THE COGNITION AND AFFECT PROJECT

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

Maintained by
Aaron Sloman
(Don't have time to respond to social media "friend" requests)

Now including the Meta-Morphogenesis Project
This File is available as http://goo.gl/U4vCKE

GOOGLE SEARCH
Use Google to search this site.
(Thanks to Dave Parker for help.)

Search bham/cogaff

GOOGLE SEARCH
Use Google to search CS.
(Thanks to Dave Parker for help.)

Search CS

MAJOR UPDATE: 3 Aug 2016
This web site has been split between:
--- **Contents lists (separate page)**
--- **CogAff Project Overview (this page).**

This file updated:
6 Sep 2014; ... 14 Dec 2015; ... 3 Aug 2016; 19 Oct 2018

CONTENTS

- Separate file listing main contents of this site.

- Origins and Overview of The Cognition and Affect (CogAff) Project
- Overview of Meta-Morphogenesis project (since 2011)
- PhD and MPhil Theses associated with the project
  (Separate file)
- Start of the CogAff (Cognition and Affect) Project in Birmingham, in 1991
  (Glyn Humphreys and Aaron Sloman)
- Links with Biology (Jackie Chappell and Aaron Sloman)
  Fig: The Meta-Configured Genome
- Requirements for architectural theories: The CogAff (generative Schema)
- A special subset of the CogAff schema: Architectures with Alarms
- Alarm mechanisms, states and processes (added 6 Nov 2013)
- Another special subset of the CogAff schema: Omega Architectures

- The CogAff Architecture Schema and the H-CogAff special case
- More detailed list of aims and topics. (Separate web page).
- The CogAff project is inherently interdisciplinary
- Web site contents (separate file)
- List of Cogaff talks/presentations (mostly pdf/postscript)
- CoSy project papers and presentations.
- Present and past collaborators:
- Software Tools Available Free Of Charge/Open Source

- Related developments elsewhere: Biologically Inspired Cognitive Architectures (BICA)

- Miscellaneous stuff (Another directory)
  (Philosophy, biology, AI, speculation, discussion, archives of some news group postings, etc.)
- Search panels (google)
- Pop-11 Eliza Chat-Bot (another web page)
- Formats of our documents
  In another file.
- The SimAgent AI toolkit
- DAM Symposium: How to Design a Functional Mind (at AISB 2000)
- A Tribute to Max Clowes

RELATED FILES

- Presentations (roughly 2000 onwards) (PDF)
  in http://www.cs.bham.ac.uk/research/cogaff/talks/ (some also on slideshare)
- Miscellaneous notes and discussion papers.
- Movies Directory: Video and audio recordings
- A few Audio recordings

- PhD and MPhil Theses associated with the project

- The Free Poplog Portal
  (Software tools and teaching materials for research and education in AI, Cognitive Science and Programming.
- Teaching examples.)

- Luc Beaudoin, who was the first PhD student on the CogAff project and made major contributions to the ideas about requirements and architectures, is continuing to develop themes related to this project. For Luc's current activities see https://cogzest.com

---

A sample of related materials on this web site

- The Meta-Morphogenesis Project
  Meta-Morphogenesis: How could our minds and the rest of life have come from a cloud of dust?

- "Could a baby robot grow up to be a mathematician?"
  Birmingham Popular Maths Lectures
  Hosted by the School of Mathematics at the University of Birmingham. 8 pm, 17 April 2013

- Symposium on AI-Inspired Biology at AISB10, Spring 2010

- List of related talks, with slides, mostly PDF, some also on slideshare.net

- http://www.cs.bham.ac.uk/research/projects/cogaff/misc/AREADME.html
  A large and growing collection of discussion papers, frequently revised or extended.

The rest of this web page gives a high level overview of the Cogaff project and related projects. It includes a roughly chronologically organised collection of papers since the 1960s grouped by year of addition to this web site.

There are also links in the table below to projects that overlap with CogAff, including, since 2004, collaborative projects in Cognitive Robotics (CoSy 2004-2008, CogX 2008-2012). Now too many to list here!

WARNING: BROKEN LINKS

Apology From time to time the central university authorities reorganise campus web pages, without taking action to ensure that references to old links are trapped and redirected. This is a common problem on university web sites.

As a result there are probably several broken links on this web site. If you find any link from or to the CogAff site that does not work, please let me know, at a.sloman[at]cs.bham.ac.uk Thanks.

Origins and Overview of The Cognition and Affect (CogAff) Project

(Gratefully acknowledging many collaborators, especially Margaret Boden, Luc Beaudoin, Ian Wright, Riccardo Poli, Brian Logan, Steve Allen, Catriona Kennedy, Nick Hawes, Jeremy Wyatt, Jeremy Baxter, Matthias Scheutz, Dean Petters, Jackie Chappell, Marek Kopicki, Dave Gurnell, Manuela Viezzer, Verónica Esther Arriola Ríos, Michael Zillich, ...)

Key Ideas

Many researchers propose a theory of THE right architecture for a system with some kind of intelligence (e.g. human intelligence).

Although this may be an appropriate way to address a specific technical problem, it is seriously misguided, if done as a contribution to our scientific or philosophical understanding, unless the specific architecture is related to a theory about THE SPACE of POSSIBLE architectures for various kinds of intelligent system.

Such a theory would need to include a survey of the possible types of component, the different ways they can be combined, the different functions that might be present, the different types of information that might be acquired and used, the different ways such information could be represented and processed, the different ways the architecture could come into existence (e.g. built fully formed, or self-assembling), and how various changes in the design affect changes in functionality.

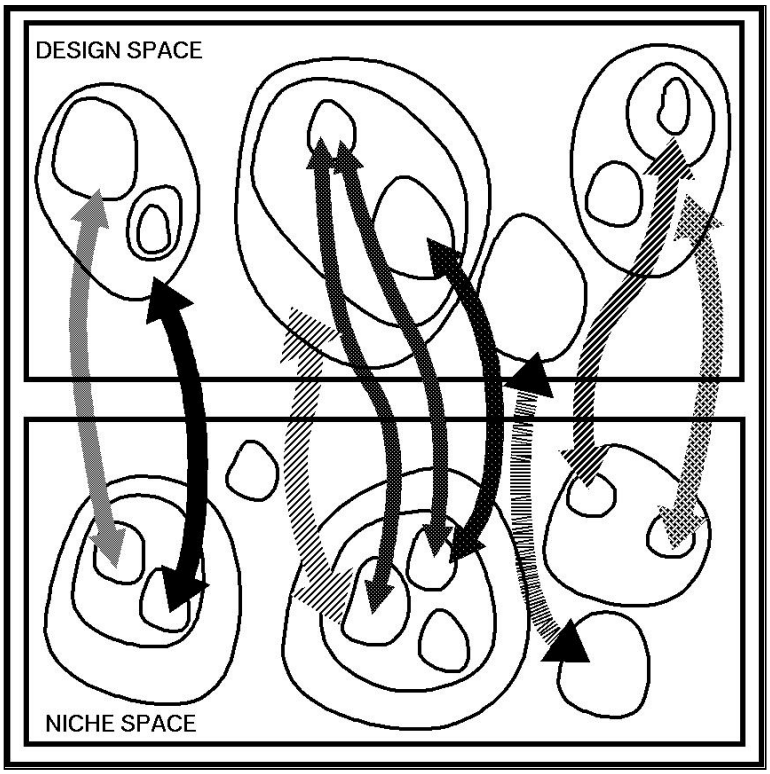
Such a theory also needs to be related to a study of possible sets of requirements for architectures (and for their components). If we don't consider architectures in relation to what they are used for or needed for (in particular types of context) then we have no way of explaining why they should have the features they have or what the trade-offs between alternative design options are.

NB
These investigations should not be restricted to *physical* architectures. Since the mid-twentieth century human engineers have increasingly found virtual machine architectures, in which multiple virtual machine components interact with one another and with physical components. It seems that biological evolution "discovered" the need for virtual machinery, especially self-modifying and self-monitoring virtual machinery, long before human engineers did. This and other "discoveries" by natural selection, and its products, are investigated in the Meta-Morphogenesis project.

Topics investigated include:

- The space of possible **designs** for architectures (design space): (what sorts of design options are available at various stages in the evolution of more complex designs and in various environments?).

- The space of possible sets of **requirements** for architectures (niche space, since a niche is a set of requirements) is distinct from, but related to the space of possible designs.



Arrows indicate that instances of possible designs may be suited to instances of sets of requirements (niches).

- The possible evolutionary, developmental, learning, social, cultural and ecological **trajectories** through these spaces, and the feedback-control loops that they form. (New requirements can lead to new designs to meet those requirements, which can produce new requirements or opportunities for that species, or new requirements and opportunities for other species -- e.g. predators, or prey.)

- How various architectures, mechanisms, and forms of representation meet, or fail to meet, or partially meet, the various sets of requirements against which they can be evaluated. This includes investigating tradeoffs among the alternatives.

- The requirements that led to different sorts of **virtual** machinery and the various designs for such virtual machinery, including evolution of "protocols" for interactions of various sorts between components, between hardware systems, between virtual machine sub-systems, ... the origins of internally self-directed conscious states and the discovery of qualia, and other mental phenomena supported by virtual machinery. Virtual-machine functionalism.

Ignoring the **variety** in these spaces, and instead proposing and studying just ONE architecture (e.g. for an emotional machine) is like doing physics by finding out how things work around the leaning tower of Pisa, and ignoring all other physical environments; or like trying to do biology by studying just one species; or like trying to study chemistry by proposing one complex molecule for investigation.

That's why, unlike other research groups, most of which propose an architecture, argue for its engineering advantages or its evidential support, then build a tool to build models using that architecture, we have tried, instead, to build tools to explore alternative architectures so that we can search the space of designs, including trying to find out which types evolved and why, instead of simply promoting **one** design. Our SimAgent toolkit (sometimes called "sim_agent") was designed to support exploration of that space, unlike toolkits that are committed to a particular type of architecture. Some videos of toy demos mostly produced in the 1990s can be found here.

Start of the CogAff (Cognition and Affect) Project Birmingham, 1991.

The project was begun by Aaron Sloman and Glyn Humphreys (then head of Psychology in Birmingham) in 1991, who later moved to Oxford. (He died suddenly in 2016 and is remembered here.)

When the work began in 1991 it was a continuation of work begun in the 1960s in the School of Social Sciences at The University of Sussex, and later continued in the School of Cognitive and Computing Sciences (COGS). (That, in turn, was a continuation of my 1962 Oxford DPhil Thesis attempting to defend Kant's philosophy of mathematics.)

Some of the earliest work was reported in this 1978 book (now out of print, but available online):
The Computer Revolution in Philosophy: Philosophy, science and models of mind

> The book was originally published in 1978. Thanks to the efforts of Manuela Viezzer and Sammy Snow it was scanned in 2001 and the chapters were made available in html format, with some notes added (since 2002). Later PDF versions of the chapters were derived from the html and copies of the whole book in PDF format made freely available here, and various other places, including the ASSC repository -- since anyone could freely copy it. In 2015, a new revised edition was produced, with all the chapters in a single internally cross-referenced HTML file, from which a PDF version was derived:
> http://www.cs.bham.ac.uk/research/projects/cogaff/crp/crp.html
> http://www.cs.bham.ac.uk/research/projects/cogaff/crp/crp.pdf
> The latest version is also freely available, with a "Creative Commons" licence.
>
> In addition, an "Afterthoughts" document was begun in August 2015, and will continue to grow, freely available here:
> http://www.cs.bham.ac.uk/research/projects/cogaff/crp/crp-afterthoughts.html
> (also in PDF).
>
> Chapter 7 on "Intuition and analogical reasoning", including reasoning with diagrams, and Chapter 8 "On Learning about Numbers" were specially closely related to the 1962 DPhil work on the nature of mathematical knowledge.

After AS moved to Birmingham, the work was partly funded by a grant to Sloman and Humphreys, from the UK Joint Council Initiative (JCI), which paid for a workstation and a studentship. An additional studentship was funded by the Renaissance Trust (Gerry Martin).

The first PhD thesis completed in the project was by Luc Beaudoin (funded by major scholarships from: Quebec's FCAR, The Association of Commonwealth Universities (UK), and the Natural Sciences and Engineering Research Council (NSERC) of Canada). The this is online here, along with others. Among other things, it offered a new, unusually detailed analysis of aspects of motives that can change over time, and introduced the important distinction between *deliberative* mechanisms (which can represent, explore, hypothesise, plan and select possible situations, processes and future actions) and *meta-management* mechanisms which can can monitor, and to some extent control internal processes (including deliberative processes). Some of the ideas are explained in more detail in
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/fully-deliberative.html

Later PhD students who built on and extended the ideas are listed here (with online theses). In particular a paper summarising some of the key ideas in the context of long term grief,

including phenomena that refute many theories of emotions/affect was published (by invitation) in the journal *Philosophy Psychiatry and Psychology* in 1996 http://www.cs.bham.ac.uk/research/projects/cogaff/96-99.html#2.

Similar work elsewhere on architectures for intelligent agents uses labels such as "reflective", "metacognitive", "executive functions", and "self-regulation", though often with different features emphasised. There is still no generally agreed ontology for describing architectures and their functions, unfortunately -- leading to much reinvention of wheels, often poorly designed wheels. (The BICA society (mentioned below) is an attempt to remedy this.)

Later extensions arose from funding by DERA which enabled Brian Logan to work here for several years, followed by a project funded by The Leverhulme Trust on Evolvable virtual information processing architectures for human-like minds, originally set up with Brian Logan, which then paid for Matthias Scheutz to work here for 13 months (2000-2001), followed by Ron Chrisley (2001-2003).

A progress report on the CogAff project was written in 2003 (separate document).

From 2004 related work was funded by the EU, in two projects on cognitive robotics CoSy and CogX.

For a while some of the work was done as part of the Intelligent Robotics research laboratory (led by Jeremy Wyatt) at Birmingham. However, the externally funded projects, partly under pressure from collaborators had to focus on more specific targets and techniques, so the more theoretical work reported here, and in the Meta-Morphogenesis project spawned around 2012, continued mostly independently, though with potential implications for future ambitious robotic projects.

Links with Biology
In 2004, Jackie Chappell arrived in the School of Biosciences (having previously worked in Oxford, in the Behavioural Ecology department led by Prof Alex Kacelnik). Our ways of thinking about intelligence in animals had significant overlap so we worked together on extending biologists' ideas about "Altricial" and "Precocial" species to robots and investigating nature-nurture tradeoffs in animals.

Our theoretical research on animal cognition then expanded e.g. to include work on varieties of understanding of causation (Humean and Kantian) in animals and machines. From 2008 this was further expanded to include studies of cognition in orangutans, in collaboration with Susannah Thorpe, and their PhD students, also in the School of Biosciences,

CogAff is really a loose, informal, collection of sub-projects, most of them unfunded at any time, including research on architectures, forms of representation and mechanisms occurring in humans, other animals, and human-like machines.

Some additional topics covered can be found in this document compiled in 2009 and this list of online discussion papers (frequently extended).

Analysing such architectures, and the mental states and processes they can support, allows us to investigate, for instance, whether consciousness or the ability to have emotional states is an accident of animal evolution or a direct evolutionary consequence of biological requirements or a side-effect of things meeting other requirements and constraints.

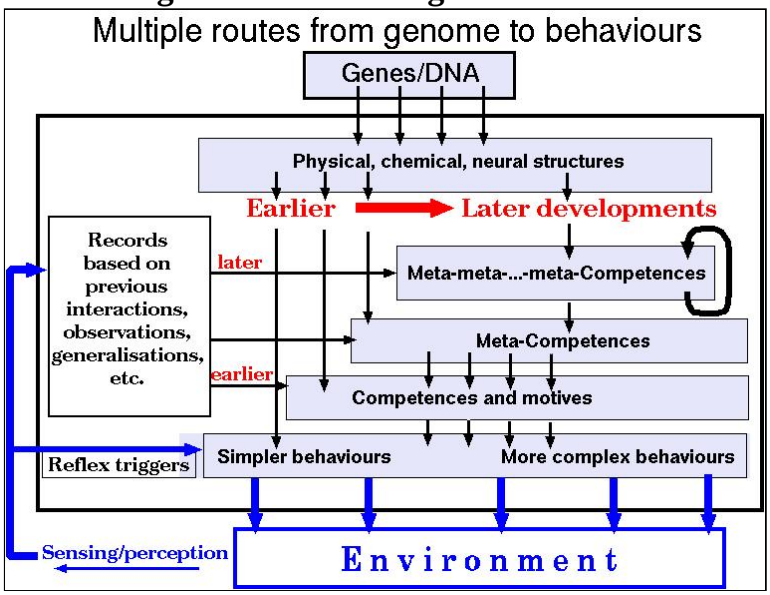
One of the outcomes of this research was development of the CogAff schema introduced above and (explained briefly in this poster). The schema (especially when elaborated beyond

those simple diagrammatic specifications) is a high level abstraction that can be instantiated in many different special case architectures. This provides a way of characterising a wide range of types of possible architecture in natural and artificial systems (in contrast with most researchers on cognitive architectures who promote a particular architecture).

A special case (or subclass) of CogAff is the H-CogAff (Human-Cogaff) architecture, described below, which is still currently too difficult to implement, though various subsets have been implemented by researchers here and elsewhere. Some "toy" versions are used in demonstration videos of student programs running.

The collaboration with Jackie Chappell, and our presentation at IJCAI 2005, led to an invited paper for a new interdisciplinary journal in 2007 "Natural and artificial meta-configured altricial information-processing systems" *International Journal of Unconventional Computing* available here. Some of our key ideas, extending Waddington's ideas about the "Epigenetic Landscape", are summarised (perhaps too obscurely) in this diagram (extending the version in the paper):

Fig: The Meta-Configured Genome



One of the implications of those ideas is that the information processing architecture of an organism can change during development in ways that depend on the influence of the environment -- insofar as the environment can influence how the genome is expressed, as indicated roughly in the above diagram. This is one of the themes of the Meta-Morphogenesis project mentioned several times in this document. For more detail see http://www.cs.bham.ac.uk/research/projects/cogaff/misc/meta-configured-genome.html (also PDF)

Requirements for architectural theories: The CogAff (generative Schema)

- 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.a sensory/perceptual processes constantly changing to represent the environment (including internal states)

2.a motor/action/effector processes constantly changing the environment and perhaps some internal states

3.a central, more slowly changing, processes

or between

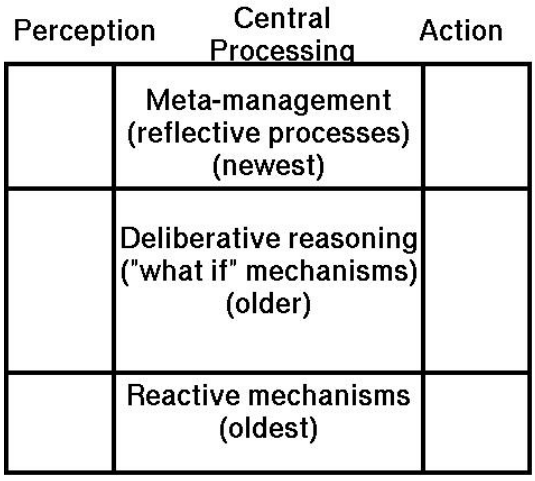
1.b Evolutionarily very old reactive processes, constantly driven by what is sensed internally and externally

2.b Newer deliberative processes able to represent what does not exist but might, e.g. future actions, unseen situations, past causes.

3.b Specialised meta-management/reflective processes capable of describing information-processes states and processes in oneself and also in others. Debates about which came first, self understanding or other understanding are futile: they almost certainly grew together in fits and starts. Some further details concerning these distinctions are available here:
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/fully-deliberative.html

By superimposing the above two classifications we get the following suggestive, but in some ways misleading, 3x3 grid of possible types of architectural component, misleading because not all the required mechanisms will fit into just one of the boxes.

The Cogaff Schema:



The CogAff schema shown above summarises this space of possible types of architectural component.

-- The first three divisions above (1.a--3.a) correspond to the vertical divisions in the schema.

-- The second three divisions above (1.b--3.b) correspond to the horizontal divisions in the schema: evolutionarily oldest functions in the bottom layer.

This is an over-simplification (a) because each layer should be more finely divided into components performing different functions, (b) because the columns and layers should overlap more (as in the diagram below), and (c) because there are mechanisms that

straddle, or link components in, the various boxes in the diagram, including the "alarm" mechanisms that can play a role in emotions and other affective states. A revised (but still inadequate) version of the diagram is presented below.

Connections with Marvin Minsky's ideas

I met Marvin Minsky briefly during a conference at MIT in 1975, but more detailed interaction began later via online discussions using Usenet/NetNews. I had previously met John McCarthy at IJCAI 1971, where I criticised his important 1969 paper with Pat Hayes. Usenet discussions of consciousness, emotions and other topics often involved both McCarthy and Minsky. Both of them accepted my invitation to take part in a special panel on philosophy and AI at IJCAI 1995, as shown in this picture http://www.cs.bham.ac.uk/research/projects/cogaff/marvin-minsky.html#phil-encounter. His 2006 book *The Emotion Machine* explicitly refers to some of our architectural ideas, but he preferred to separate out the architectural scheme into more layers, e.g. dividing our reactive layer into reflexes and other reactions.

NOTE - A student video on the CogAff schema

Jonathan Metallo and Daniel Lohmer gave a short and entertaining tutorial video presentation on some of the architectural ideas summarised below. The video is available here:
http://www.youtube.com/watch?v=Twzw9iFOspI
Jonny M and Dani L talk about AI architecture and Sloman. This appears to be an assignment for a course on "Perspectives on Artificial Intelligence, Robotics, and Humanity", in The Department of Computer Science and Engineering at the University of Notre Dame.

NB A Schema for architectures is not an architecture.

It is more like a grammar. Instances of the schema are like sentences in the grammar. However the CogAff schema is a grammar whose 'sentences' are not strings but quite complex networks of concurrently active mechanisms with different functions, as discussed in this paper on Virtual Machine Functionalism (VMF).

I have begun to discuss ways in which these ideas could shed light on autism and other developmental abnormalities, in
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/autism.html

---

A special subset of the CogAff schema: Architectures with Alarms

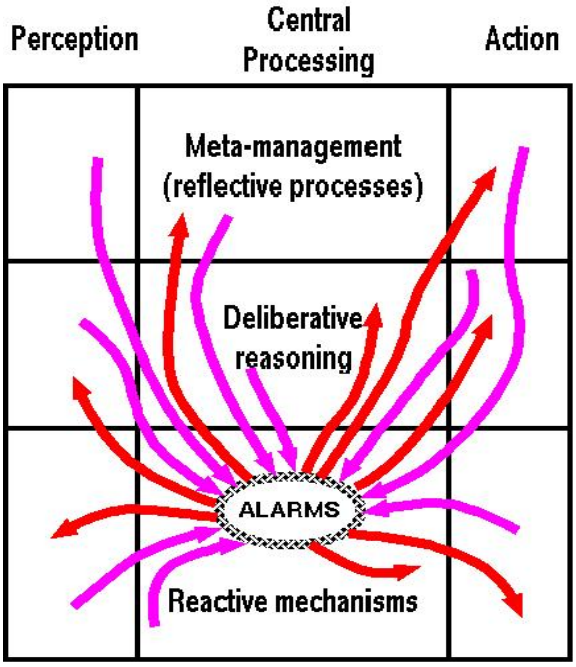


Fig Alarms

Alarm mechanisms, states and processes (added 6 Nov 2013)
Many organisms seem to have, and many robots and other intelligent machines will need, an "alarm" mechanism, which receives input from many of the internal and external sensors and is capable of recognising patterns that require very rapid global reorganisation of ongoing processes, for example switching into states like fleeing, attacking, freezing, or attending closely to what may or may not be a sign of serious danger or some opportunity.

This kind of mechanism seems to be very old in animal evolution and can be observed in a woodlouse, for example, when it reacts to being touched by rolling itself up in a ball, or a fly which reacts to the rapid approach of a fly-swat by stopping whatever it is doing (e.g. feeding) and switching to an escape action.

The kinds of effects of alarm mechanisms will obviously depend on various factors such as

- Where their inputs come from (e.g. direct from low-level perceptual subsystems, or indirectly via more sophisticated processing layers);
- Where the outputs go (e.g. direct to low level motor control subsystems or to some other part of the architecture, e.g. awakening some memory, or re-activating a dormant motive, or grabbing control of many different resources);
- How long the 'alarmed' states last -- some are very short (e.g. being started) others may last a long time as in extended anxiety or grief, possibly enduring in parallel with other, unrelated, short term alarms, e.g. being startled by a noise;
- Whether the outputs from the alarm system actually interfere with other subsystems or merely have a tendency to interfere, but fail to have an effect (we use the label "perturbance" to include these dispositions to perturb or disturb, which may be real but ineffective, or partially effective, etc.
- Many theories of emotion (especially theories that over-emphasise bodily states, or action tendencies, like the James-Lange theory and its many derivatives) imply or state that an individual can be in only one emotional state at a time, whereas this architecture allows several different sorts of perturbance to coexist, including different sorts. It is also possible for an enduring state to be temporarily dormant, as in grief or jealousy.
- We do not claim that humans are born with such an architecture fully developed: it obviously grows over time. (A newborn human infant is incapable of feeling despair at the long term effects of human population growth.) That implies that at different stages of development different affective states are possible, including different emotions.

- Since emotional states can very in their complexity, in the architectural resources, conceptual apparatus and potential interactions with other states and processes, no theory of emotions that places them in a space of some fixed set of dimensions can be correct. Rather we need something like **grammar** for types of emotion and for types of emotional process extended in time. See Sloman(1982)
  See also: Aaron Sloman, Review of **Affective Computing** by Rosalind W. Picard, 1997, in **The AI Magazine,** 20, 1, 1999, pp. 127--133, http://www.cs.bham.ac.uk/research/projects/cogaff/Sloman.picard.review.pdf That and other CogAff papers in the 1990s proposed that the then popular distinction between "primary" and "secondary" emotions based on different criteria used by different authors should be expanded to include at least the following:
  - Primary emotions: involving only mechanisms at the reactive level.
  - Secondary emotions: involving some mechanisms at the deliberative level (in addition to reactive mechanisms).
  - Tertiary emotions: involving some mechanisms at the meta-management level, making use of meta-semantic concepts possibly referring to one's own mental states and processes, or those of others. (Recently other researchers have been using the primary/secondary/tertiary labels in different ways, not usually architecture-based.)
  - In addition there are many affective states and processes that are not emotions, including desires, preferences, intentions, attitudes, ambitions, moods,..., and probably many more components of the architecture in different species.

  Note:
  It can be difficult to get 'professional' emotion researchers to take such theories seriously, since they are educated to use observation, measurements, questionnaires, statistical packages, but not to think about how to design, test and debug complex **working** information-processing systems, which is what humans and other animals are. See Margaret A. Boden, **Mind As Machine: A history of Cognitive Science (Vols 1--2)** Oxford University Press, 2006,

  A symptom of that educational inadequacy is attempting to find a few "dimensions" in which features of emotions can be classified and then taking the resulting 2 or 3 (or sometimes higher) dimensional grid as providing a principled classification of emotions.

  But complex information processing systems can get into states that vary in far more complex ways than such simple-minded tables or graphs can accommodate: as any system designer knows.

Special cases of the CogAff schema

A very crude depiction of an insect-like information processing architecture with alarms could be something like this:

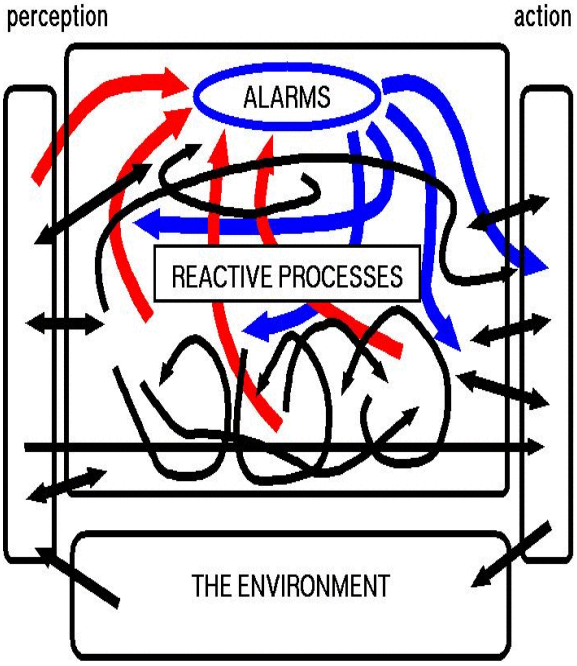


Fig Insect

An insect-like special case of the CogAff schema is purely reactive -- none of the deliberative or meta-management functions are provided, though reactive mechanisms may be layered, as indicated crudely in the diagram.

[Modified: 7 Dec 2012] A purely reactive system that always merely reacts to particular stimuli could be modified to include "proto-deliberative" mechanisms (unfortunately labelled "deliberative" by Michael Arbib at a conference in 2002). In a proto-deliberative system, reactive mechanisms can simultaneously trigger two incompatible response-tendencies. Since in general a blend of two incompatible responses is worse than either response, it can be useful to have mechanisms for choosing one of them, e.g. using a comparison of strength, or some other mechanism such as always letting escape reactions win over feeding reactions, or using a self-organising neural net capable of achieving a variety of potentially-stable states, then adopting one.

For more more on different intermediate cases see this discussion of "Fully Deliberative" systems.

In such a (relatively) simple architecture, alarm mechanisms can trigger simple emotions (e.g. in the woodlouse that rapidly curls up in a ball if touched while walking).

Another special subset of the CogAff schema: Omega Architectures

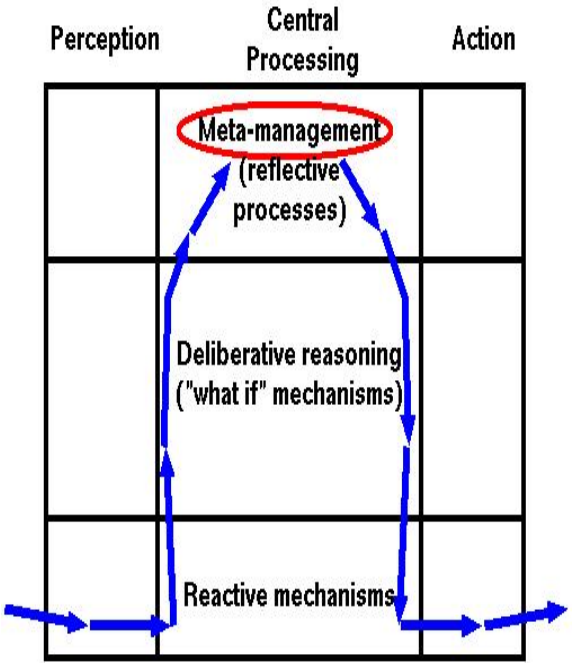


Fig Omega

Architectures of this general type where the flow of information and control can be thought of as roughly like the Greek capital letter Omega **Ω** (not necessarily presented in this sort of diagram) are often re-invented.

The assumption is that perception consists of detection of low level physical signals that are processed at increasing levels of abstraction until the processing generates new goals or preferences, at which point some selection mechanism (e.g. contention scheduling) chooses the best motive or action and the then signals propagate downwards to the motor subsystems which then produce behaviour.

This "peephole" (or "peep-hole") view of perception and action contrasts with the "multi-window" view of both perception and action as involving concurrent processing at different levels of abstraction, partly under the control of the environment and partly under the control of various layers of central information processing, operating in parallel.

So some of the more abstract perceptual or motor processing can be thought of as both cognitive insofar as they make use of forms of representation and ontologies shared with more central processing mechanisms, and also as peripheral (e.g. perception or action processes) because the information structures used are maintained in registration with perceptual input signals (or the optic array in the case of visual input) or in registration with motor signal arrays, and also because the processing at those more abstract levels is bi-directionally closely coupled with the lower level perceptual or motor signals.

These extra layers of perceptual or motor processing are fairly obviously needed for language production or perception because it is now well understood that linguistic expressions have structures at different levels of abstraction that all need specialised processing. Our claim is that that is a special case of a far more general phenomenon (as illustrated in the POPEYE program described in Chapter 9 of The Computer Revolution in in Philosophy 1978).

A much more complex special case (or subset of special cases) of the CogAff schema: H-CogAff (Human-inspired CogAff).

A poster summarising some of the main theoretical ideas is here (PDF).

Some dimensions in which architectures can vary were presented at the Designing a Mind Symposium on in 2000 in "Models of models of mind." However, that paper is inadequate in several ways, e.g. because it does not clearly distinguish the CogAff schema from the H-CogAff special case, presented briefly below.

It has other flaws that need to be remedied, in part by extending the analysis of ways in which architectures can differ, in part inspired by the diversity produced by biological evolution, and in part by inspiring deeper analyses of that diversity as proposed at the AIIB symposium in 2010.)

The CogAff Architecture Schema and the H-CogAff special case

The name "CogAff" is used both for the project and as a label for a generic schema proposed several years ago for a wide variety of architectures, natural and artificial. (We don't claim it is general enough to cover all cases: some of the distinctions are not fine-grained enough. But it illustrates a style of research on architectures that is unfortunately rare.)

This Schema, as explained above, classifies requirements for the major components of an architecture into nine broad categories on a 3x3 grid which can be connected together in different ways (depending on how various kinds of information - factual information, queries, control information, etc, flow between subsystems).

This is just a first crude sub-division, requiring more detailed analysis and further decomposition of cases (as illustrated here). However it does cover many different types of architecture, natural and artificial, depicted rather abstractly above.

Architectures vary according to what mechanisms they have in the boxes, and how they are connected. Also more complex architectures may have important subdivisions and possibly may require functions that don't fit neatly into any of the boxes. (For example, it is arguable that the mechanisms concerned with production and understanding of language, and use of language for thinking and reasoning, are scattered over many different subsystems.)

The generic CogAff schema includes an important sub-class of architectures that include mechanisms capable of producing what might be called "emotional" or "alarm" reactions, as shown in the "insect-like" special case, above.

A much more complex special case is the H-CogAff architecture, which we suggest provides a very high level "birds-eye view" of the architecture of a typical
(adult) human mind, depicted crudely here (as a first approximation):

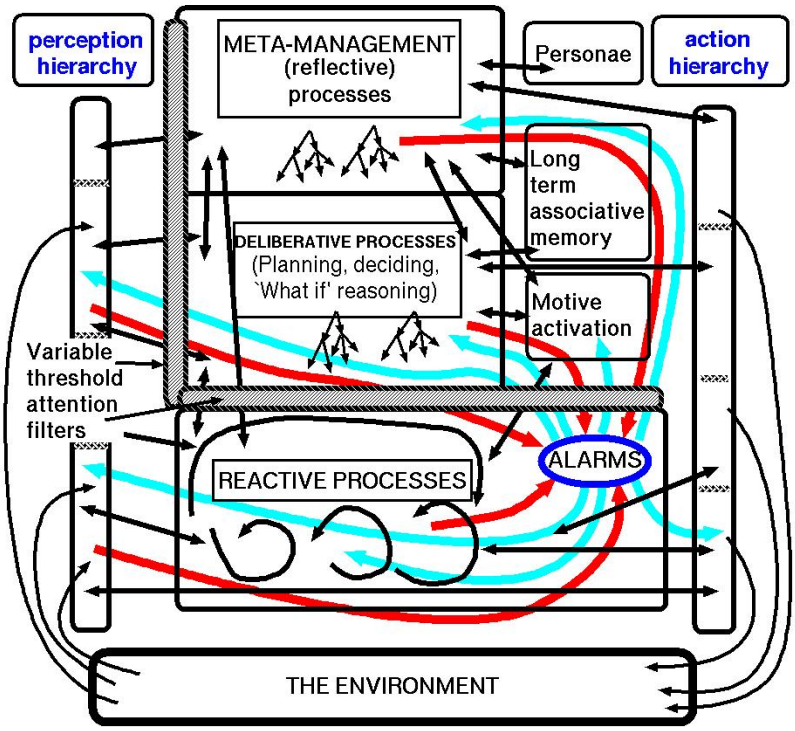


Fig H-Cogaff

It includes concurrently active sub-architectures that evolved at different times in our evolutionary history, in addition to sub-architectures that grow themselves during individual development (as discussed in this paper by Chappell and Sloman.)

A paper summarising the ideas behind the CogAff schema and the H-CogAff architecture is this 2003 progress report on the Cogaff project.

A paper published in 1996 (published with commentaries) explained how emotional phenomena like long-lasting grief could be accommodated within this framework

> I.P. Wright, A. Sloman, L.P. Beaudoin,
> Towards a Design-Based Analysis of Emotional Episodes, *Philosophy Psychiatry and Psychology,* 3, 2, pp. 101--126, 1996,
> http://www.cs.bham.ac.uk/research/projects/cogaff/96-99.html#2

Further details are provided in other papers, including for example this polemical piece:

> Some Requirements for Human-like Robots:
> Why the recent over-emphasis on embodiment has held up progress (2008).
> Now published in *Creating Brain-like Intelligence,*
> Eds. B. Sendhoff, E. Koerner, O. Sporns and H. Ritter and K. Doya,
> Springer-Verlag, 2009 Berlin,
> http://rapidshare.com/files/209786694/Creating_Brain-Like_Intelligence.zip

An incomplete survey of types of architecture that include various types of "deliberative layer" can be found in "Requirements for a Fully Deliberative Architecture"
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/fully-deliberative.html

Some designs described as "deliberative" by other authors include only what we call "proto-deliberative" mechanisms.

Most of the hypothesised architectures are still too difficult to implement though some of the simpler ones have been implemented using the SimAgent toolkit, and demonstrated here.

More complex examples were developed within the EU-funded CoSy robot project (2004-2008), and are being extended in its sequel the CogX robot project (2004-2012).

Tutorial presentations of how ideas like "qualia" and some of the vexing problems of consciousness ("the explanatory gap") can be understood in this framework are presented here.

In 1998 Gerd Ruebenstrunk presented some of our ideas for German readers in his diploma thesis in psychology on "Emotional Computers" (Bielefeld University, 1998). See especially sections 9 and 10 of his thesis.. His 2004 presentation on emotions, at a workshop on "Affective Systems" (in English) is here.

Some of the ideas presented here, including what has been referred to as the use of multi-window perception and action seem to be closely related to some of the architectural ideas in this book (though we have some serious disagreements about the notion of 'self' and about consciousness):

```
Arnold Trehub,
The Cognitive Brain, MIT Press, Cambridge, MA, 1991,
http://www.people.umass.edu/trehub/
```

A more accurate but more obscure(?) version of the schema
(inserted 21 Mar 2013)

The previous diagrams do not make it clear enough that perceptual and action/motor mechanisms overlap. E.g. (as J.J.Gibson pointed out in *The Senses Considered as Perceptual Systems* (1966)), mechanisms of vision depend on the use of saccades, head movement, and whole body movements, and haptic sensing depends on controlled movements of hands, tongue, lips, etc.

The following diagram is an attempt to remedy this deficiency in the previous diagrams.

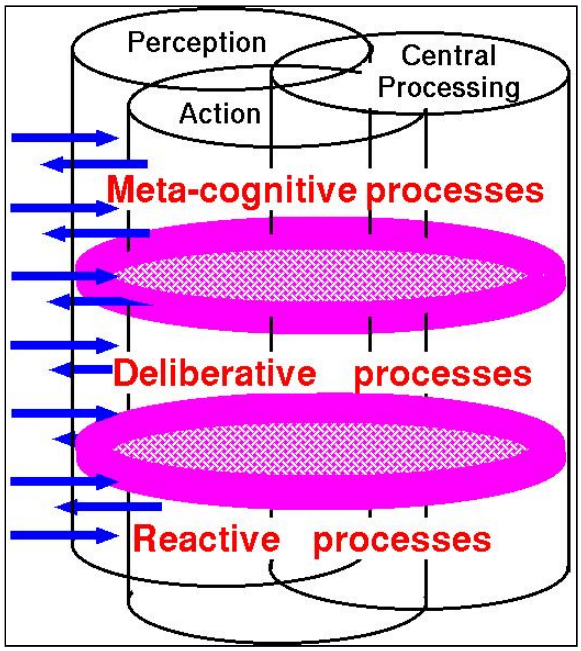


Fig CogArch

Note: the above diagram does not show the "alarm" processing routes and mechanisms described below.
(With thanks to Dean Petters, who produced a first draft of the above diagram.)

Some of the missing structural and functional relations in the above diagram are included in the next diagram, which shows the "alarm" processing routes and mechanisms described in other CogAff papers (allowing asynchronous interruption or modulation of ongoing processes, e.g. to meet sudden threats, opportunities, etc.)

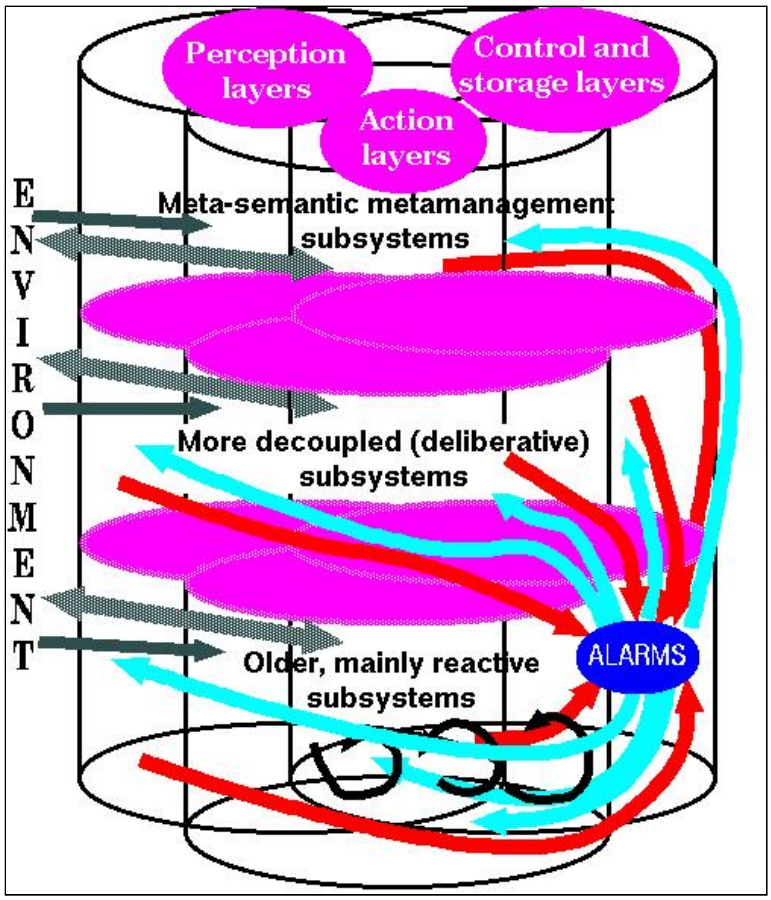


Fig NewCogArch

(Also with help from Dean Petters.)
Compare the BICA (Biologically Inspired Cognitive Architecture) web site:
http://bicasociety.org/cogarch/

There are additional complexities not shown in the above diagrams, including the architectural decomposition at each layer, the complex sub-architectures straddling layers, e.g. for several different kinds of long term memory, for vision, for behaviour initiation and motor control, for language use, for learning, for many kinds of motivation, for personality formation, for social and sexual interaction, and many more.

A Dynamical Systems view of H-CogAff

To be added.

See also:

- Aaron Sloman, The mind as a control system, in *Philosophy and the Cognitive Sciences*, Eds. C. Hookway and D. Peterson, CUP 1993, pp. 69--110, http://www.cs.bham.ac.uk/research/projects/cogaff/81-95.html#18

- A Multi-picture Challenge for Theories of Vision
  http://www.cs.bham.ac.uk/research/projects/cogaff/misc/multipic-challenge.pdf

---

The CogAff project is inherently interdisciplinary

This work has (surprisingly?) many links with other disciplines, including several branches of philosophy, for example:

- Philosophy of mathematics,
- Philosophy of Science
- Ethics
- Meta-ethics
- Theories of meaning
- Philosophy of language
- Philosophy of mind
- Metaphysics and ontology
- The nature of conceptual analysis.
- The analysis of concepts of causation (Humean and Kantian)
- Added 2011-12: The Meta-Morphogenesis project:
  http://www.cs.bham.ac.uk/research/projects/cogaff/misc/meta-morphogenesis.html

---

We have links with several other groups of researchers at Birmingham

An interdisciplinary Centre for Research in Computational Neuroscience and Cognitive Robotics,
led by the Schools of Psychology and Computer Science, was approved by the University in 2009.

- Researchers in The Intelligent Robotics Lab and various robotic projects, including especially
  - The EU-funded CoSy Project (2004-2008)
  - The EU-funded CogX Project (2008-2012)
- Research in animal cognition, especially cognition in birds that manipulate things,
- Various groups in Psychology, including
  - PRISM Group
  - The Symon Group
  - The Cognitive Neuroscience and Neuropsychology Group,
  - The Language, Cognition and Perceptual Systems Group
  - The Cognitive NeuroImaging Group
- We also interact with Philosophers here and elsewhere.
- There are overlaps with research in several other research groups in The School of Computer Science.
- Related work in the University of Birmingham

COGAFF PAPERS AND PRESENTATIONS

Managed by Aaron Sloman.

Associated with the CoSy Robotics Project since 2004

WEB SITE CONTENTS

Cognition and Affect Project Papers, Presentations, Theses, Software

- SLIDE PRESENTATIONS on COGAFF TOPICS
  (PDF.)
  in http://www.cs.bham.ac.uk/research/cogaff/talks/
- Various audio recordings relevant to the project
- Movies Directory: Video and audio recordings relevant to the project
- Chronological lists (above)
  PAPERS published or inserted here
- Search panels (google)
- Intelligent Robotics Lab in Birmingham
- The EU-funded CoSy (Cognitive Systems Robotics) Project at Birmingham
  Consortium web site
- The EU-funded CogX project
  The consortium web site
- CoSy papers, presentations and discussion notes
- Draft list of papers and presentations on requirements for human-like AI systems, robots, etc.
- PhD and MPhil Theses associated with the project
- **The Computer Revolution In Philosophy: Philosophy Science and Models of Mind**
  Readable online in HTML and PDF formats.
  Also available as a complete PDF book.
  This 1978 book is available online free of charge, in a new electronic edition 2015, under a Creative Commons licence.
  An Afterthoughts document will have notes and comments added from time to time, including responses to some criticisms by reviewers:
  http://www.cs.bham.ac.uk/research/projects/cogaff/crp/crp-afterthoughts.html
  http://www.cs.bham.ac.uk/research/projects/cogaff/crp/crp-afterthoughts.pdf
- Miscellaneous stuff
  (Philosophy, biology, AI, speculation, discussion, archives of some news group postings, etc.)

- **Present and past collaborators:**
  - Luc Beaudoin: Author of the first CogAff PhD thesis Goal processing in autonomous agents (1994)
    He introduced the concept of "meta-management" as involving both concurrent self-observation and self control (including monitoring and control of deliberative processes) -- central to our work on varieties of architecture, as reported here.
  - Ian Wright
    https://sites.google.com/site/ianwrightoxford/
    Worked closely with Luc (and other students) and later obtained from Sony the studentship that funded Nick Hawes' PhD. Ian was the lead author of the (invited) paper (partly about long term grief) also mentioned above:
    Towards a Design-Based Analysis of Emotional Episodes,
    and also contributed ideas about loop-closing semantics. See his new research site
    https://sites.google.com/site/ianwrightoxford/research/philosophy-of-mind-ai

  - Jeremy Wyatt (Birmingham)
  - Nick Hawes (Birmingham)
  - Richard Dearden (Birmingham)
  - Dean Petters (Birmingham)
  - Jackie Chappell (Biosciences, Birmingham)
  - Susannah Thorpe (Biosciences, Birmingham)
  - Margaret Boden, at the University of Sussex
  - Ron Chrisley at the University of Sussex
  - Matthias Scheutz now at Tufts University
  - Brian Logan at The University of Nottingham
  - Darryl Davis at the university of Hull
- An older list of people involved
- Students and Theses

- **Some influences on this work**
  - Immanuel Kant also on wikipedia
    (who would be doing AI if he were alive now).
  - Gottlob Frege also on internet encyclopaedia of philosophy and wikipedia
  - Karl Popper, also on Stanford Encyclopedia and BBC discussion led by Melvyn Bragge
  - Imre Lakatos. See a summary of his philosophy of science.
  - Gilbert Ryle and Some pictures and links
  - Marvin Minsky
  - John McCarthy
  - Max Clowes

NEWS: AUDIO BROADCAST ONLINE:
Audio discussion broadcast on Deutschlandradio on 'Emotional Computers' online
(mostly in German), chaired by Maximilian Schönherr.
The audio link is on the right, under 'AUDIO ON DEMAND'. Click on 'Emotionale Agenten'.

Audio interview on grand challenge (December 2004)

RESEARCH GRAND CHALLENGE:

In 2002, the UK Computing Research Committee (UKCRC) initiated a discussion of research grand challenges. One of these is Grand Challenge 5: 'Architecture of Brain and Mind' For more information see http://www.cs.bham.ac.uk/research/cogaff/gc/

OUR SOFTWARE TOOLS ARE AVAILABLE FREE OF CHARGE/OPEN SOURCE

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

- http://www.cs.bham.ac.uk/research/projects/poplog/examples
  (Introduction to teaching materials using AI programming.)
- The SimAgent Toolkit and some
- demonstration movies of the Simagent Toolkit.

Related developments elsewhere: Biologically Inspired Cognitive Architectures (BICA)

The organisers of the BICA (Biologically Inspired Cognitive Architectures) workshops/conferences have begun to address this problem in a promising way.

Here are some links (BICA and related sites):

- The BICA Society
- Past and future BICA meetings
- BICA Journal
  (Unfortunately published by Elsevier.)
- CogArch Repository
  Comparative Repository of Cognitive Architectures, Models, Tasks and Data.
- Ron Sun's Architectures Page
- umich.edu Cognitive Architectures page
  By Bill Lemon, David Pynadath, Glenn Taylor and Bob Wray.
- Report on AIIB symposium, Spring 2010

Other links

- Requirements for a Fully Deliberative Architecture.
  Discussion note on some possible architectural sub-divisions.

- A First Draft Analysis of Some Meta-Requirements for Cognitive Systems in Robots (With David Vernon.)

- Architecture-Based Motivation vs Reward-Based Motivation

- The Design-Based Approach to the Study of Mind
  (in humans, other animals, and machines).

THIS WEB SITE CONTAINS PAPERS AND PhD/MPhil THESES ON MANY TOPICS, INCLUDING:
(Use google and "CogAff" to search for more.)

- Philosophical foundations of computing and AI
- Architectures for intelligent agents of various sorts (natural and artificial)
- Emotions, moods, attitudes, and other affective states and processes
  (E.g. motives, moods, desires, preferences, pains, pleasures, attitudes, ec.)
- Computers and emotions, synthetic agents and emotions.
  Deep and shallow models.
- What representations are, and how many varieties there are
- Diagrammatic/spatial reasoning
- Consciousness - what it is and isn't, and how various types might have evolved
  (consciousness in humans, in other animals, and in machines)
- Evolution and Co-evolution
  (Evolution via concurrent interacting trajectories in ``design space" and ``niche space": a non-mystical version of Gaia?)
- Evolvable architectures for human-like minds.
  (The evolutionary history of an architecture has implications for what is in the architecture.)
- Tools for exploring agent designs, Tools for Artificial Intelligence
- Philosophy of mind
- Philosophy of Artificial Intelligence
- Philosophy of Cognitive Science
- Philosophy of Mathematics (and its links to robotics)
- Ontology, including how a young child or robot needs to develop the ontologies it uses.
- The nature of causation and whether events in virtual machines can be real causes
  (Including "downward" causation")
- Philosophy of Computation
  E.g. the nature of virtual machines, and how the implementation of virtual machines in physical machines is related to the philosopher's notion of supervenience of minds on matter.
- Emergence of various kinds
- Nature-Nurture tradeoffs
  E.g. Multiple routes from genome to behaviour in precocial and altricial species, the latter a mixture of preconfigured and meta-configured competences.
- Nature-nurture tradeoffs for robots.
- Evolution of language
- Free will (the good the bad and the ugly kinds)
  .....

For more details see

- Online versions of presentations given since 2001.
- A large and growing collection of discussion notes.
- Papers and presentations related to the CoSy Robotics project (2004-2008)
- Work related to the project, since about 1962.

Related links

- Research in CS and AI at Birmingham
- Robotics at Birmingham
- The CoSy Project at Birmingham and CoSy papers
- The ATT-Meta project
- Image understanding and Computer vision at Birmingham
- Presentations on cognition and affect, architectures, philosophy, vision, AI, tools, etc. (PDF)

- The UKCRC Grand Challenge project on Architecture of Brain and Mind
- Audio interview on grand challenge (December 2004)
- The Free Poplog web site with tools, teaching materials and demonstrations.
- AISB'06 Symposium on Grand Challenge 5: Architecture of Brain and mind.
- IJCAI'05 Tutorial on 'Representation and learning in robots and animals'
  and its online booklet (to be expanded)
- The School of Computer Science
- Projects funded in the European Commission's 6th Framework Cognitive systems initiative
- Marvin Minsky's web page
  Including a draft version of his 2006 book *The Emotion Machine*, which overlaps with much of the work done here, on architectures, forms of representation, philosophy of AI and varieties of affect, but focuses on a single architecture rather than the space of architectures.
- John McCarthy's Web Page
- John McCarthy -- Some Reminiscences
  Original version was in AISB Quarterly, January 2012.

The SimAgent AI toolkit

Our toolkit is available within the Birmingham Free Poplog Web directory with full system sources. For information about the toolkit, which is available see
http://www.cs.bham.ac.uk/research/poplog/packages/simagent.html

The toolkit is mostly implemented in Pop-11, which is part of Poplog, which used to be an expensive commercial product, but is also now available free of charge with full system sources, at
http://www.cs.bham.ac.uk/research/poplog/freepoplog.html

Symposium: How to Design a Functional Mind (at AISB 2000)
(The DAM -- 'Designing a Mind' -- symposium)

A symposium on ``How to Design a Functional Mind'' was held at
the AISB'00 Convention at the University of Birmingham 17-20 April 2000.

Information about the symposium, including abstracts and full papers can
be found here http://www.cs.bham.ac.uk/research/projects/cogaff/dam00

A book of papers related to the workshop edited by Darryl Davis was published in 2004
Visions of Mind: Architectures for Cognition and Affect.
IGI Publishing
(I won't publish with IGI because I object to their copyright requirements.)

A Tribute to Max Clowes,

one of the pioneers of AI in the UK, who died in 1981.
His ideas played an important role in the early development of this work.

PhD SUPERVISION

For information on how to apply to be a PhD or MSc student
(in Computer Science, Software Engineering, AI, or Cognitive Science)
in this School, see the School's study opportunities web page.

Please read that file information BEFORE writing to individuals asking for advice or information.
Please note: I do not deal with student admissions.
Being retired, I am also no longer able to supervise PhD students.

See also the School of Computer Science Web page.
The Birmingham Centre for Computational Neuroscience and Cognitive Robotics.

Creative Commons License

This file is maintained by Aaron Sloman,
Email A.Sloman@cs.bham.ac.uk