Expanded version of slides presented at
Grand Challenges Conference Newcastle, March 29-31 2004

Revised for Edinburgh Seminar Nov 2004, EC Meeting Luxembourg Oct 2005

GRAND CHALLENGE 5

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

http://www.tech.plym.ac.uk/soc/staff/mikedenh/home.htm

and others involved in discussions of GC5

See: `http://www.ukcrc.org.uk/grand_challenges/`

These slides, and other information about GC5, can be found here:
`http://www.cs.bham.ac.uk/research/cogaff/gc/`

Background

- In 2002 Robin Milner and Tony Hoare proposed through the UKCRC that the UK computing research community should identify a number of long term "grand challenge" projects that would advance science.

- A number of grand challenge proposals emerged, summarised in a booklet available via the UKCRC web site: `http://www.ukcrc.org.uk/grand_challenges/`

- One of the proposals was Grand Challenge 5: Architecture of Brain and Mind.

- No funding has specifically been allocated to them (compare the EC, below) but the EPSRC has acknowledged the proposed Grand Challenges.

- Any proposals submitted (in 'responsive mode') will be judged on merit.

- Already grant proposals and job advertisements are referring to GC5 and science journalists have been enquiring about it. E.g. I was interviewed for Danish Radio. See `http://www.cs.bham.ac.uk/research/cogaff/audio/gc/`

- More detailed information about this proposal, and related web sites, can be found here: `http://www.cs.bham.ac.uk/research/cogaff/gc/`

THE GRAND CHALLENGE PROBLEM

Can we understand brains and minds of humans and other animals well enough to build convincing functional robot models?

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 – at different levels of abstraction.

- This is an enormously difficult long-term task, requiring cooperation between many disciplines, e.g.
  - Neuroscience
  - Psychology
  - Computer science
  - Software engineering
  - AI (including Robotics)
  - Linguistics
  - Social sciences
  - Biology
  - Ethology
  - Anthropology
  - Philosophy
  - Mathematics / Logic

CONJECTURE

We need a far better understanding of how natural intelligence works, at different levels of abstraction, if we are to build more intelligent (e.g. robust, autonomous, adaptive) artificial information-processing systems.

In particular, building a working robot requires us to develop an architecture integrating many types of functionality with much richer interactions than current AI systems (usually designed to work as self-contained mechanisms) allow.

Doing this will advance the study of biological systems and biological evolution, by producing new hypotheses and new, deep, empirical questions.

This is a computing grand challenge

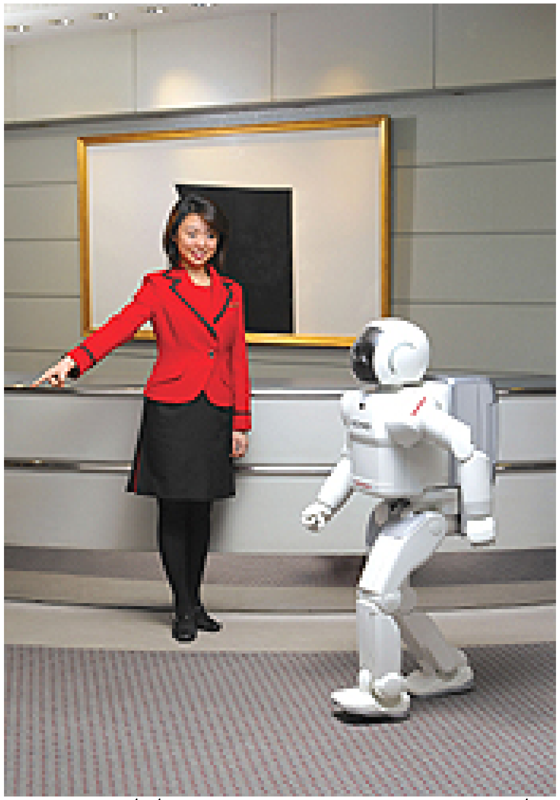
Some people argue that explaining how humans and other animals work is a problem for biology, neuroscience and psychology, not computer science, e.g. because brains are not computers

But all organisms, including humans, are information-processing systems and there is no other discipline that has tools, techniques and theories for modelling and explaining a wide range of information-processing capabilities, especially in virtual machines.

So we can't just leave this to other disciplines,
e.g. biology, neuroscience, psychology —
but we must learn from and cooperate with them.

Is work on robots already doing this?

THE STATE OF THE ART IN 2002



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

http://www.aibo.com/

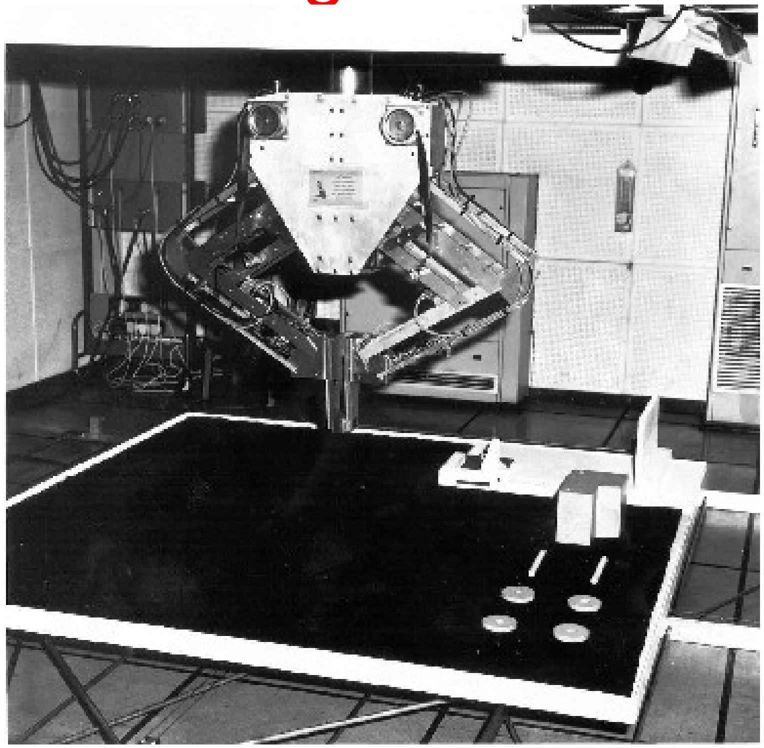
Despite very impressive engineering, present day robots look incompetent if given a task that is even slightly different from what they have been programmed to do – unlike a child or crow or squirrel.

Mostly they have (less than) insect-like purely reactive behaviours, lacking the deliberative ability to wonder 'what would happen if...'.

They also lack self-knowledge or self-understanding, e.g. about their limitations, or why they do things as they do, or why they don't do something else, or what they cannot do.

Compare Freddy the 1973 Edinburgh Robot

Some people might say that apart from 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 II 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: partly because visual processing was extremely slow.

Imagine using a computer with 128Kbytes RAM for a robot now.

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 (including humans) with the eyes of (software) engineers.

Freddy's 'heap smasher'

If the parts to be assembled were jumbled in a heap so that Freddy could not recognize them, then it would look for something long and thin protruding from the heap, grasp it, and then move it left and right.

That would typically cause the objects to be separated so that they could then be recognized by their outlines, and the rest of the task completed.

This idea was developed and implemented by Chris Brown during a visit to Edinburgh around 1973 – it was not invented by Freddy.

In fact Freddy had no idea why it was looking for the protruding object. It had no understanding of what moving the object would cause the items in the heap to be rearranged so that their outlines could be seen and recognized.

Freddy used affordances that Chris Brown had perceived and understood, but Freddy neither noticed nor understood them: it merely had an innate disposition to move into heap-smashing mode when it failed to recognize objects.

Like a precocial animal that is very competent but only because evolution has pre-programmed it to be competent — e.g. infant deer running with the herd.

For more on the altricial-precocial spectrum (for robots) see

`http://www.cs.bham.ac.uk/research/cogaff/05.html#200502`
`http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0609`

Having capabilities
vs
understanding what you are doing

One of the research challenges is to explain the difference between merely having a collection of capabilities and which are successfully deployed to fulfil biological or other functions and

UNDERSTANDING WHAT YOU ARE DOING

One way to clarify this is to compare many different cases, analysing their similarities and differences, including evolutionary and developmental sequences.

Is understanding shown by merely being able to do deliberative reasoning: e.g. being able to consider multi-stage branching alternatives, compare their costs and benefits and choose one: i.e. having deliberative mechanisms?

Consider various kinds of meta-level capabilities including being able to compare two cases of deliberative reasoning and notice costs and benefits of thinking in different ways.

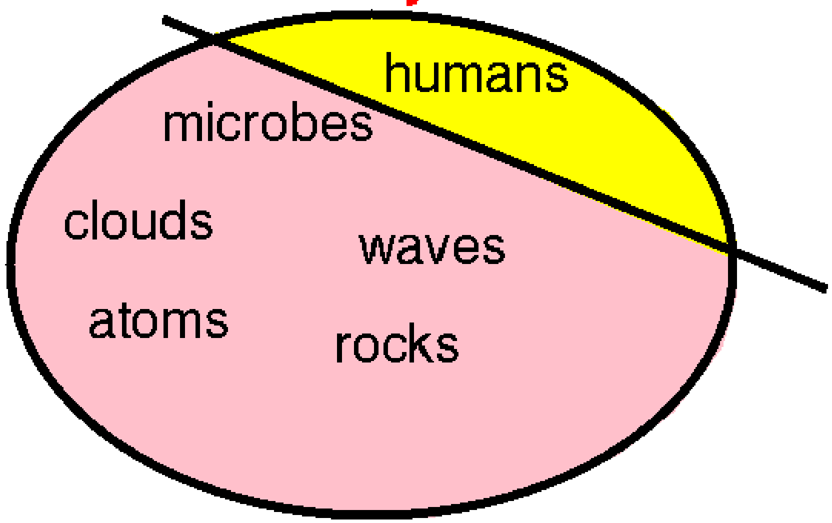
(At what stage does a child start to be able to do that?)

Don’t look for simple dichotomies

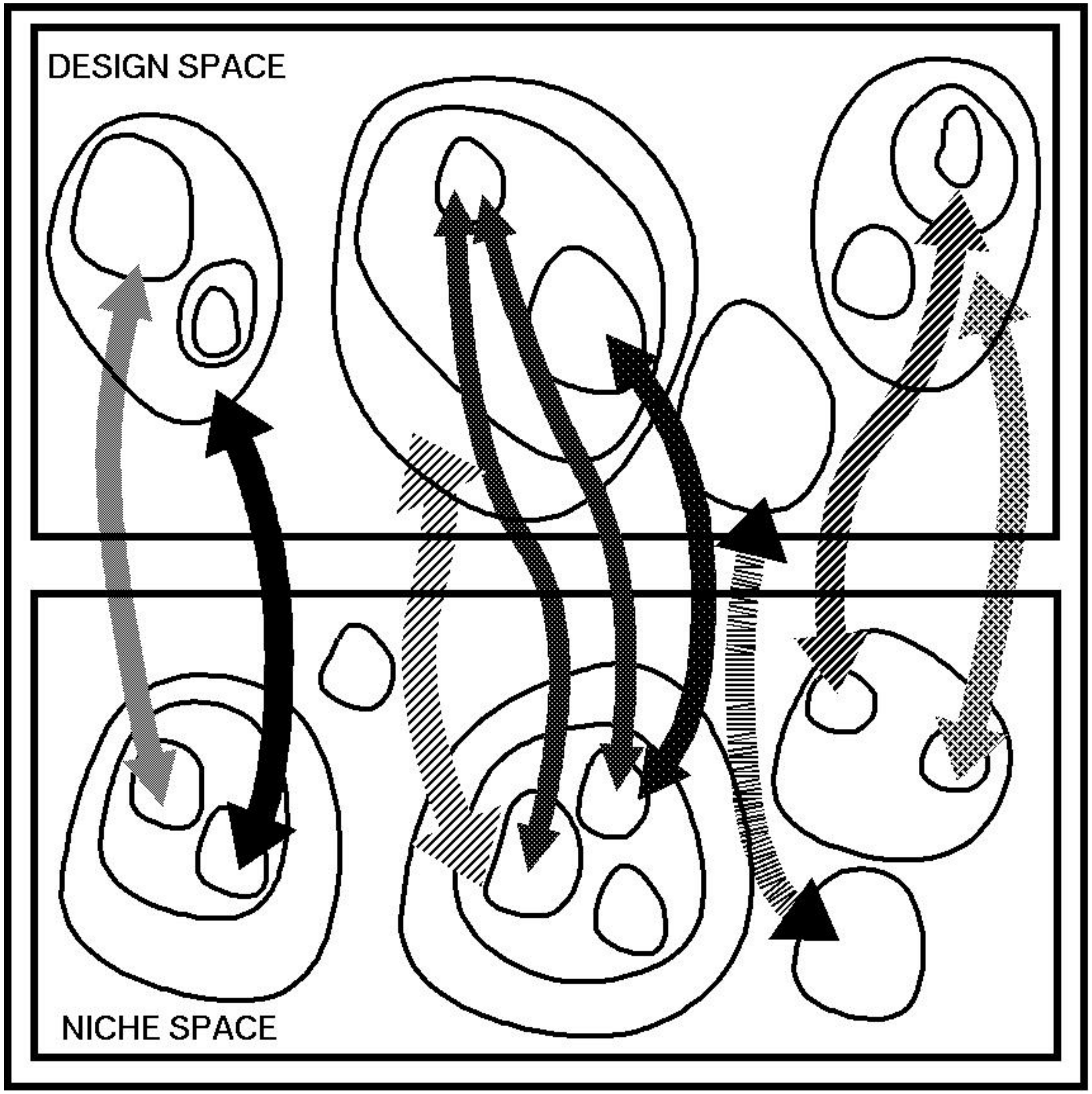
Some tempting wrong ways to think about consciousness:

1. There's no continuum from non-conscious to fully conscious beings

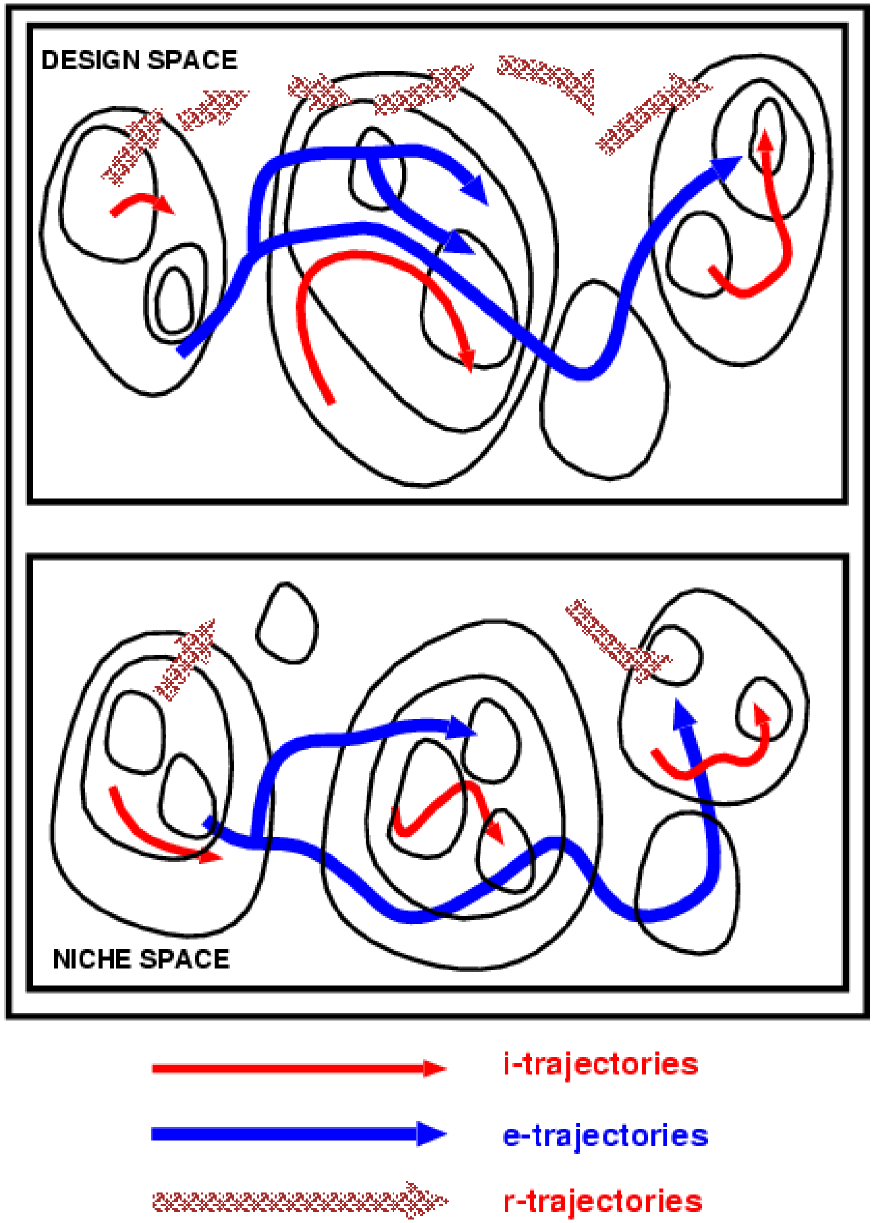
2. It's not a dichotomy either



Design space and niche space



Trajectories in design space and niche space



Science or Engineering?

This is primarily a scientific challenge,
not an applications challenge aimed at
producing some useful new machines.

But the research has two aspects,
theoretical and practical,
which inform each other.

POTENTIALLY THERE ARE MANY APPLICATIONS –
BUT THEY ARE NOT THE MAIN MOTIVATION.

The engineering goal of getting a machine to play chess as well as the best human players has been achieved, but not the scientific goal of clarifying requirements and designs for a machine that understands what it is doing when it plays chess, and can describe its strategy, explain things to a weaker player, etc.

Background to challenge No 5

There is a very long history to the 'Architecture of Brain and Mind' challenge, with three aspects

- Theoretical

  Humans have been attempting to understand what we are, and how we fit into the natural world, for many centuries, perhaps millenia.

  This curiosity has driven research in philosophy and more recently the various sciences that study man: psychology, neuroscience, biology, linguistics, sociology, anthropology, ...

- Practical

  For many centuries attempts have been made to produce machines with various subsets of human characteristics, some using mystical and religious inspiration (e.g. stories about golems), some based on engineering design (e.g. calculating machines, game-playing machines, automatic looms, etc.)

- Imaginative

  Many stories, myths, films, ...

Past Obstacles to Progress

Both the theoretical and the practical goals are still far from being achieved. Progress has been impeded by:

- Limitations of available technology

  E.g. speed, storage, kinds of design tools, programming languages, inadequate ways of thinking about complex systems that are strongly coupled with complex, ill-understood, evolving, environments, ...

- The extreme complexity of the mechanisms and the limits of our theoretical grasp of the space of possible mechanisms and what they can do.
- Over-simple theories of learning (e.g. ignoring learning about thinking)
- Mistaken views of the functions of language

  (e.g. not taking account of the role of language in thinking).

- Inadequate theories of motivation in children,
- Lack of understanding of the problems: what needs to be explained, or replicated, e.g.
  - mistaken theories of the functions of vision (e.g. thinking it is about segmentation, object recognition, or depth estimation, or ....
    ignoring affordances[*] and the uses of vision.)

  [*] Seeing affordances is seeing opportunities and obstacles relating to one's goals and capabilities – it involves seeing which actions are and which are not possible, and why.

Perhaps we have now learnt the difference between blind and shallow optimism and disciplined long-term ambition?

We are able to make faster progress now:

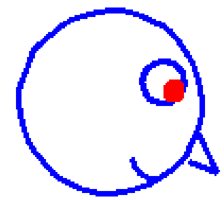
- Limitations of available technology have been reduced, as a result of:
  - Advances in electronic computing (speed, memory, clusters),
  - New languages and tools
  - New advances in materials science, mechanical and electronic engineering,
  - Perhaps nanotechnology and other things before long???
- The extreme complexity of the mechanisms and the limits of our theoretical grasp of the space of possible mechanisms and what they can do remain, but new theoretical understanding of possible mechanisms comes from
  - New thinking about architectures
  - Neuroscience
  - Computer science and AI
  - Dynamical systems theory
  - Physics???
- Enriched understanding of the problems: what needs to be explained, or replicated: In the last half century, progress in characterising the problems has come from
  - Huge amounts of empirical research in many areas of psychology, ethology, linguistics, and other sciences.
  - 50 years of attempts in AI to implement working systems and analysing the reasons for their failure or inadequacies

How should the research be done?

Many (though not all!) agree that because of recent advances in science and technology, and expected new advances in the next 20 years, the time is ripe for a new initiative.

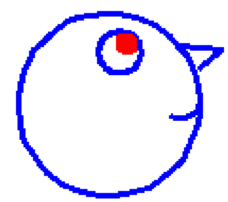
But there are different views as to how to proceed

- Some prefer implementation-neutral “top-down” research strategies, attempting to explain known kinds of human and animal competence using any low-level mechanisms that work, whether biologically plausible or not.



Top down

- Some prefer to work “bottom” up from brain mechanisms (e.g. chemical and neural) along with physical properties of bodies.



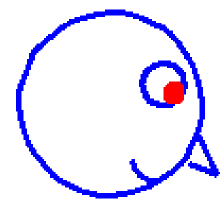
Bottom up

How should the research be done?

Many (though not all!) agree that because of recent advances in science and technology, and expected new advances in the next 20 years, the time is ripe for a new initiative.

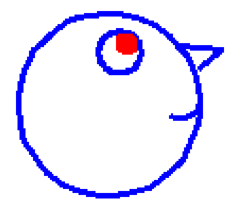
But there are different views as to how to proceed

- Some prefer implementation-neutral “top-down” research strategies, attempting to explain known kinds of human and animal competence using any low-level mechanisms that work, whether biologically plausible or not.



Top down

- Some prefer to work “bottom” up from brain mechanisms (e.g. chemical and neural) along with physical properties of bodies.



Bottom up

- This grand challenge proposal recommends
  – a combined bottom-up, top-down and middle-out strategy,
  – using requirements at different levels of abstraction in combination, to constrain the search for good designs.

Understanding what the problems are

Part of the scientific goal is to understand the requirements for integrated, multi-functional, human-like systems.

MANY EARLIER PROJECTS MISTAKENLY ASSUMED THE REQUIREMENTS WERE CLEAR.

BUT

- Biological systems are products of many different sorts of processes
  - Species evolution
  - Individual development
  - Cultural evolution
  - Interactions between species
    - of similar scales
    - of very different scales
    - with very different relations of competition, coexistence, symbiosis
  - Non-biological processes (geological, meteorological, planetary)
- Operating over different time-scales
- Subject many different sorts of constraints
- With causation over very different time-scales (within actions, within the life of an individual, over generations, and beyond)

Ignorance of history, and constraints, may hamper understanding of their products – solutions to many different sorts of problems, at different levels of abstraction.

Integrating parallel approaches

Individual researchers can adopt different detailed research strategies, and different intermediate goals

provided that people using different strategies communicate with one another and share the same long term goals, namely

- identifying unsolved problems and constraints on their solution
- explaining integration of diverse kinds of functionality
- understanding biological implementations of the architectures and mechanisms
- demonstrating our new understanding by building working systems.
- feeding results, hypotheses, questions to other disciplines, and learning from them

PEOPLE WORKING ON THIS CHALLENGE MUST NOT USE IT AS AN EXCUSE TO DO WHAT THEY WERE GOING TO DO ANYWAY,

e.g.:

- speeding up existing algorithms,
- extending a working system to handle more cases of the same general sort
- extending their favourite architecture for some limited system.

Most researchers are already doing those things.

The need for integration I

Over the last half-century there has been much fragmentation, within each of: AI, psychology, neuroscience — most researchers focus only on a limited sub-field, e.g.

- vision (usually low-level vision nowadays)
- language (text, speech, sign-language)
- learning (many different kinds)
- problem solving
- planning
- mathematical reasoning
- motor control
- emotions

etc....

COMPARE
THE FORESIGHT
INITIATIVE

The need for integration II

There is no guarantee that a technique, or form of representation, or algorithm, etc. that works for an isolated task will also work when that task has to be integrated with many other kinds of functionality in an integrated system.

For example, in humans what you see can help you understand words you hear and vice versa — and the interactions can be asynchronous.

Can we take existing vision systems and language understanding systems and combine them to demonstrate this phenomenon?

The need for integration III

Our previous considerations show that merely combining many capabilities is not enough: an insect combines many capabilities.

Just giving it the ability to add linguistic descriptions, answer questions posed in language, would not necessarily add self-understanding of a kind that gradually grows in a human child and produces an intelligent human adult.

Integration needs to include combining different sorts of capabilities along with additional capabilities to notice, reflect on, reason about, and learn from the integration.

This has been called ‘reflection’ by some ‘meta-management’ by others.

Putting the pieces back together:

- This grand challenge aims to understand and model brains/minds as integrated robotic systems functioning at different levels of abstraction, including
  - Physiological properties of brain mechanisms (how many different sub-types are there?)
  - Neural information processing 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, ...
- EXAMPLE: allowing components to interact fruitfully at any time: asynchronous anytime incremental influences

  This requires new styles of programming!

  Show hybrid reactive deliberative demo.

  (Movie here: http://www.cs.bham.ac.uk/research/poplog/figs/simagent)

Finding requirements through deep analysis of tasks

Most 3-D scenes have multiple objects with multiple parts that stand in multiple relationships forming a richer web than we can express in language, though we can summarise salient aspects, e.g.

THE CAR MOVES INTO THE GARAGE

At every moment there are

- many parts of the car
- in multiple relations to one another
- many parts of the garage and things in the garage
- many relationships between parts of the car and parts of the garage (and other things in it),
- where parts and relationships exist at different levels of abstraction (physical, geometrical, topological, causal, functional, aesthetic, ...)

In short what we can see, and think about, and learn about, and talk about, are multi-strand relationships, and when multi-strand relationships change we get multi-strand processes.

How to understand requirements for integration

One good way (not the only one) to study integration of multiple kinds of functionality is to attempt to design

- a biologically-inspired, physically embodied robot
- combining different human-like capabilities,
- e.g. perceiving, acting, learning, and communicating
- in the real world, in real time
- driven and modulated by affective states and processes e.g. wanting, preferring, disliking, ...

This will require new theories about how humans and animals understand space, time, motion, actions in 3-D space, and causality.

Conjecture: that understanding, present in a typical 5 year old underlies many other kinds of human abilities, including mathematical reasoning abilities.

Attempting to produce such a robot, on the basis of a theory of how it works and the trade-offs between different designs and implementations, is a major feature of this challenge.

Later we mention alternative possible target demonstrations.

An old idea: progressive deepening

The tasks are VERY difficult.

To explore integration of different components
it may be necessary to simplify some of the components
temporarily.

Compare Joe Bates on “broad architectures”.
http://www-2.cs.cmu.edu/afs/cs.cmu.edu/project/oz/web/papers/sigart_2_4.ps

Progressive deepening can follow.

Architectural challenges

One requirement for progress is specification of a virtual machine architecture that can combine many known kinds of human capabilities, including

- evolutionarily very old reactive mechanisms
- newer deliberative mechanisms and
- biologically rare reflective, meta-management mechanisms with meta-semantic capabilities (the ability to represent processes in things that themselves represent other things, unlike rocks, trees, levers, wheels, blocks, ...).

Papers and presentations in the Cognition and Affect project provide more detailed analyses of these architectural features, illustrated on the next slide. See

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

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

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

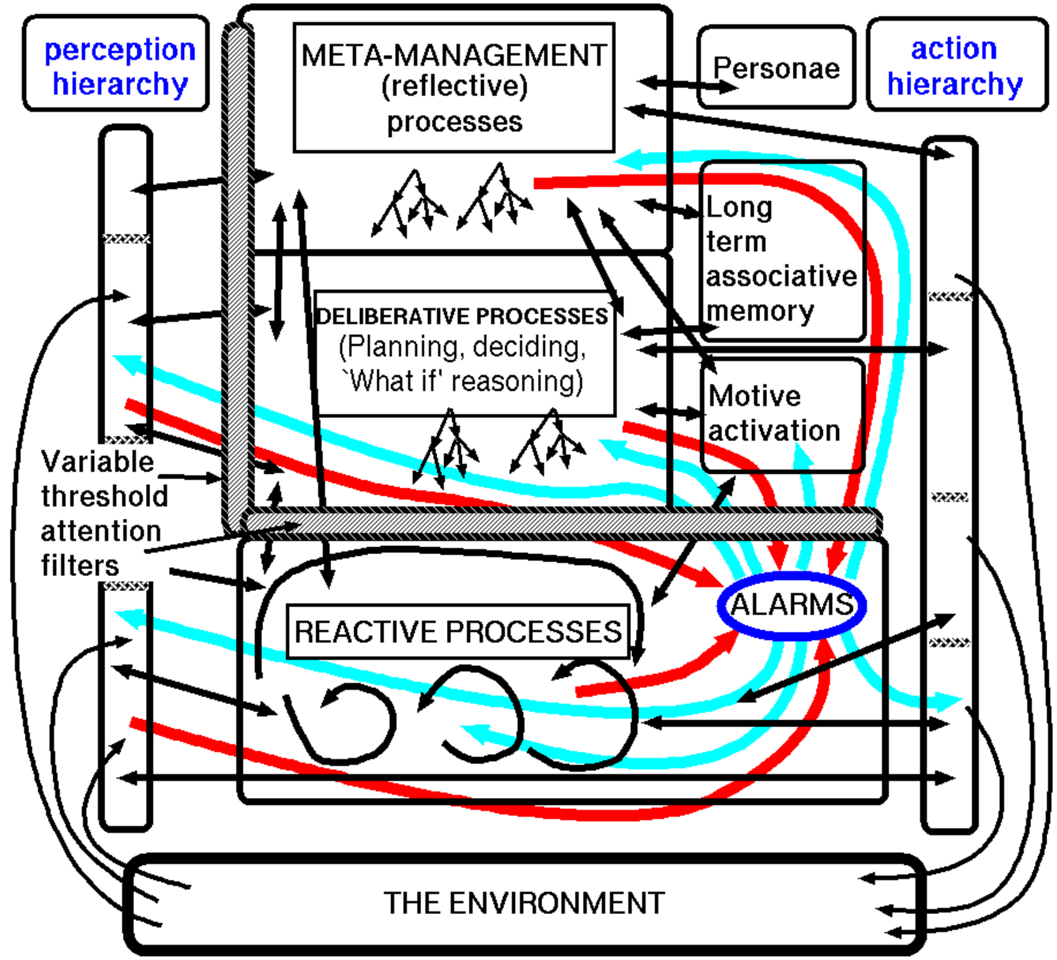
Many required sub-systems are not shown.

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...



(This is an illustration of some recent work on how to combine things: much work remains to be done. This partly overlaps with Minsky's *Emotion machine* architecture.)

But that's just one example

WE NEED LOTS MORE WORK ON

A TAXONOMY OF TYPES OF ARCHITECTURE,
REQUIREMENTS FOR ARCHITECTURES,
DESIGNS FOR ARCHITECTURES,

HOW ARCHITECTURES CAN DEVELOP,

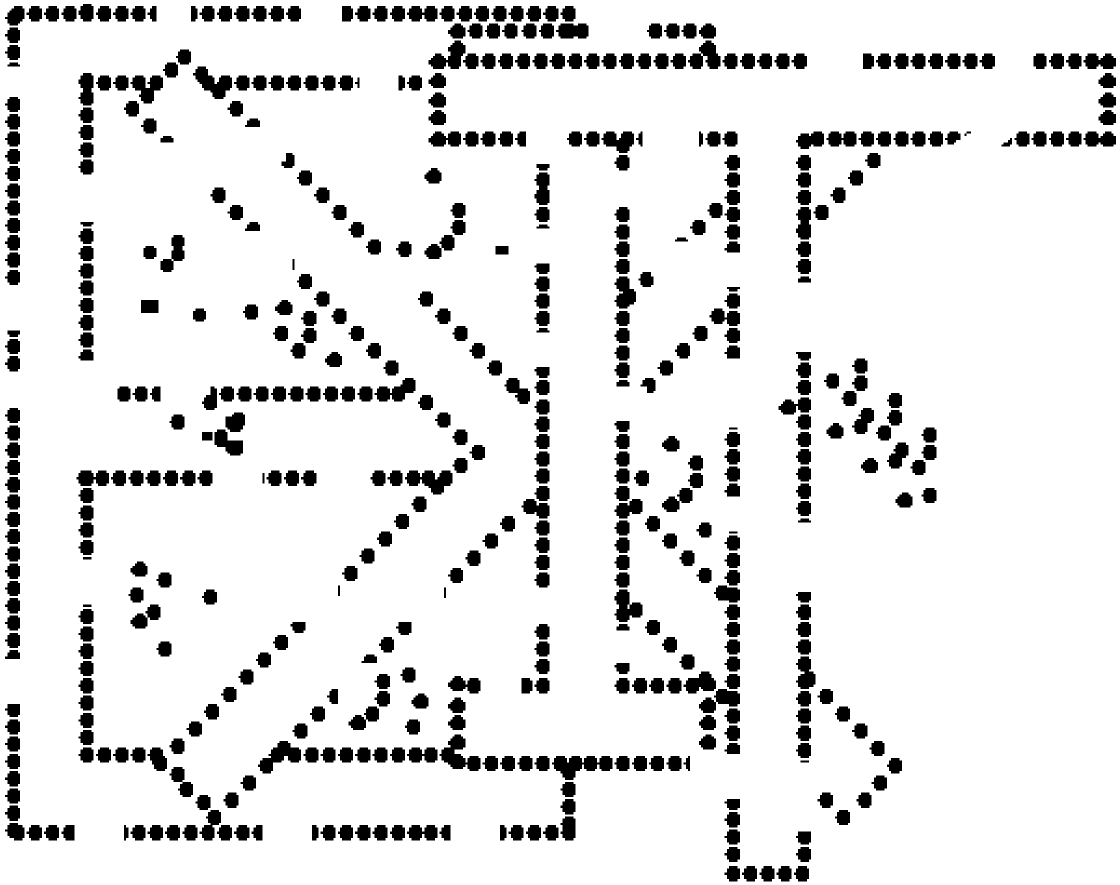
TOOLS FOR EXPLORING AND EXPERIMENTING WITH ARCHITECTURES

Vision is an example

Do we know what sort of architecture is required for human-like visual system?

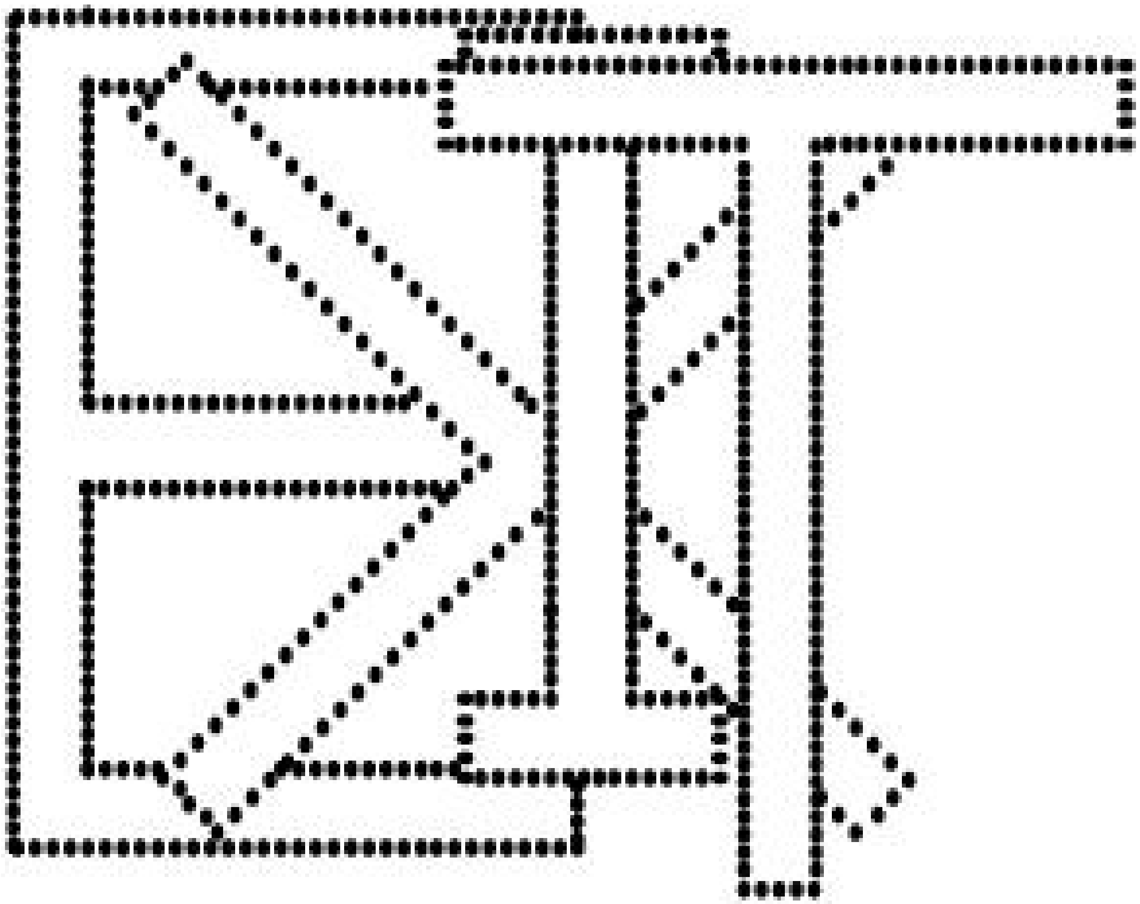
How quickly do you see the word
in the next slide??

Try to view it for less than a second.



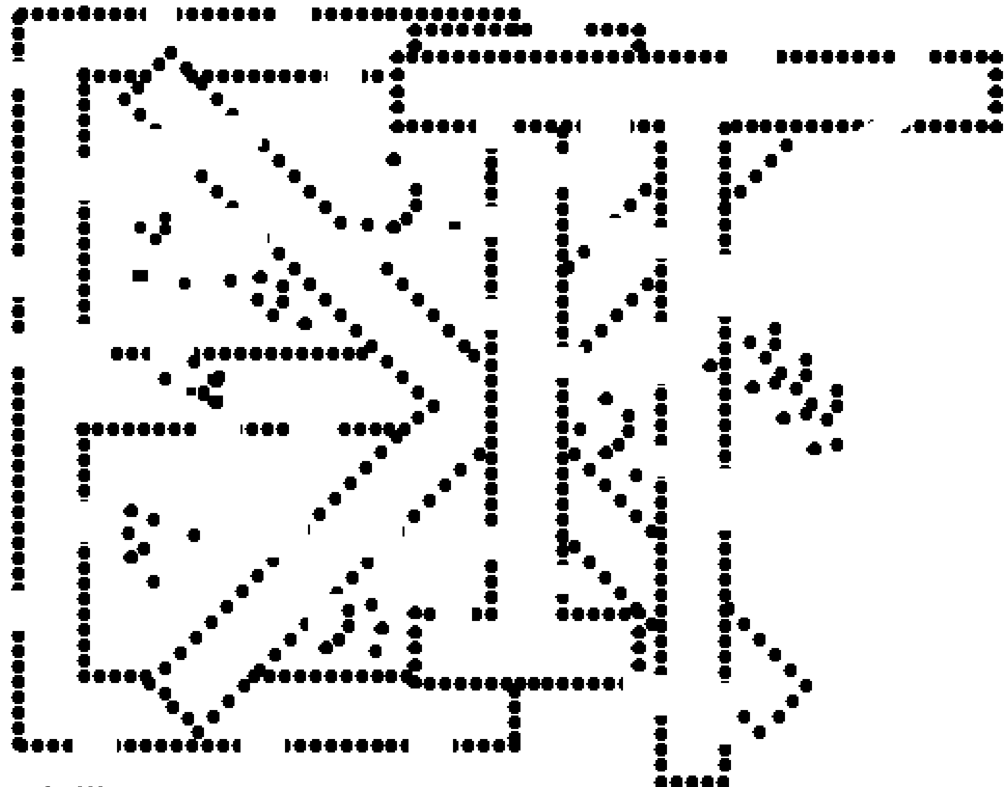
If you did not see a word, try looking back for about two seconds.

Did you see this?



Why do human-like systems need concurrent multi-level perception?

Answer: In order to cope with rapid recognition of high level structures in complex and messy scenes.



Despite all the clutter, most people see something familiar.
Some people recognize the whole before they see the parts.

Animal visual systems are not presented with neatly separated images of individual objects, but with cluttered scenes, containing complex objects of many sorts often with some obscuring others.

The objects may be moving, may be hard to see because of poor lighting, or fog, or viewed through shrubs, falling snow, etc.

How do we do it?

Real seeing is often much harder than the tasks most artificial vision systems can perform at present (or tasks presented in vision research laboratories)

Humans (and other animals?) are not always perfect, but they degrade gracefully.

A 30-year old idea may help.

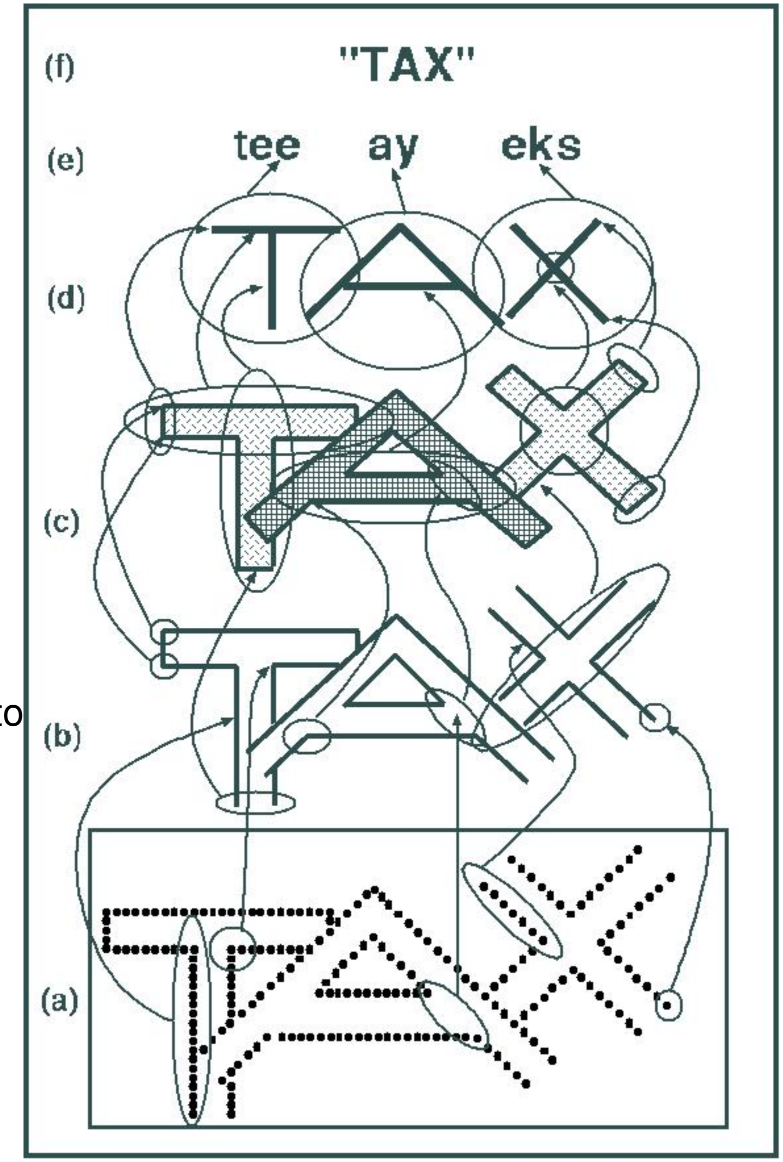
Multiple levels of structure perceived in parallel

Conjecture: Humans process different layers of interpretation in parallel.

Obvious for language. What about vision?

Concurrently processing bottom-up and top- down helps constrain search. There are several ontologies involved, with different classes of structures, and mappings between them.

- At the lowest level the ontology may include dots, dot clusters, relations between dots, relations between clusters. All larger structures are agglomerations of simpler structures.
- Higher levels are more abstract – besides grouping (agglomeration) there is also interpretation, i.e. mapping to a new ontology.
- Concurrent perception at different levels can constrain search dramatically (POPEYE 1978) (This could use a collection of neural nets.)
- Reading text would involve even more layers of abstraction: mapping to morphology, syntax, semantics, world knowledge



From *The Computer Revolution in Philosophy* (1978)
http://www.cs.bham.ac.uk/research/cogaff/crp/chap9.html

Perceiving structures vs perceiving combinations of features

It is important to understand the difference between

- Categorising
- Perceiving and understanding structure.

  You can see (at least some aspects of) the structure of an unfamiliar object that you do not recognise and cannot categorise: e.g. you probably cannot recognise or categorise this, though you see it clearly enough.

```
 Oooo
 Oooooo-------+
 OOOoooOOO    +
|oooOOOooo----+
+-------------+
```

What is seeing without recognising?

There's a huge amount of work on visual recognition and labelling e.g. statistical pattern recognition.

But does that tell us anything about perception of structure?
Much work on vision in AI does not get beyond categorisation.

There is something even more subtle and complex than perception of structure.

VIDEOS

The presentation in Newcastle started with two short videos:

- One showed Betty, the new caledonian crow, who surprised researchers at the Oxford University Zoology department when she displayed an ability to make a hook out of a straight piece of wire, in order to fish a bucket containing food out of a tube: (http://news.bbc.co.uk/1/hi/sci/tech/2178920.stm)

- The other showed an 18 month old child attempting to join two parts of a toy train by bringing two rings together instead of a ring and a hook, and showing frustration and puzzlement at his failure. (http://www.cs.bham.ac.uk/~axs/fig/josh34_0096.mpg)

  The crow seemed to have an understanding of hooks which the child lacked.

  A few weeks later the child was able to solve the problem: what had changed?

Perceiving structures vs perceiving affordances

Structures

things that exist, and have relationships, with parts that exist and have relationships

Affordances (positive and negative)

things that could or could not be made to exist by the agent, with particular consequences for the perceiver's goals, preferences, likes, dislikes, etc.:

modal, as opposed to categorical, types of perception.

- Betty looks at a piece of wire and (maybe??) sees the possibility of a hook, with a collection of intervening states and processes involving future possible actions by Betty.
- The child looks at two parts of a toy train remembers the possibility of joining them, but fails to see the precise affordances and is mystified and frustrated: presumably he sees parts and structural relationships because he can grasp and manipulate them in many ways. But he appears not to see some affordances.

How specialised are the innate mechanisms underlying the abilities to learn categories, perceive structures, understand affordances, especially structure-based affordances.

Millions of years of evolution were not wasted!

Affordances and empty space

What are affordances?

An example: How do you see empty space?

Here is a sample:

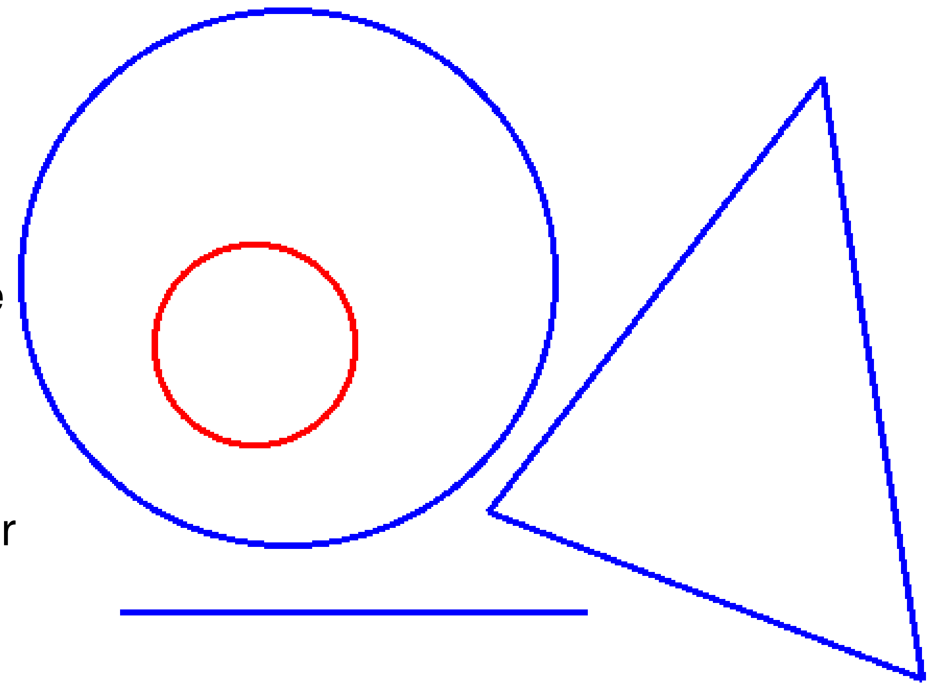
Empty space is a rich collection of possibilities.

(Possible objects, structures, processes, events.)

Affordances and empty space - examples

One way to see space: it affords motion: things that are in one place can move to other places, through (possibly empty) space. This could be important for an agent that moves things.

Consider figures in a Euclidean plane: there are empty spaces to which, and through which, those figures could move, with consequential changes in relationships, including creation of new points of contact or intersection.



- Where will intersections occur if you push the blue line vertically up the page?
- What events will occur if the blue circle moves to the right?
- How many blue circle/triangle intersection points can there be at any time? (Blue circle and blue triangle moving arbitrarily.)
- What changes if the blue circle can grow or shrink?
- If the red circle moves around without changing its size, what is the maximum number of intersection points between circle and triangle?
  (How do you know you have considered all possibilities?)

Visual reasoning in humans

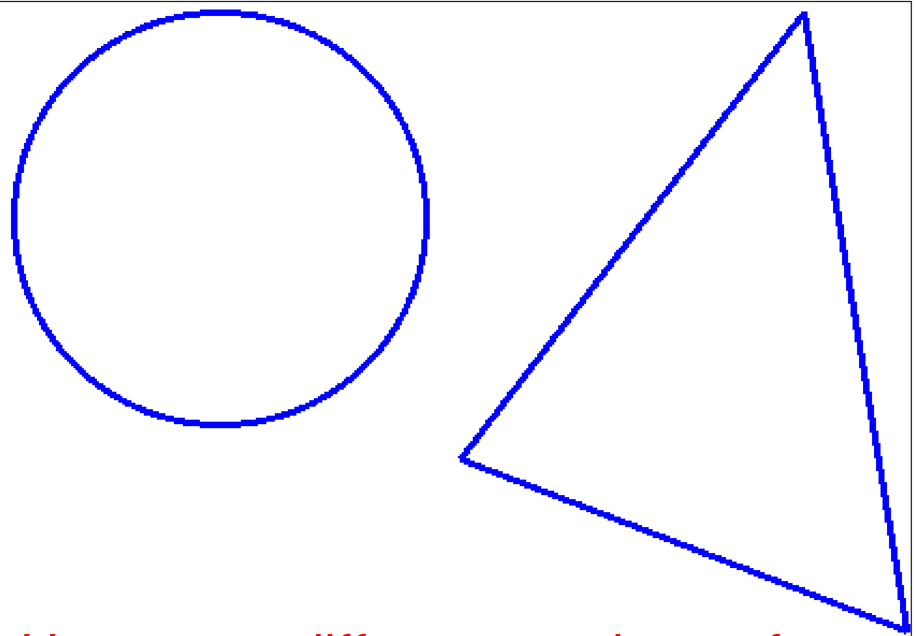
How do humans answer the question on the right?

Except for the minority who use logic, this requires the ability to see empty space as containing possible paths of motion, and fixed objects as things that can be moved, rotated and deformed.

Does it require *continuous* change?

Perhaps: but only in a virtual machine!
(To be discussed another time.)

Some people (e.g. Penrose) have argued that computers cannot possibly do human-like visual reasoning e.g. to find the answer 'seven' to the question.
(Compare Mateja Jamnik's PhD thesis)



How many different numbers of contact points can there be as circle and triangle slide around, shrink and grow?

Can we find a way to integrate visual perception, imagination, reasoning, problem-solving, language-understanding, action....?

Can we make an artificial mathematician?

Seeing empty space: potentially colliding cars

The two vehicles start moving towards one another at the same time.

The racing car on the left moves much faster than the truck on the right.

Whereabouts will they meet?

Where do you think a five year old will say they meet?

Five year old spatial reasoning

The two vehicles start moving towards one another at the same time.

The racing car on the left moves much faster than the truck on the right.

Whereabouts will they meet?

Where do you think a five year old will say they meet?

One five year old answered by pointing to a location near ‘b’

Me: Why?

Child: It’s going faster so it will get there sooner.

What is missing?

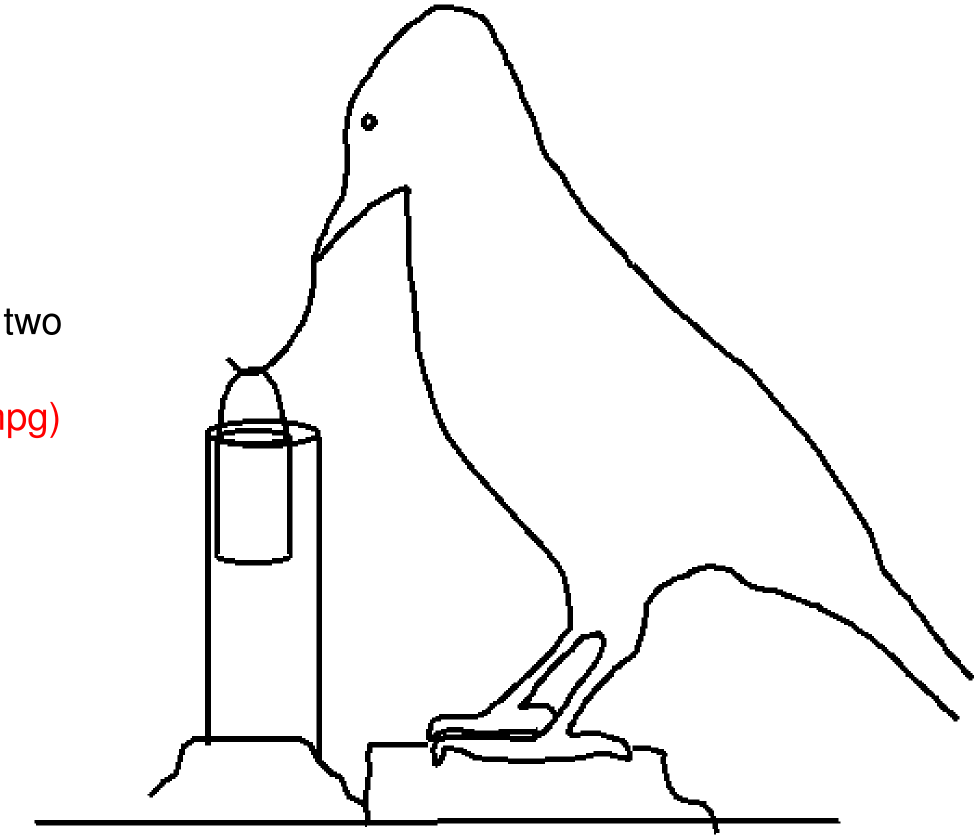
- Knowledge?
- Appropriate representations?
- Procedures?
- Appropriate control mechanisms in the architecture? ? ?

Betty

Betty the hook-making crow.

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

Contrast the 18 month old child attempting to join two parts of a toy train by bringing two rings together
(http://www.cs.bham.ac.uk/~axs/fig/josh34_0096.mpg)

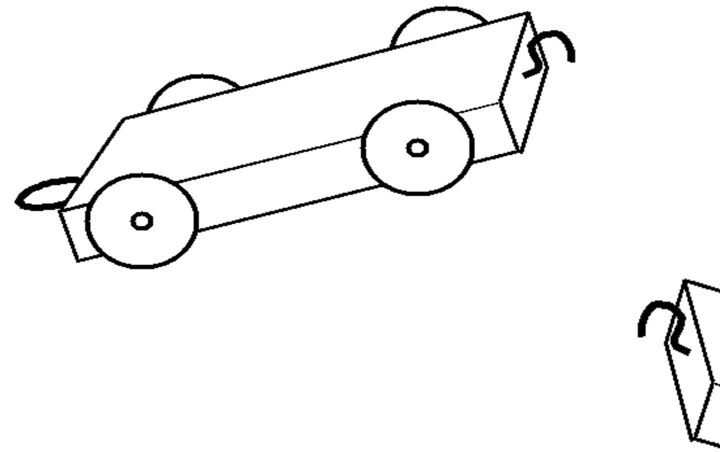


Does Betty see the possibility of making a hook before she makes it?

She seems to. How?

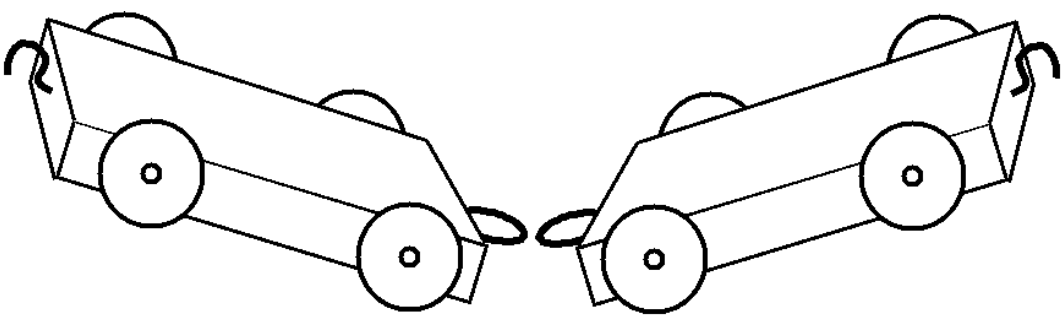
Affordances required for building trains

Perceiving affordances requires high levels of abstraction and representation of what might happen or be done.



How might you move the trucks on the right to join them together?

What capabilities are required in order to see why the following will not work?



How might the perceived affordances be represented – in a computer, or in a brain?

What changes between a child not understanding and understanding?

See the video of a child aged about 18 months failing to understand the affordances in two rings as opposed to a ring and a hook:

http://www.cs.bham.ac.uk/~axs/fig/josh34_0096.mpg

A few weeks later, he seemed to understand. WHAT CHANGED?

Could talking to a child speed up the change? How?

What sort of integrated robot is a good research target?

Different sorts of demonstrator targets have been proposed for this grand challenge: they have different pros and cons, and can be pursued in parallel. Understanding the pros and cons of the different options is more useful than arguing over which sort of target is “best” – e.g.

1. Model a (simplified) new-born human infant with general human neonate learning capabilities (if they can be discovered), and then train it.
2. Design and implement an adult-like intelligent personal assistant (DARPA cognitive systems initiative)
3. Attempt to replicate a significant subset of capabilities of typical human children aged somewhere between 2 and 5 years:

PEOPLE PURSUING THE DIFFERENT TARGETS MUST TALK TO ONE ANOTHER, AND LEARN FROM ONE ANOTHER

Option 3 could be approached via option 1 or directly.

More subtly, option 1 could follow option 3 – after understanding a later stage work on the earlier system that produces it.

Opinions differ as to which would be more likely to succeed.

Both could be tried in parallel.

For this project we do not recommend option 2.

What sort of robot?

Some comments on the different proposed targets.

- Model a new-born human infant with general human neonate learning capabilities, and then train it.
  - Several recent projects are like this (MIT COG, Grand's Lucy, Michigan project, de Garis).
  - It is very hard to determine a human neonate's capabilities: they are highly inscrutable.
  - It does not seem likely that totally general learning mechanisms suffice: evolution did a lot of prior work — this approach will have to make guesses about innate mechanisms.

- Design and implement an adult-like intelligent personal assistant (DARPA)
  - Adult human expertise is a very complex mixture of general (biological) mechanisms and culture-specific information, combined with vast amounts of idiosyncratic individual history. Any working system is likely to include much that is ad-hoc (less science more application-oriented work).

- Attempt to replicate a significant subset of capabilities of typical human children aged about 2-4 years:
  - We cannot replicate anything as sophisticated as a child: much simplification is needed
  - There is a huge amount of information available from developmental psychology
  - Such children engage in complex structured tasks, including linguistic tasks, with varying success, so that behaviour gives a fairly rich "window" into their competence
(what it is, not how it works)
  - The ratio of generic biologically determined capabilities to idiosyncratic and cultural information is still high: such children can be transplanted to a wide range of cultures and learning environments, and will develop and fit in. (Greater generality, Less ad-hocery?)
  - Information about these stages may help us understand requirements for neonates.

Analysing requirements — 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 chair to fetch a fragile object off a shelf.
- Communicate with other intelligent systems
  - Asking questions and giving answers
  - Requesting or providing explanations, and using them
  - Giving advice or help on how to do something
  - Warning someone who is about to sit on or climb onto a broken chair.
  - Reporting what it did last week.
- Explain how a pendulum clock works.
- Explain what a clock is for.
- Become more skilled at catching a ball, and handing things to people.
- Tie shoelaces.
- Reason about geometrical relations.
- Learn to count, then later use that ability in many tasks,
  as discussed in http://www.cs.bham.ac.uk/research/cogaff/crp/chap8.html
- Learn to think about numbers, then later about infinite sets.
- Feel fear, pity, shame, pride, jealousy, ... enjoy dancing, music, painting, poetry.
- Care about how another feels.
- Help a less able person with everyday tasks?

  E.g. a younger child, or someone physically infirm, or blind, or deaf, or ...

Milestones can be based on task analysis

Here is one way to identify milestones/intermediate targets for the 'top-down' robotic project (i.e. initial exploration unconstrained by neural or other implementation requirements)

- The tasks listed above, and many more, can be analysed in depth and groups of competences identified, and related to design principles.
- Progressive and demanding milestones can be identified by specifying progressively more demanding sets of scenarios and meta-scenarios
(the approach of the EC CoSy project: http://www.cs.bham.ac.uk/research/cogaff/),
- Meta-scenarios demonstrate the robot's ability to reflect and talk about what it does and does not do within scenarios, including explaining why something works, and offering another agent sensible advice and help.
- Negative scenarios help to specify what will not be attempted.

  E.g. can the robot understand and explain why order does not matter when you count things?

Some illustrations of this approach can be found here:

http://www.cs.bham.ac.uk/research/cogaff/gc/targets.html

Applications for a child-like robot.

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.

Example applications include
guide-robots for the blind,
more flexible assembly robots,
entertainment robots (and synthetic agents in virtual worlds),
search and rescue robots,
home assistants for elderly or disabled people
who don't want to be a burden on other people
(a growing subset of the population).

Applications are not the main point!

The main goal is
advance of science
not
a new engineering achievement.

But it is not a goal with a well-defined endpoint, so there is a question of how milestones can be set up and progress demonstrated.

Tests for the bottom-up (brain-modelling) component

- It should produce an overview of varieties of neural and chemical mechanisms in brains and how their functional roles differ
  (e.g. the different types of neurons, different neurotransmitters and neuromodulators),
- It should explain how physical subdivisions in the brain are (or are not) related to functional sub-divisions of the sort that the 'top-down' investigations identify,
- It should explain not only low-level sensory processes (as much current research does) but also indicate ways in which more abstract perceptual, reasoning, planning, decision-making, functions might be implemented in brain mechanisms,
  e.g. perception of affordances, learning new concepts, acquiring new values,
- It should explain how the architecture develops over time in an individual.
- It should provide ideas about designs and mechanisms that reform neuroscience!
- It should show how biological mechanisms provide better implementations for kinds of functionality investigated in the 'top down' sub-project.
  (Compare Grand Challenge 7: 'non-classical' forms of computation.)

  All this is unlikely to be achieved for several decades at least, but we need to identify some intermediate, simplified milestones.

This is not a grand challenge with a well-defined endpoint

COMPARE THE GRAND CHALLENGE OF FINDING A CURE FOR CANCER: THERE ISN'T ONE BUT THERE ARE MANY MILESTONES, INCLUDING CURES FOR PARTICULAR CANCERS.

The task of producing a robot that has all the characteristics of a typical human child, including

- similar physical skills
- similar patterns of learning and development
- similar likes and emotional responses

would require so many advances in so many areas of science and engineering, including materials science and mechanical engineering, that it is unlikely to be achieved in the foreseeable future.

However, a combination of child-like abilities based on a deep understanding of how multiple human abilities can be integrated, might be achieved in 15-20 years by a judicious choice of simplifications of human competences.

If such robots can be given useful applications as helpers for blind, or disabled people, as described above, that will be evidence that the simplifications were not a form of 'cheating' to make the goals achievable.

A MAJOR PROBLEM

Most of the best known, most able, most experienced AI researchers in the UK have so far shown no interest in this challenge.

There are probably several explanations

- Announcements about the GC proposals may not have reached everyone relevant.
- Pressure of work, including existing demanding projects already funded.
- No guarantee of any funding for this area.
- Time spent in many consortia submitting proposals to the EU Framework 6 and 7 'Cognitive systems' initiative (many proposals included UK partners).
- Belief that this proposal is too ambitious, and fear of repeating previous failed predictions of AI promoters.

We need need to encourage some of them to take a closer look, and also develop links with psychologists, neuroscientists, linguists, philosophers, etc.

We can use closely related recently funded projects as a nucleus.

- Mike Denham's EPSRC-funded cluster project (involving several neuroscientists) see: www.xxx.xxx. (URL to be provided)
- Owen Holland's EPSRC 'adventure fund' project on machine consciousness.
- The CoSy project funded by the EC 'Cognitive Systems' initiative (http://www.cs.bham.ac.uk/research/projects/cosy/)
- Other EC 'cognitive systems' projects, people in the UK Foresight 'cognitive systems' initiative, researchers in the USA e.g. workshop in Washington in April on self-aware systems.

Should we try to get funding for a major international cross-disciplinary conference in this area during 2005?

A challenge for Computer Scientists

GC-5 raises many challenges for computer science:

- can these systems ever be modelled formally?
- can the requirements or the designs be expressed in available formalisms?
- if not can those formalisms be extended?
- Can such robot designs be reasoned about, proved to work, ... etc.
  (This will require explicit formal theories about the environments in which they operate, i.e. physical, psychological, social, legal, ...)

NOTE:

Some of the other grand challenge proposals (e.g. those concerned with ‘Memories for life’, ‘Ubiquitous computing’, ‘Dependable computing’) involve designing reliable new complex systems providing many kinds of functionality and services.

It is not always realised that if such systems are to interact with humans, designers will need to have good theories of how humans work, e.g. how they learn, what sorts of interactions are more or less cognitively demanding, stressful, etc.

So progress in Grand Challenge 5 may be a prerequisite for success in others.
See `http://www.ukcrc.org.uk/grand_challenges/`

The EC Frameworks 6 and 7 'Cognitive Systems' initiatives have significant overlap with GC5

June 2003 document: IST Work Programme 2003/2004 states:

- Objective: To construct physically instantiated or embodied systems that can perceive, understand ... and interact with their environment, and evolve in order to achieve human-like performance in activities requiring context-(situation and task) specific knowledge.
- Focus is on: methodologies and construction of robust and adaptive cognitive systems integrating perception, reasoning, representation and learning, that are capable of interpretation, physical interaction and communication in real-world environments for the purpose of performing goal-directed tasks.
- ....emphasis is on systems and on interrelation between functions and not component methods for specialised tasks...

This does not require implementation in biologically plausible mechanisms. Nevertheless projects funded contribute substantially to the goals of GC5.

First tranche of funds: 25M Euro, over 4 years

A new call went out with about 50M Euro available, deadline March 2005.
See Colette Maloney's October 2004 presentation here:

`http://www.cordis.lu/ist/directorate_e/cognition/projects.htm`

Since then there have been several more calls, with large amounts of many allocated.

FP6 on applications

The EU calls emphasise the need to expand science rather than to produce new useful applications.

ROLE OF APPLICATIONS, ACCORDING TO FP6 CALL

Not about application development!

Applications serve to:

- provide research questions
- demonstrate impact of conceptual/technical innovation

(Presentation by Colette Maloney, Brussels June 2003.
Contrast DARPA Cognitive Systems initiative.)

The projects funded by the EU Cognitive Systems initiative provide a substantial part of the research required for the grand challenge on ‘Architecture of Brain and Mind’.

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/child.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.

Demos used the SimAgent Toolkit
http://www.cs.bham.ac.uk/research/poplog/packages/simagent.html

For more on this Grand Challenge: `http://www.cs.bham.ac.uk/research/cogaff/gc/`

Early press reaction:

`http://www.timesonline.co.uk/printFriendly/0,,1-50-811189,00.html`