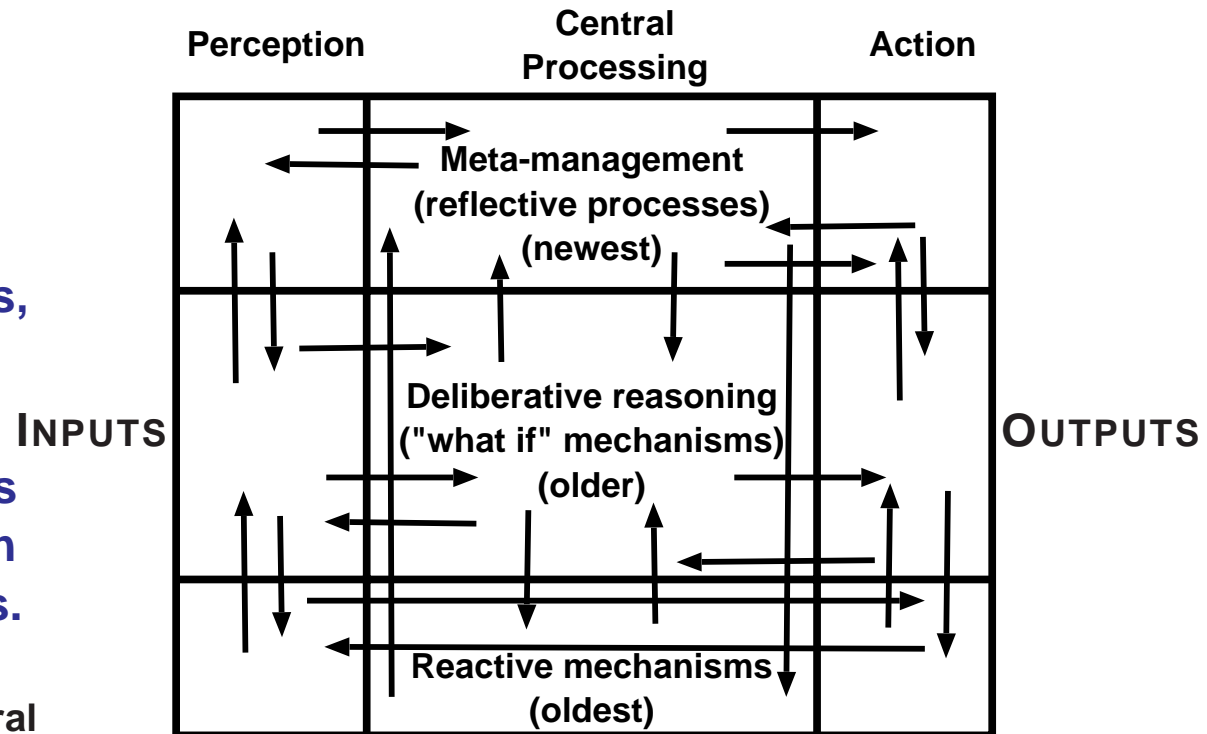
There are also differences in forms of representation and types of mechanisms.

There are many particular architectures that fit this general framework. E.g. it seems that insects have architectures containing only mechanisms in the reactive layer.

The **Cognition and Affect** papers and presentations explain this in more detail.

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

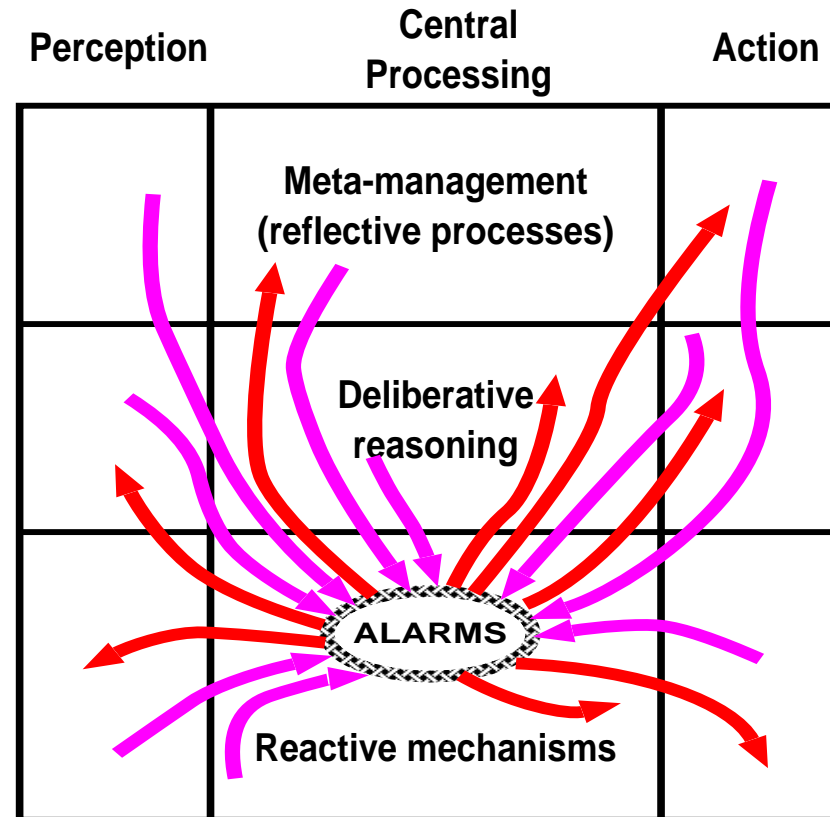
<http://www.cs.bham.ac.uk/research/cogaff/talks/>



**Not all possible information flows are shown!**

# With alarm mechanisms

- 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.



**The architectural basis  
for several types of emotions**

# A hypothetical Human-like architecture: H-CogAff (See <http://www.cs.bham.ac.uk/research/cogaff/>)

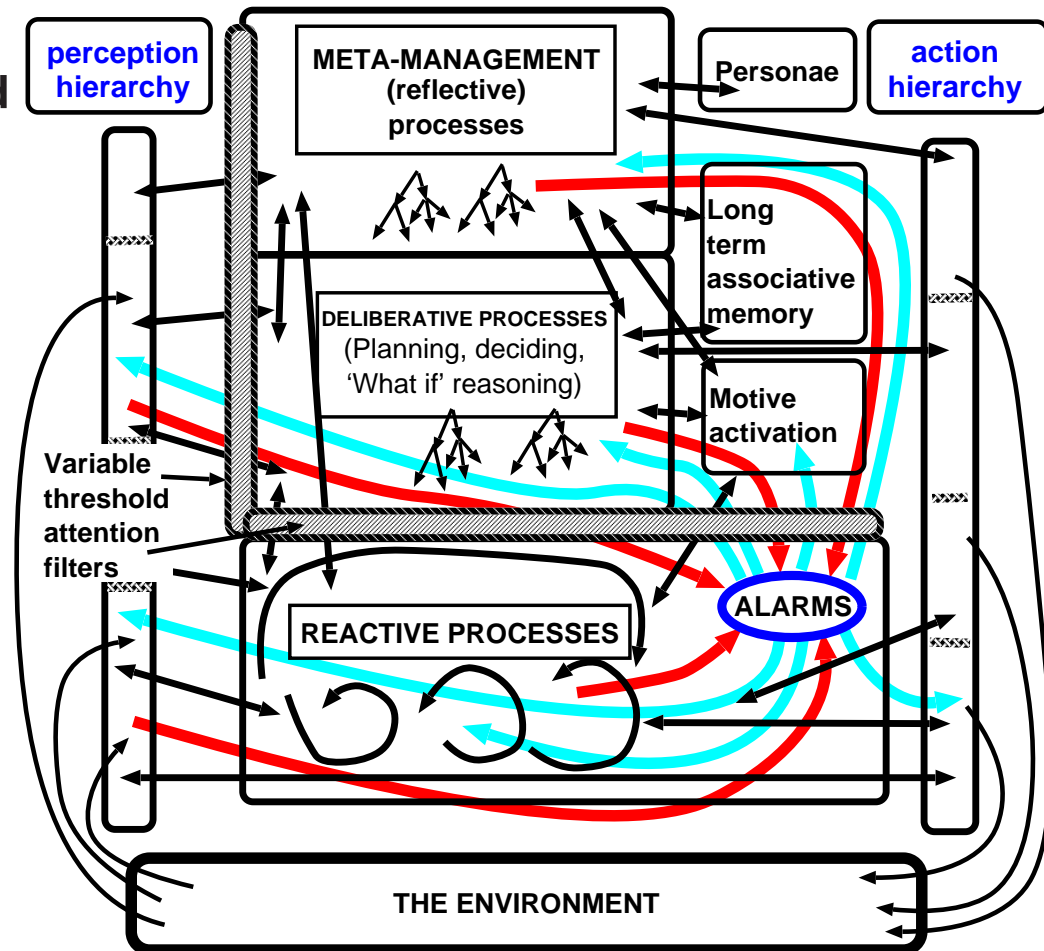
This partly overlaps with Minsky's *Emotion machine* architecture.

This is an instance (or specialised sub-class) of the architectures covered by CogAff schema.

Where could it come from?

Various trajectories:

- evolutionary,
- developmental,
  - **Altricial species build their architectures while interacting with the environment?**
- adaptive,
- skills developed through repetition (how?)
- social learning, including changing personae...



**Most computing power is in the evolutionarily old reactive mechanisms. Why?**

# Related things happening – some examples:

**EU Cognitive Systems programme** (There will soon be a call for proposals).

**DARPA Cognitive systems programme** (Online workshop reports)

November 2002, Virginia:

<http://www.dsic-web.net/meetings/oy8guwod/presentations.html>

<http://www.dsic-web.net/meetings/oy8guwod/papers.html>

March 2003, Stanford workshop on cognitive architectures:

[href="http://www.isle.org/symposia/cogarch/](http://www.isle.org/symposia/cogarch/)

**Robocup and Robocup Rescue**

<http://www.robocup.org>

<http://robomec.cs.kobe-u.ac.jp/robocup-rescue/>

**NurseBot (Martha Pollack and others)**

<http://www.eecs.umich.edu/pollackm/nursebot/index.htm>

<http://www-2.cs.cmu.edu/nurse-bot/>

**Minsky and Singh at MIT**

<http://www.media.mit.edu/~minsky/>

<http://taffy.media.mit.edu/Push.PhD.Proposal.pdf>

**IBM Architecture project**

March 2002 Workshop reported in their systems journal

<http://www.research.ibm.com/journal/sj41-3.html>

**Many toolkits being developed** (often too restricted, however).

**Many conferences and workshops, e.g.**

**ASSC7 in Memphis May/June**

**EU workshops on consciousness, Birmingham Sept, Torino Oct.**

**We must not be parochial, or we risk being left behind.**

## Two Warnings

---

**The end is nowhere near being in sight,  
despite many extravagant claims.**

**But we can make significant progress,  
provided that we select sub-goals with great care.**

**Those who are ignorant of philosophy are  
doomed to reinvent it, usually badly.**

**E.g. “symbol grounding” theory is just a reincarnation of  
concept empiricism refuted by Immanuel Kant long ago.**

**See <http://www.cs.bham.ac.uk/research/cogaff/talks/#talk14>**

# THANKS

---

The ideas presented here owe a lot to interactions  
with Marvin Minsky and Push Singh at MIT,  
e.g. see Minsky's draft chapters for *The Emotion Machine*

<http://www.media.mit.edu/~minsky/>

and some of the work of John McCarthy,  
e.g. his paper on the well-designed child:

<http://www-formal.stanford.edu/jmc/child1.html>

and many other people...

**Thanks also to the developers of Linux**  
**and other free, portable, reliable,**  
**software systems,**

**e.g. Latex, Tgif, xdvi, ghostscript, Poplog/Pop-11, etc.**