The Cognition and Affect (CogAff) Project

School of Computer Science,
The University of Birmingham

Towards A Theory of What Minds Are and How They Work

Contact: Aaron Sloman

Email: A.Sloman@cs.bham.ac.uk
WEB: http://www.cs.bham.ac.uk/~axs/

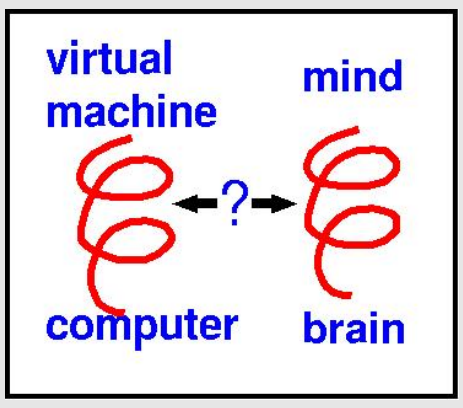
CoSY Project: http://www.cs.bham.ac.uk/research/projects/cosy/
TALKS: http://www.cs.bham.ac.uk/research/cogaff/talks/

Some background to our work

- **Human minds combine many functions, and many parallel operations, including perception, learning, generating motives, making decisions, changing affective states (e.g. attitudes, emotions and moods), controlling actions, communicating through language, ...**

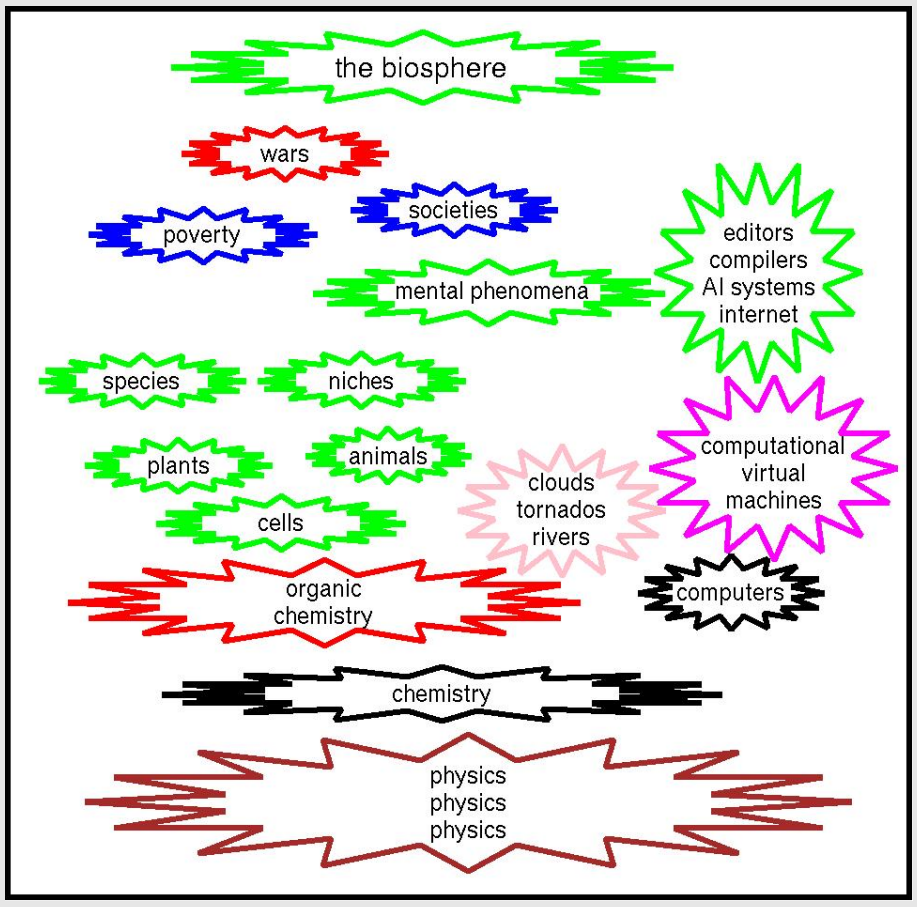
- **These processes all acquire, use, manipulate and generate information (including control information) – some of it about the environment (including other animate entities) and some about internal states and processes.**

- **Computational virtual machines (running operating systems, email systems, airline booking systems, etc.) are implemented in man-made machines. Likewise, we can view minds as implemented in brains.**

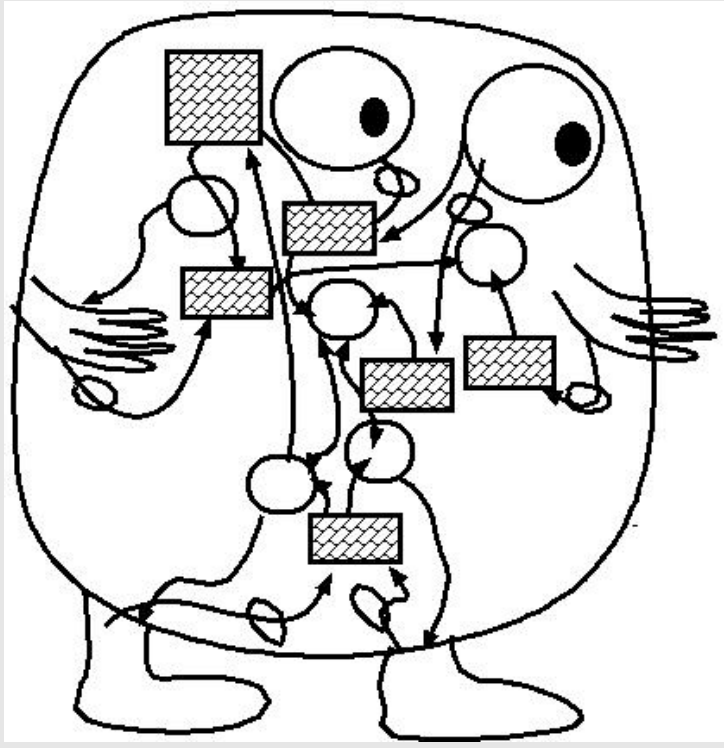


In both cases there are several layers of virtual machinery: including quantum physical processes, molecular level (e.g. chemical) process, physiological or electronic processes, information manipulation processes, and functional processes based on and to some extent controlling all the others.

- **Words of ordinary language referring to mental phenomena, such as emotion, desire, belief, consciousness are systematically ambiguous. But we can 'rationally reconstruct' the concepts on the basis of a design-based theory: a theory of the architecture of brain and mind at various levels of abstraction – some implemented in others.**



Reality has many levels: ***science investigates them all and their interactions, though most scientists peek only at a small subset.***



Any flat model of how humans or any other animals work (e.g. a 'wiring diagram' of the brain) will be incomplete: we need to understand levels of abstraction and implementation.

Levels of implementation

Levels in computing systems	Levels in minds/brains
Networks: internet addresses, email, web, security, ...	Ontologies, languages, beliefs,desires, intentions, skills, preferences, values, attitudes, emotions, moods, ...
Packages: uses, user interface, bugs, ...	Functional roles: kinds of information, connections, control relations,learning, inputs, outputs (of components)
Languages: data-types, procedures, compilers, interpreters, ...	Brain organisation: major functional divisions, ...
Operating system: scheduler, memory management, file-system, ...	Physiology: neurons, blood-vessels, neuro-transmitters, pathways, ...
Computer: instructions, data, devices,...	
Electronics: circuits, signals, timing,..	
Physics: atomic, molecular, materials,...	Physics: atomic, molecular, materials,...

Virtual machines are shown at the top and at intermediate layers, and physical machines at the bottom. This is not meant to be an accurate model, merely an indication of the scope of the concept of layered virtual machines in man-made and natural systems.

There is no simple relation between the two columns.

Requirements for architectural theories

- **Natural architectures evolved to fit many different biological niches. We need, but don't yet have, an agreed conceptual framework for describing both architectures and requirements/niches.**
- **We can move towards an agreed ontology for architectural designs by making some high level distinctions, e.g. between**
  1. ***sensory/perceptual processes constantly changing to represent the environment (including internal states)***
  2. ***motor/action/effector processes constantly changing the environment and perhaps some internal states***
  3. ***central, more slowly changing, processes***

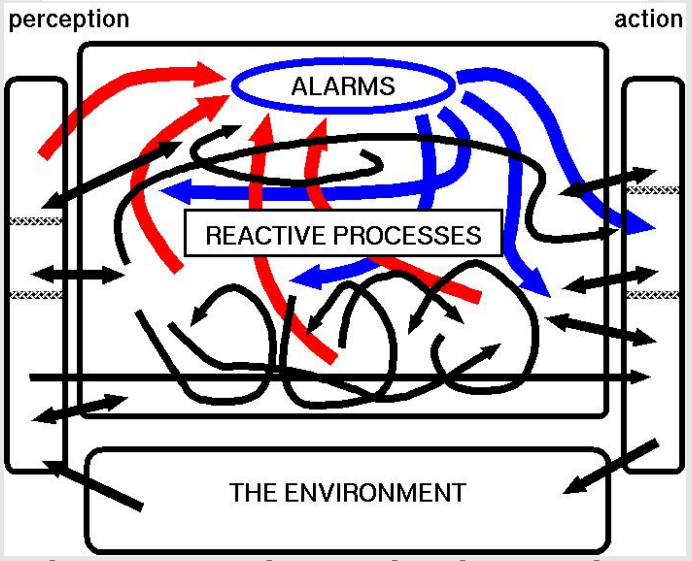
  or between

  1. ***Evolutionarily very old reactive processes, constantly driven by what is sensed internally and externally***
  2. ***Newer deliberative processes able to represent what does not exist but might, e.g. future actions, unseen situations, past causes.***
  3. ***Specialised meta-management/reflective processes capable of describing information-processes states and processes in oneself (and therefore also others).***

The CogAff schema shown, above right, summarises this space of possible types of architectural components. An insect-like special case is on the right (purely reactive).

The CogAff Grid

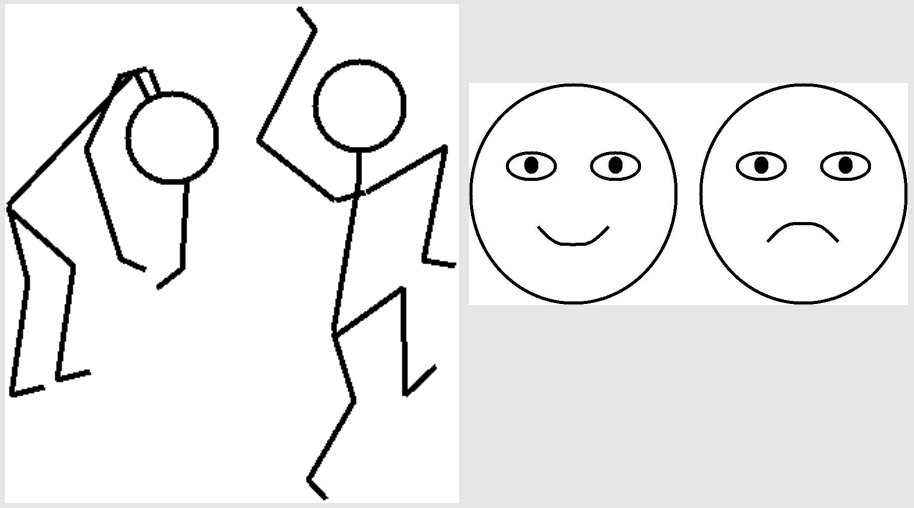
Perception	Central Processing	Action
	Meta-management (reflective processes) (newest)	
	Deliberative reasoning ("what if" mechanisms) (older)	
	Reactive mechanisms (oldest)	



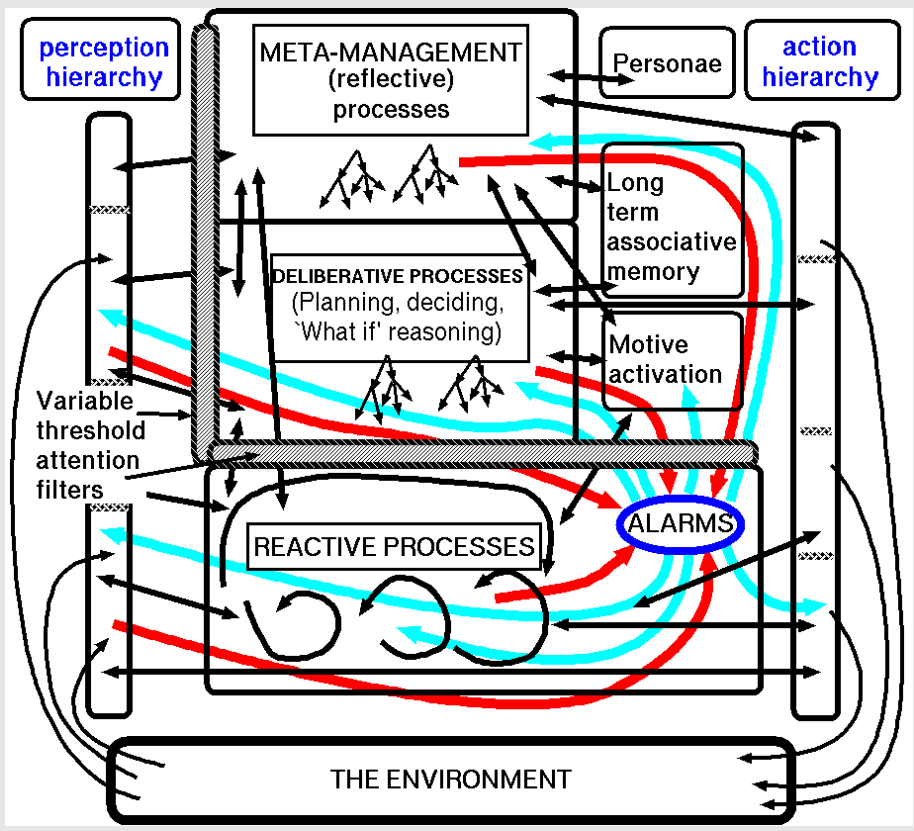
Alarms can trigger simple emotions (even in insects?)

Towards a human-like architecture

- By superimposing the two three-way divisions listed previously we get a 3x3 grid of nine possible types of mechanisms, states and processes in an architecture, which may or may not be linked to other such processes. We call this generic schema CogAff, shown on previous page.
- An example of an instance of CogAff might be an insect-like organism with only reactive mechanisms, including a fast-acting "alarm" sub-mechanism, also shown on previous page.
- Other examples of the schema will be architectures that include specific components from the various categories, with specific information flow (including control flow) connections between them.
  E.g. microbes and insects all seem to use only components at the reactive level, though interacting reactive components can produce very complex behaviours (e.g. termites building 'cathedrals').
- Fully deliberative mechanisms with the ability to construct structural descriptions of both perceived and hypothetical situations, including sequences of possible future actions, evolved much later, and require biologically expensive mechanisms, like discretizing perceptual mechanisms, a large extendable store of associations, short-term memories for constructing temporary structures and comparing, them, etc. This also supported high level 'chunking' of actions to simplify plan structures.
- Even more sophisticated animals evolved reflective abilities to represent information-processing mechanisms, states and processes, whether in other individuals (e.g. predators, prey, conspecifics) or in themselves. This enabled evolution of perceptual and action mechanisms making use of this information-processing ontology, as indicated in the figure top right. Where the perception and action are internally directed (e.g. towards deliberative and other processes) we call this meta-management (following Luc Beaudoin).
- The HCogaff architecture sketched on the right, is a special case of the CogAff schema which is conjectured as a (crude, first draft) model of a typical human information-processing architecture (perhaps after four or five years of development from infancy, and later), combining components in all the CogAff boxes all acting concurrently.



Multi-layer perception and action: some high level perceptual and action mechanisms, linked to central mechanisms using an ontology of information states, may be able to detect affective states in others (e.g. sadness, and joy), and to produce affectively expressive behaviours.
(Instead of merely perceiving 2-D and 3-D structure and motion, recognizing types of physical objects, identifying individual entities, and assembling low-level actions.)



The HCogaff (Human-like) architecture, an instance of the CogAff schema, supporting many varieties of motivation, learning, perception, deliberation, emotion, mood, etc.
The alarm mechanisms are reactive components which take information from various parts of the system and use fast (and therefore possibly stupid) pattern recognition to detect a need for rapid global re-organisation: a type of emotion.
Similar mechanisms with long term global control functions (not shown) produce moods.

Architecture-based analysis accounts for a rich collection of affective states and processes arising from interactions of various information-processing subsystems, including motivation, emotions, moods, and personality. Emotion categories depend on: time scales, where inputs come from, where disruption or modulation occurs, whether episodic or dispositional, whether detected by meta-management or not, etc.

Our online papers elaborate on all this: http://www.cs.bham.ac.uk/research/cogaff/
Our freely available SimAgent Toolkit helps us explore different architectures.
A multi-site EC-Funded project (four years from Sept 2004) will apply these ideas to design of an intelligent robot: http://www.cs.bham.ac.uk/research/projects/cosy/