What is Artificial Intelligence?
How is it doing?
Where might it go?

Aaron Sloman

School of Computer Science,
The University of Birmingham, UK

ME: http://www.cs.bham.ac.uk/~axs/

PAPERS: http://www.cs.bham.ac.uk/research/cogaff/
TALKS: http://www.cs.bham.ac.uk/research/cogaff/talks/
TOOLS: http://www.cs.bham.ac.uk/research/poplog/freepoplog.html

Last updated (August 29, 2005)

WHAT IS AI?

It is an attempt to make things

- in order to understand things SCIENCE
  - humans (children, adults, processes of learning and development)
  - other animals
  - possible designs for machines
- in order to make new useful tools, and techniques ENGINEERING

In both science and engineering we learn through the processes of

- observing
- thinking
- designing
- testing
- debugging
- extending
- using

all done in parallel

Not just Science: Applications also

As we progress, we can use our increased understanding to produce a succession of robots and other machines doing more and more interesting things, including some that approach some human capabilities (or exceed them)

- **As with all new technologies we shall have to take care that they are put to good, not bad, uses.**

Besides making new machines we can learn to understand old ones better, including ourselves

- **Finding out (through many decades of trial and error)**
- **how to design and build ever more human-like robots and other machines,**
- **and learning what does not work and why,**
- **will help us acquire a deeper understanding of what we are and how we work than we have ever had before,**
- **and this will have profound effects on our approaches to education, counselling, and therapy, for example.**

Example: 'toy' emotional agents.

Note the diversity of types of virtual machines.

AI – the true story

Despite many exaggerated claims, AI is still a long way away from explaining or replicating, the competence of a nest building or hook-making bird, or the learning capabilities of a human toddler.

There are several reasons for this. The main one is immense complexity of bodies, brains and minds produced by several billion years of biological evolution – followed by extraordinary growth of competence produced by individual learning and cultural evolution.

Other reasons why progress has been slow include:

- **our inadequate understanding of what animals and children can do, and**
- **our inability to account for the information-processing involved in biological organisms.**
- **Moreover we are still far from being able to make bodies and body parts with the mechanical capabilities of animal bodies, including their manoeuvreability, their controllability and their power/weight ratios.**

For more on this see the papers prepared for this tutorial at IJCAI-05
http://www.cs.bham.ac.uk/research/projects/cosy/conferences/ijcai-booklet

and the online presentations at
http://www.cs.bham.ac.uk/research/cogaff/talks/

BUT THERE IS CONSTANT PROGRESS, DESPITE THE DIFFICULTIES

See

www.aaai.org/aitopics

for news of AI developments

Show example web page

BEWARE OF OVER-OPTIMISM

Optimistic predictions about rapid progress in AI turned out to be false.

About 20 years ago, AI started betting a bad reputation for 'hype'.

But that merely shows that the people who made the predictions did not understand the problems.

- **I.e. they did not deliberately exaggerate**

There have been many successes in AI, both as science and as engineering.

Despite those successes, AI is still a science in its infancy.

We need to analyse the problems with great care, and try to get people from different disciplines and different subfields working together instead of trying to solve all their problems in isolation.

If we do that, there is hope for sustained continuing progress on one of the greatest scientific challenges of all time, understanding the variety of types of mind produced by evolution.

A key idea: our ordinary concepts are deceptive

We often think we know what we mean when in fact we don't

- Our ordinary concepts such as 'emotion', 'consciousness', 'feeling' are really too full of muddle and confusion to be useful in posing scientific questions or formulating explanatory theories
- We can refine, extend, and subdivide them to produce new more precise, more theoretically-based concepts if we can specify explanatory architectures and see what kinds of states and processes they can generate (as happened when physics and chemistry revised and extended our concepts of kinds of matter.
- We need to understand a wide range of architectures – not just humans: e.g. can an insect have emotions?

For a start, we need to be able to think about architectures of various kinds.
Different architectures will support different kinds of mental states and processes – requiring different collections of concepts.

Example: A simple (insect-like) architecture

A reactive system does not construct complex descriptions of possible futures, evaluate them and then choose one.

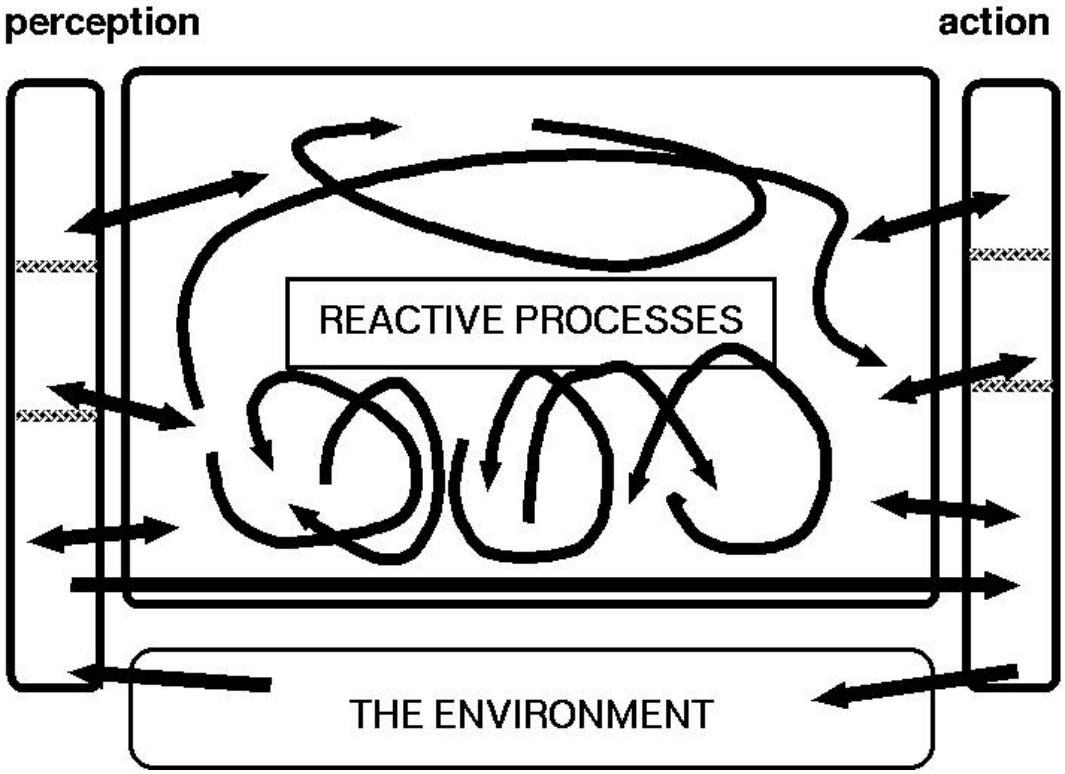
It simply REACTS: internally or externally.

Several reactive sub-mechanisms may operate in parallel.

The arrows indicate information flow (including control).

Processing may use a mixture of analog and discrete mechanisms.

Sensors and motors may operate at different levels of abstraction.



CogAff: A schema for a variety of architectures.

'CogAff' is our label, not for an architecture (like 'H-Cogaff' – presented later), but for a way of specifying architectures in terms of which sorts of components they include and how they are connected.

Think of a grid of co-evolved types of sub-organisms, each contributing to the niches of the others, each performing different functions, using different mechanisms, etc.

We could add lots of arrows between boxes indicating possible routes for flow of information (including control signals) – in principle, mechanisms in any two boxes can be connected in either direction.

This allows many possible architectures, including insects and humans

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

This is not a final taxonomy, just an indication of the sort of thing I am trying to do.

It has many flaws.

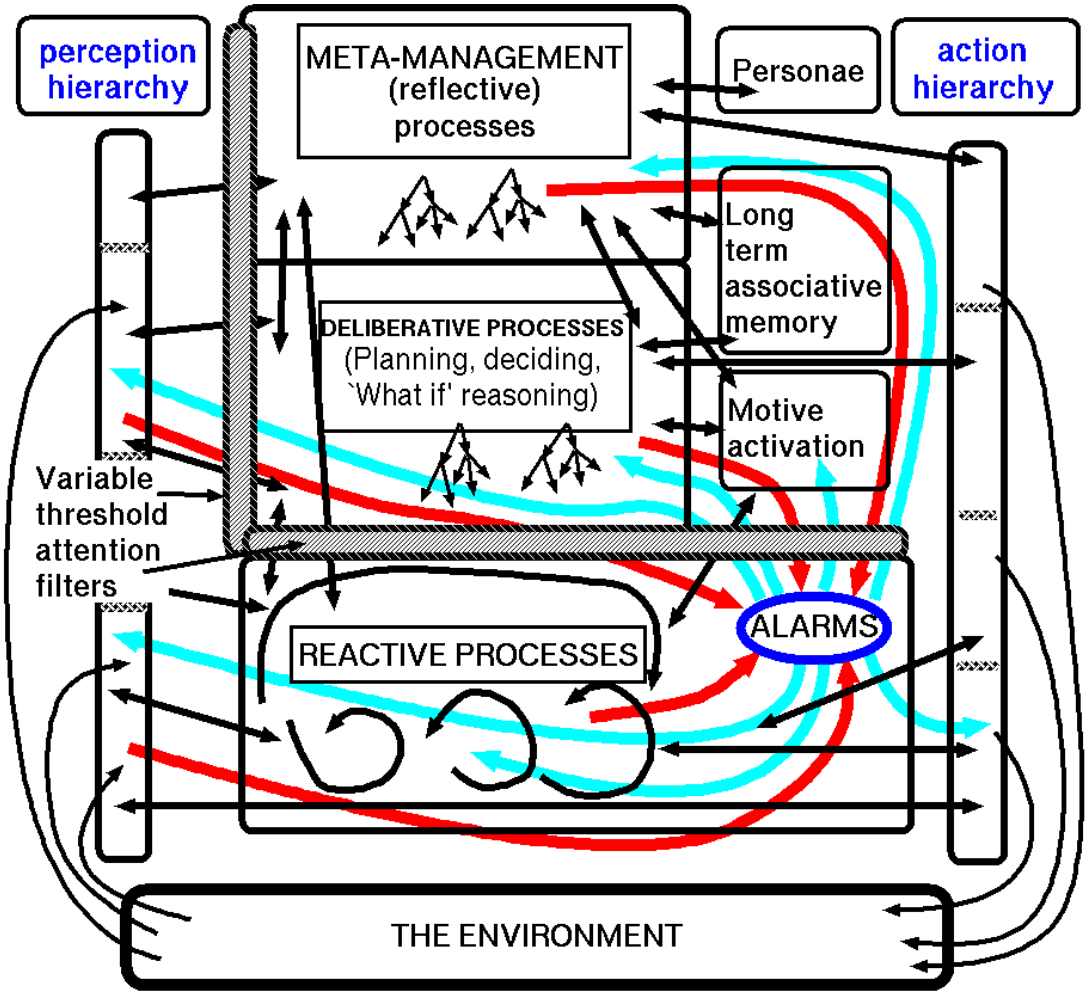
H-CogAff — Your Mind?

(very sketchy first draft – see http://www.cs.bham.ac.uk/research/cogaff/)

The H-Cogaff (Human Cogaff) architecture is a (conjectured) special case of the CogAff architecture schema, containing many different sorts of concurrently active, mutually interacting components.

It includes

- **'old' reactive components shared with many other animals (most species are purely reactive)**
- **'newer' deliberative mechanisms (for considering non-existent possibilities) and**
- **relatively rare meta-management capabilities for inspecting, evaluating, and influencing internal information-processing.**



See other Cognition and Affect papers and talks for details

Multi-disciplinary Co-operation

Making progress requires some of the brightest young minds to study several disciplines that need to be combined in this endeavour, including not only AI, but also

- **philosophy,**
- **psychology,**
- **linguistics,**
- **neuroscience,**
- **biology,**
- **anthropology,**
- **ethology,**
- **engineering and, of course,**
- **mathematics.**

Researchers must not allow administrative and funding constraints, or narrow-minded teachers, to keep their work focused within the borders of specific disciplines or research groups.

Progress requires our vision to be both broad and deep.

It's very difficult.

And very exciting.

I AM OFTEN ASKED

Aren’t you afraid of what super-intelligent machines may one day do to us?

My answer:

The things they do will probably not be nearly as bad as the things humans do to one another.
If they are more intelligent the machines may be less cruel and vicious.

Human cruelty seems to be a result of deep flaws in human minds — by-products of biological evolution, as a result of which people are nationalistic, intolerant, excessively vindictive, religious (superstitious), jealous, etc.)

THANKS

I am very grateful to
the developers of Linux
and other free, open-source,
platform-independent, software systems.

LaTex was used to produce these slides.

Diagrams are created using tgif, freely available from
http://bourbon.cs.umd.edu:8001/tgif/

Demos are built on Poplog

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