BMVA MEETING

Reverse Engineering: the Human Vision System
Biologically inspired Computer Vision Approaches

Royal Statistical Society, 12 Errol Street, London

29 January 2003

Human vision — a multi-layered multi-functional system

Aaron Sloman

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

School of Computer Science
The University of Birmingham

Related papers can be found at
http://www.cs.bham.ac.uk/research/cogaff/

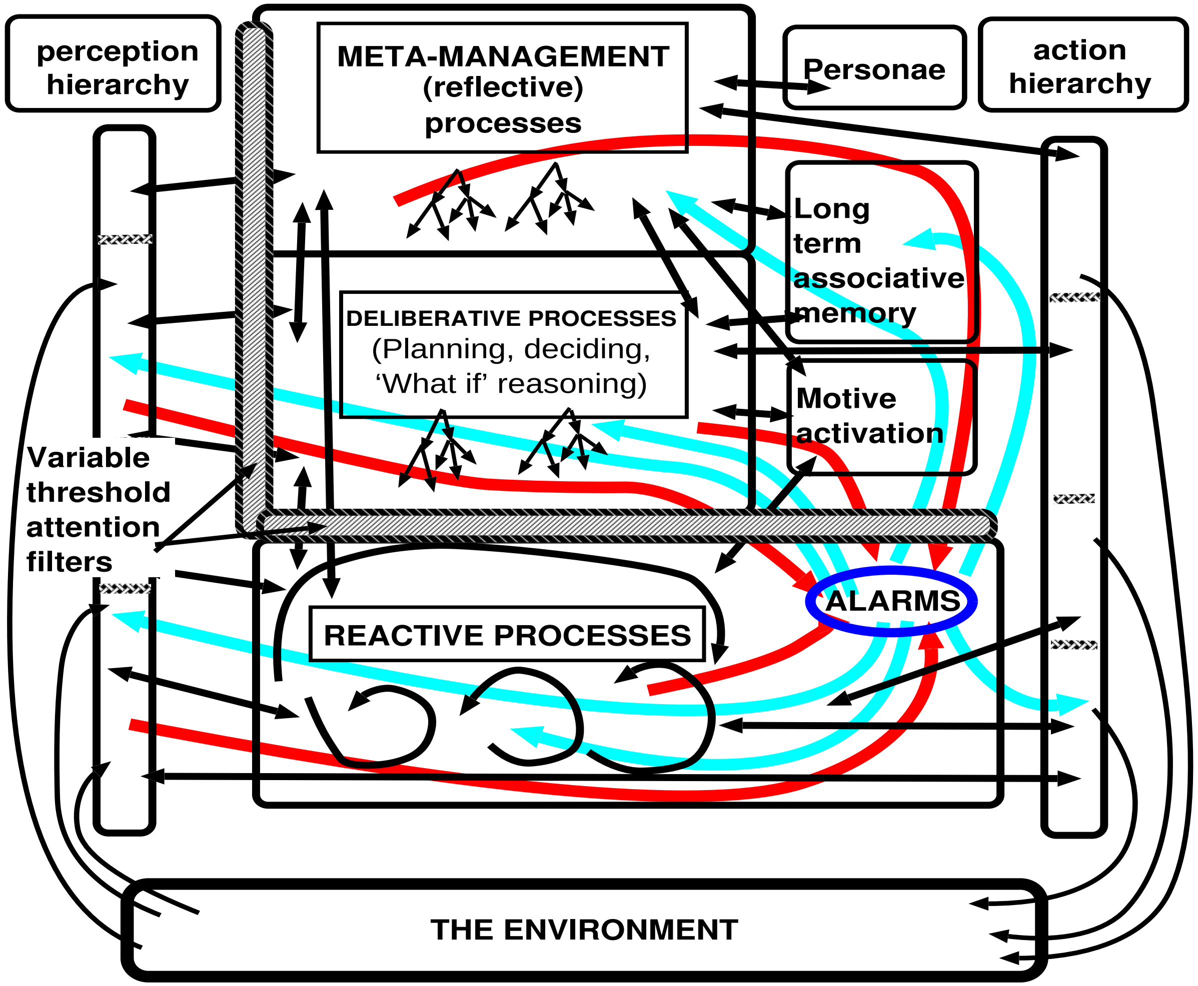
Related slide presentations can be found at
http://www.cs.bham.ac.uk/~axs/misc/talks/

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

The "Human-like" sub-schema H-Cogaff

Our conjectured architecture for human-like systems:

The H-Cogaff (Human-Cogaff) architecture is a special case of the CogAff architecture scheme/framework described in various papers and presentations



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

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

Requirements

- Cluttered scenes
- Variable amounts and kinds of occlusion
  by solid object, mist, fence, dirty glass, shrubbery
- Variable lighting
- Variable motion
- Many levels of structure
- Relevance to different kinds of expertise, different goals, different kinds of body parts
- Seeing possibilities for action and constraints on actions
  affordances
- Need for speed
- Need to learn
- Use for communication, recording information, reasoning
  Writing, pictures, diagrams, proofs
- Doing the same thing in your head
  mathematical thinking, planning

Has there been real progress in AI vision? The most demanding test is in robotics.

- AI researchers have tried for decades to produce intelligent machines (like those in movies).
- So far, despite many impressive achievements, they are nowhere near producing a robot with the visual and other capabilities of most animals.
- Current robots have general intelligence far below that of a young child, or even a squirrel, or crow, even though some of them may be very good at performing some narrowly circumscribed set of tasks.

For examples, see the Honda Asimo robot page
http://world.honda.com/news/2002/c021205.html
(Including some movies showing the robots doing things.)

the Sony AIBO page
http://www.aibo.com/

There's a lot more information at the AITOPICS web site
http://www.aaai.org/AITopics/html/robots.html

And the Birmingham University Robot page
http://www.cs.bham.ac.uk/research/robotics/cbbc/

Impressive robots made by Honda and Sony

THE STATE OF THE ART IN 2002

http://www.aibo.com/

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

Limitations of current robots

In many cases the engineering is very impressive.

But present day robots look incompetent if given a task that is even slightly different from what they have been programmed to do – unlike a child or chimp or squirrel.

Mostly they have purely reactive behaviours, lacking the deliberative ability to think or wonder 'what would happen if...'.

They also have very little self-knowledge or self-understanding, e.g. about their limitations.

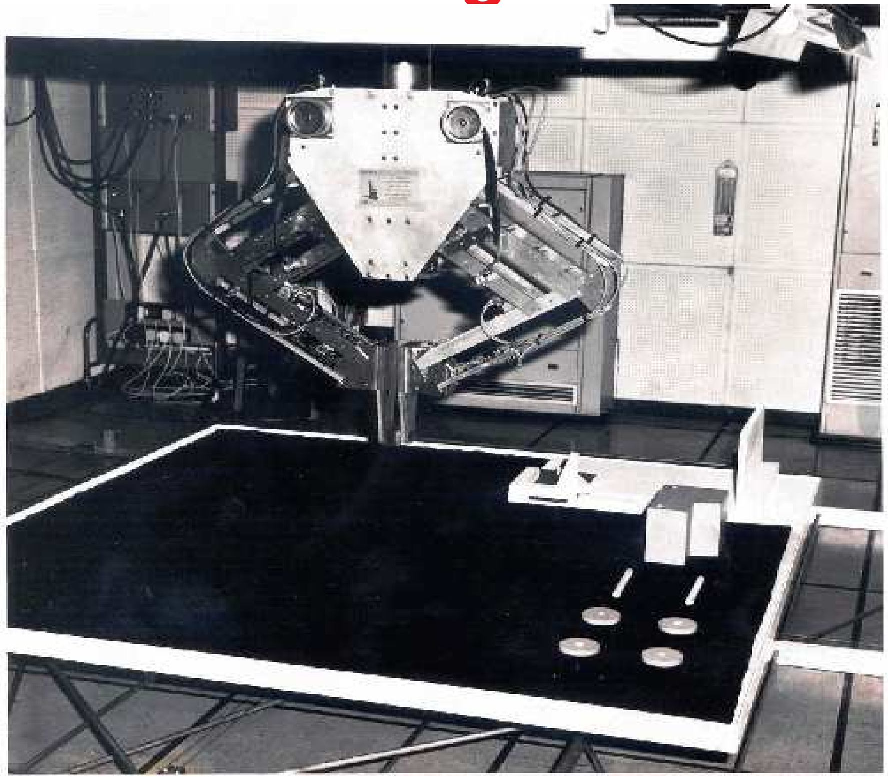
In particular, what they see is very limited: they lack the ontology to think about what we do, let alone perceive it, and they lack the perceptual architecture to provide the sort of information that ours does.

What information?

Do we know what information human vision provides?

Compare Freddy the 1973 Edinburgh Robot

Some people might say that apart from the wondrous advances in mechanical and electronic engineering there has been little increase in sophistication since the time of Freddy, the ‘scottish’ Robot, built in Edinburgh around 1972-3.



Freddy could assemble a toy car from the components shown on the table. They did not need to be laid out neatly as in the picture. However, Freddy had many limitations arising out of the technology of the time. E.g. Freddy could not simultaneously see and act.

There is more information on Freddy here

http://www.ipab.informatics.ed.ac.uk/IAS.html
http://www-robotics.cs.umass.edu/ pop/VAP.html

In order to understand the limitations of robots built so far we need to understand much better exactly what animals do: we have to look at animals with the eyes of software engineers, not psychologists.

Why are current robots still so limited?

1. Because animal intelligence has many features we don't understand yet.
2. Because mechanical engineering and materials science are still far behind what biological evolution has produced.
3. Because the computing power required to match animal brains, along with constraints of size, weight, energy consumption, etc. are so great.
4. *Above all because we don't yet know what needs to be done to replicate animal intelligence.*

   I.e. We don't yet have a understanding of what animal intelligence includes.

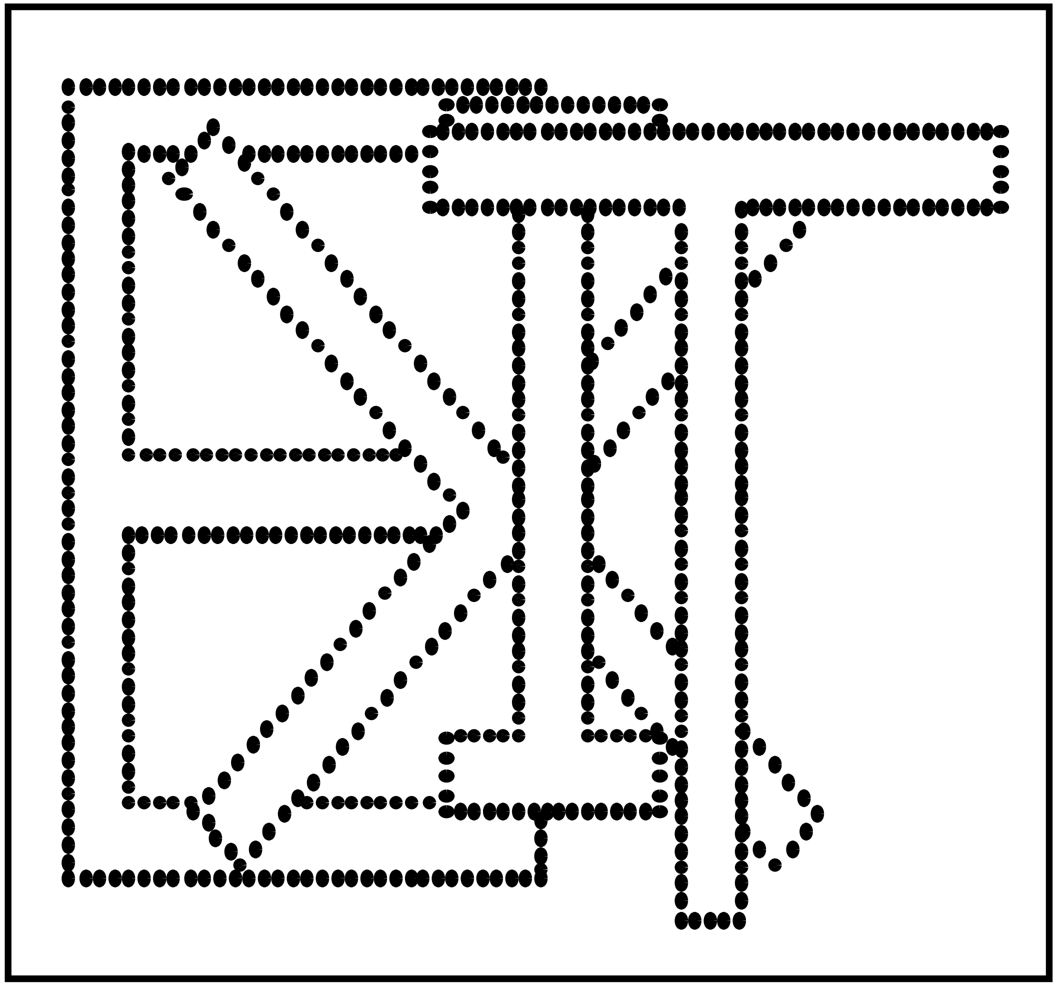
   In engineering terms: we don't know the functional requirements for the designs we are trying to produce – we don't really know what sort of software is required, let alone how to design and implement the software.
5. *This point is missed by people who proclaim that because computers are becoming much more powerful very rapidly we shall soon be able to produce super-intelligent machines.*

   That view assumes that we shall know how to use all the new power.

Some work done in the 1970s: POPEYE

The Popeye project (using Pop2 since Pop-11 did not exist when it started!) investigated how it is possible for humans to see structure in very cluttered scenes, where structure exists at different levels of abstraction.

Pictures used were like this, with varying degrees of clutter and with varying amounts of positive and negative noise.



Human performance degrades gracefully, and we often recognize the word before the individual letters have been recognized.

HOW?

See: *The Computer Revolution In Philosophy* (1978) Chapter 9

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

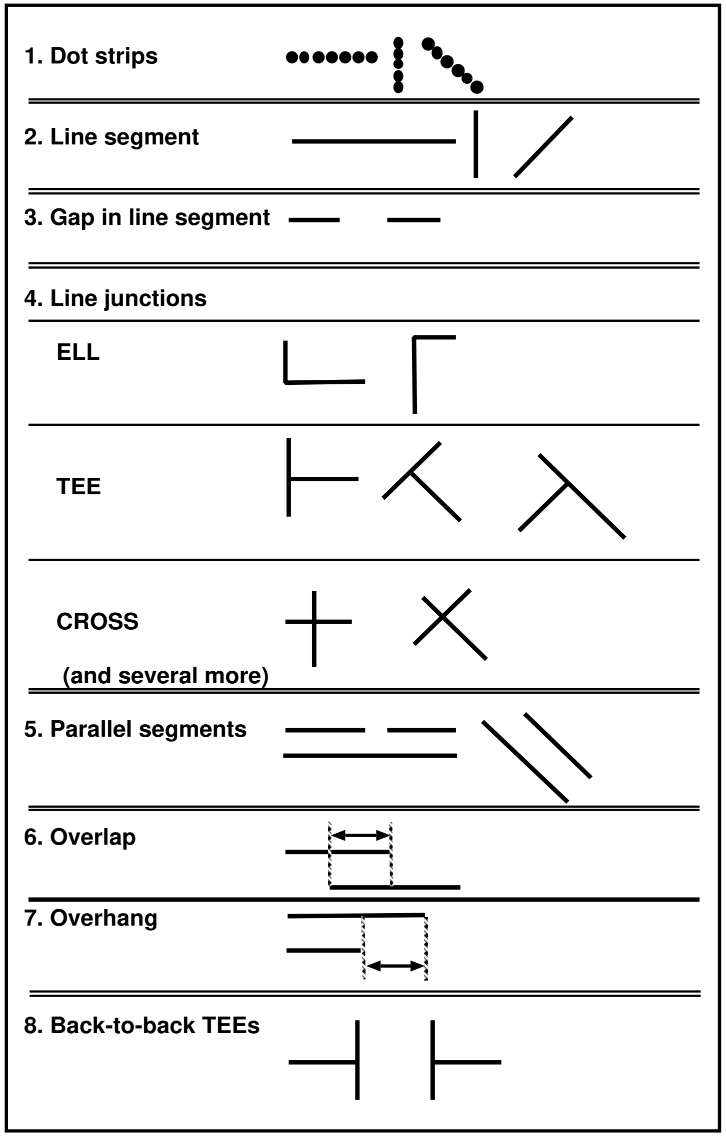
Conjecture:
Assemble fragments at different levels of abstraction

We seem to make use of structures of different sorts,

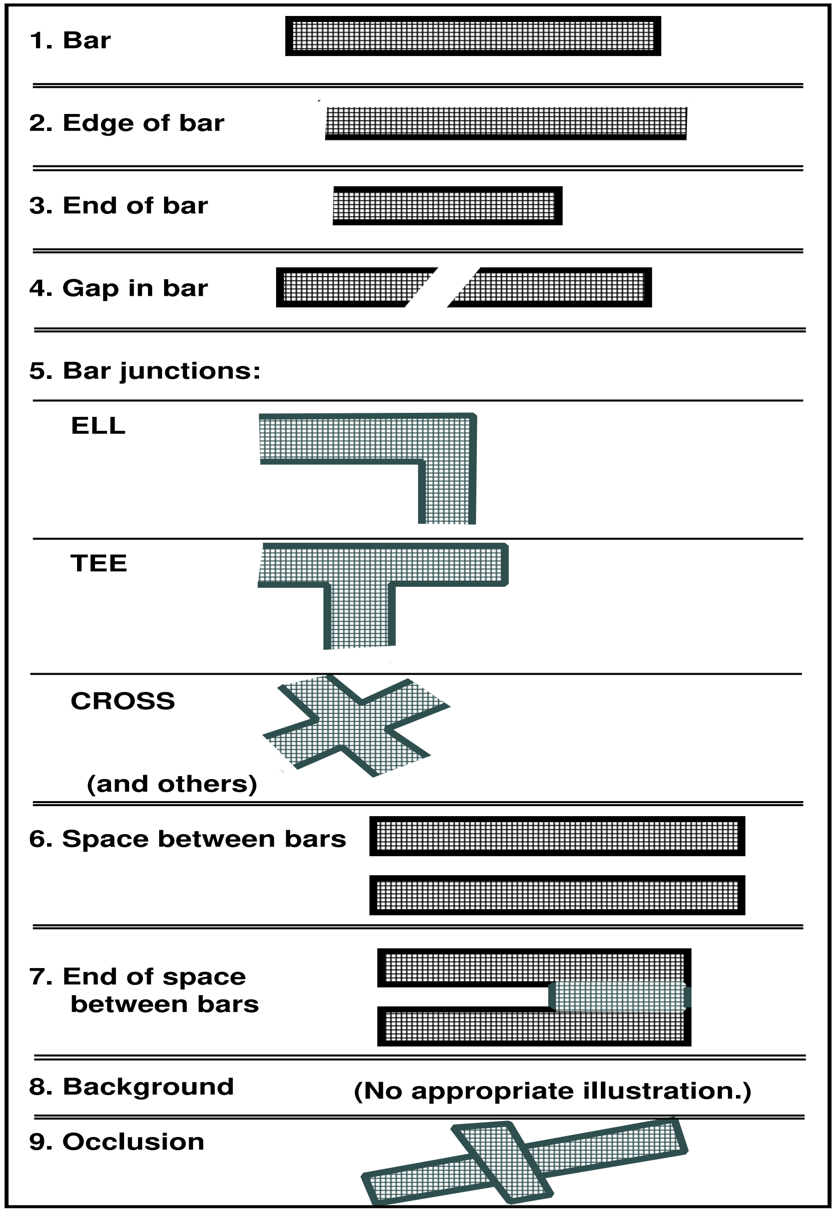
- **some of different sizes at the same level of abstraction,**
- **others at different levels of abstraction i.e. using different ontologies.**

Various fragments are recognised in parallel and assembled into larger wholes which may trigger higher level fragments, or redirect processing at lower levels to address ambiguities, etc.

Here we have some of the fragments at the level of configurations of dots, and the next abstraction level, configurations of continuous line segments



Useful fragments at different levels of abstraction



Here are some of the significant fragments detectable in the domain of overlapping laminas made from merged rectangular laminas.

These might be worth learning as useful cues if the system can detect that they occur frequently.

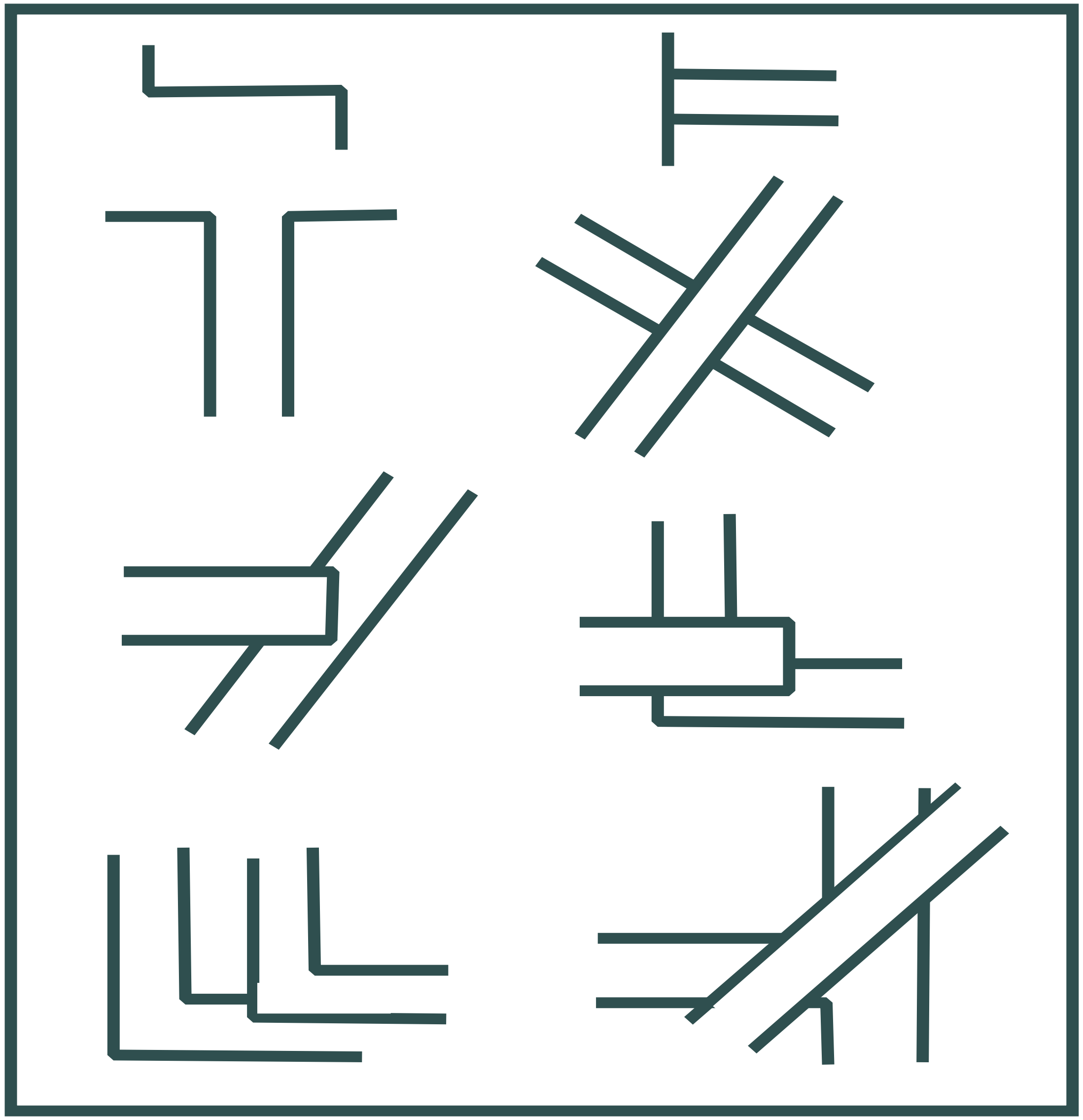
|X| Larger scale line "phrases"

J. Becker proposed in 1975 that fluency in use of language involved the use of a "phrasal lexicon".

This idea generalises to vision also.

Larger "phrases" in the "language" of line fragments. These might be worth learning as useful cues if the system can detect that they occur frequently.

Could a neural net learn such things?
Are there any known mechanisms that are appropriate?



Putting it all together

The Popeye architecture specified concurrent processing at all these different levels of abstraction.

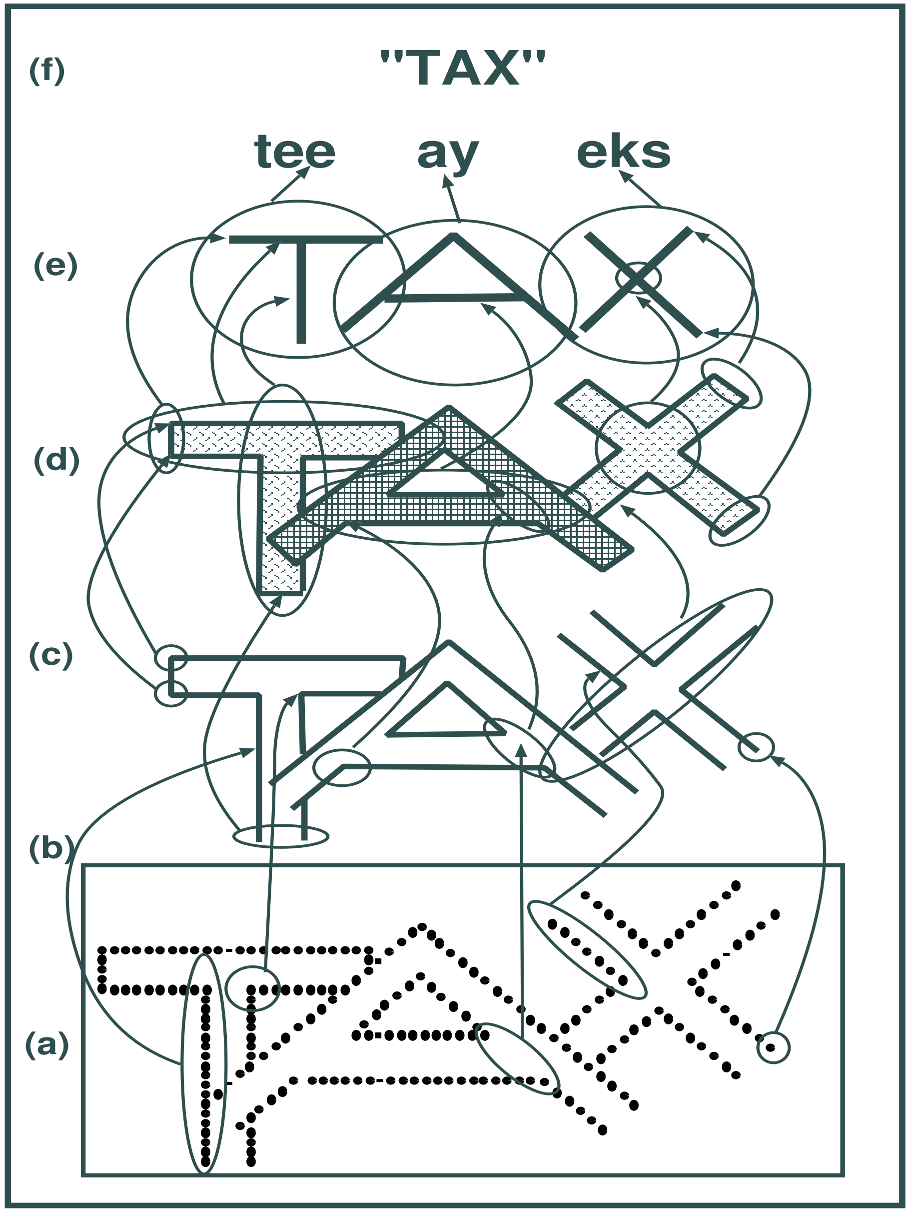
Sub-systems at different levels could interact with other sub-systems, including interrupting them by providing relevant new information or redirecting “attention” or altering thresholds.

Sometimes a higher level subsystem (e.g. word recogniser) would reach a decision before lower levels had finished processing.

Sometimes the decision was wrong!

Perhaps a network of neural nets could learn such things?

How many known mechanisms are appropriate for this sort of purpose?



Seeing beyond structures

Besides structures of different sizes at a given level of abstraction, and structures at different levels of abstraction, perception can also involve other things than perceived structures.

- So far we have summarised only what is seen that is present to be seen.
- But much of what is seen involves affordances and these go beyond what is there, since they involve what might be there.
- Seeing graspability, bendability, obstruction, passage, ...

  (Think of Betty the crow.)
- Seeing functions, functional and causal relationships.
- Seeing mental states or "intended actions" of other agents.

What does seeing these things involve?

What does it mean to say that an animal or machine sees them, or even can think about them, or can use the information?

Answering this requires specifying the larger architecture within which the perceptual processes form a part.

Building a self contained visual system is not enough

The POPEYE system could identify components in the scene and their relationships and use that to guide the recognition of other components and relationships, at different levels of abstraction.

(Some critics tended to confuse this with the then fashionable notion of "heterarchic" processing strongly criticised by David Marr.)

But vision does not occur in isolation.

A visual system is part of a whole organism, or robot.

What the visual system needs to do will in part depend on what the organism needs, and on what other components there are in the system.

Other components

- can ask the visual system questions,
- can use information provided by vision,
- can help to train the visual system,
- can provide information for the visual system,
- ....

Chaos in architecture-land

Unfortunately, there is much confusion in discussions of architectures.

E.g. different people use apparently similar diagrams and descriptions, to refer to different architectures.

E.g. “multi-layer” architecture means different things to different people. (Compare our three layers below.)

There is also too much factionalism (narrow vision).

Many people commit themselves to one or other type of mechanism (e.g. neural nets) or one type of architecture (e.g. subsumption) without having any really clear idea what the alternatives are or what the trade-offs between them are, ignoring the history of the field.

Some also teach their students to be too narrow-minded.

Contrast Minsky’s analysis of trade-offs between neural and other forms of computation:

‘Future of AI Technology’, 1997,
http://www.media.mit.edu/people/minsky/papers/CausalDiversity.html,
Original version in Toshiba Review, Vol.47, No.7, July 1992.

|X| Another problem: what needs to be explained?

It is too easy to assume we know what capabilities need to be explained, for they are our capabilities.

Problems with this assumption:

- **We are not necessarily aware of *which* capabilities we use in many tasks, or even *that* we are performing them, e.g. posture control, recognising features, analysing structures, solving image correspondence problems, reacting to facial expressions, doing visual learning.**
- **In particular, we may not always be aware of the role of visual processing in some of those tasks, e.g. in doing abstract mathematics**
  (See Talk 7 here: http://www.cs.bham.ac.uk/~axs/misc/talks)
- **What may appear to be *one* task, e.g. estimating distance, or seeing shape, or comparing angles, may actually be different tasks in different contexts, performed in different ways in different parts of the information processing architecture, using different forms of representation, e.g. judging distance in preparing to jump across a ditch, and judging distance in selecting a plank to lay across the ditch.**

We still need to identify the diverse functions of vision: a requirement for building adequate explanatory theories or working models.

|X| In humans there is great diversity of visual capabilities

E.g.

- What we see goes far beyond geometric/spatial structures and properties.
- We see many things that are abstract, some of them possibly shared with other species, others unique to humans.
- We can train ourselves to see and interpret things more quickly and fluently (e.g. learning to sight-read music).

Some human visual capabilities are culture-specific or location specific (e.g. in snow or in forests), while others are more general, e.g. the ability to see symmetries.

Non-geometrical visual percepts are harder to explain:

– seeing causal and functional relations like ‘holding up’, ‘obstructing’,

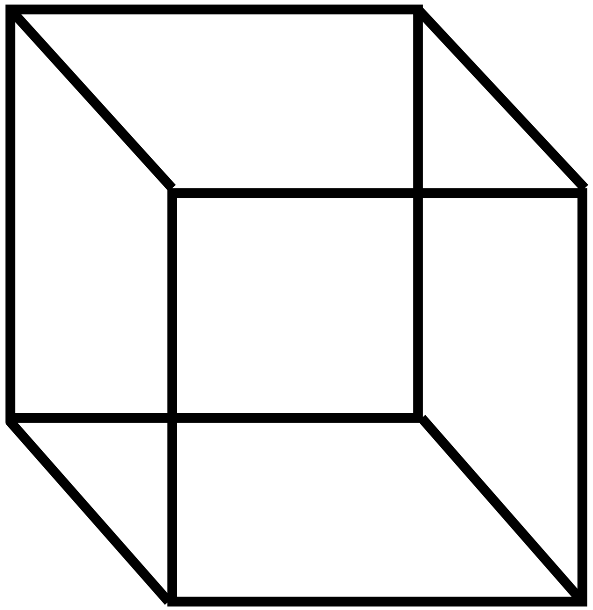
– seeing which way someone is facing,

– seeing how someone feels

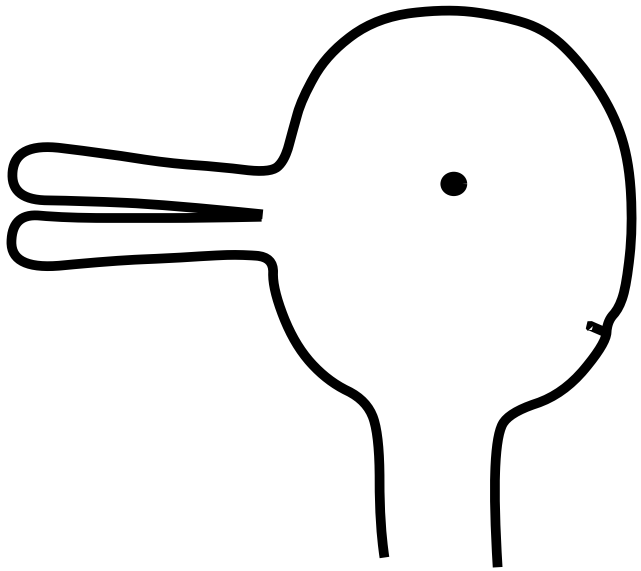
Examples: non geometric percepts

It is often thought (e.g. by Marr, or some of his followers) that visual systems provide only information about geometrical properties and relationships of objects in the environment, plus surface properties like colour and texture; and also physical changes.

But some visually ambiguous figures suggest otherwise:



Necker cube



Duck-rabbit

|X| Necker Cube and Duck-rabbit

When the Necker cube figure flips, all the changes are geometric.

They can be described in terms of relative distance and orientation of edges, faces and vertices.

When the duck-rabbit flips the geometry does not change:

- The functional interpretation of the parts changes
- More subtle features change, attributable only to animate entities.

  E.g. Which way is the animal looking?

These differences are visual, not simply inferential.
The examples occur in textbooks on vision, not reasoning

What does it mean to say that you "see the rabbit facing to the right".

Perhaps it involves seeing the rabbit as a potential mover, more likely to move right than left.

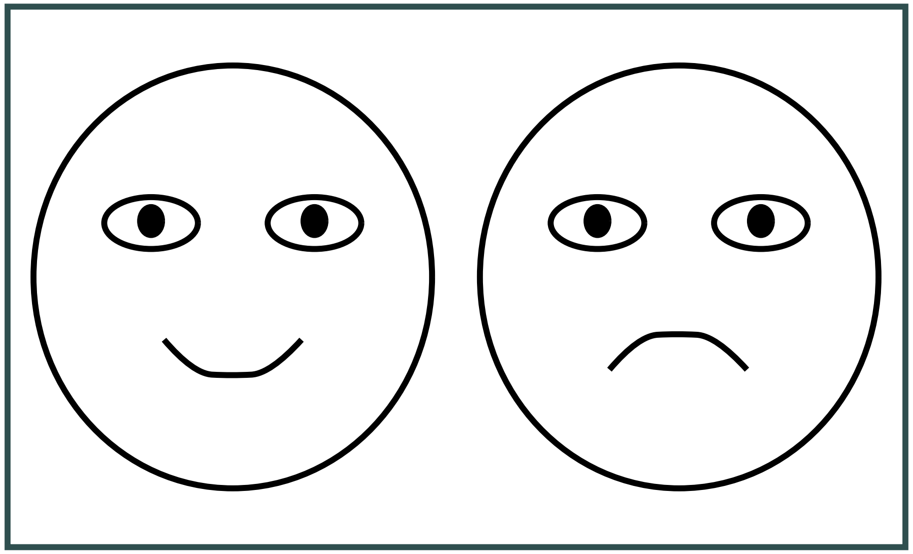
Or seeing it as a potential perceiver, gaining information from the right.

What does categorising another animal as a perceiver involve? How does it differ from categorising something as having a certain shape?

|X| Seeing Faces

Seeing facial expression as we do may just be a very old and simple process in which features of the face trigger reactions in a pattern-recognition device.

Or it may also involve deployment of sophisticated concepts that developed only through the evolution of meta-management. (Explained later)

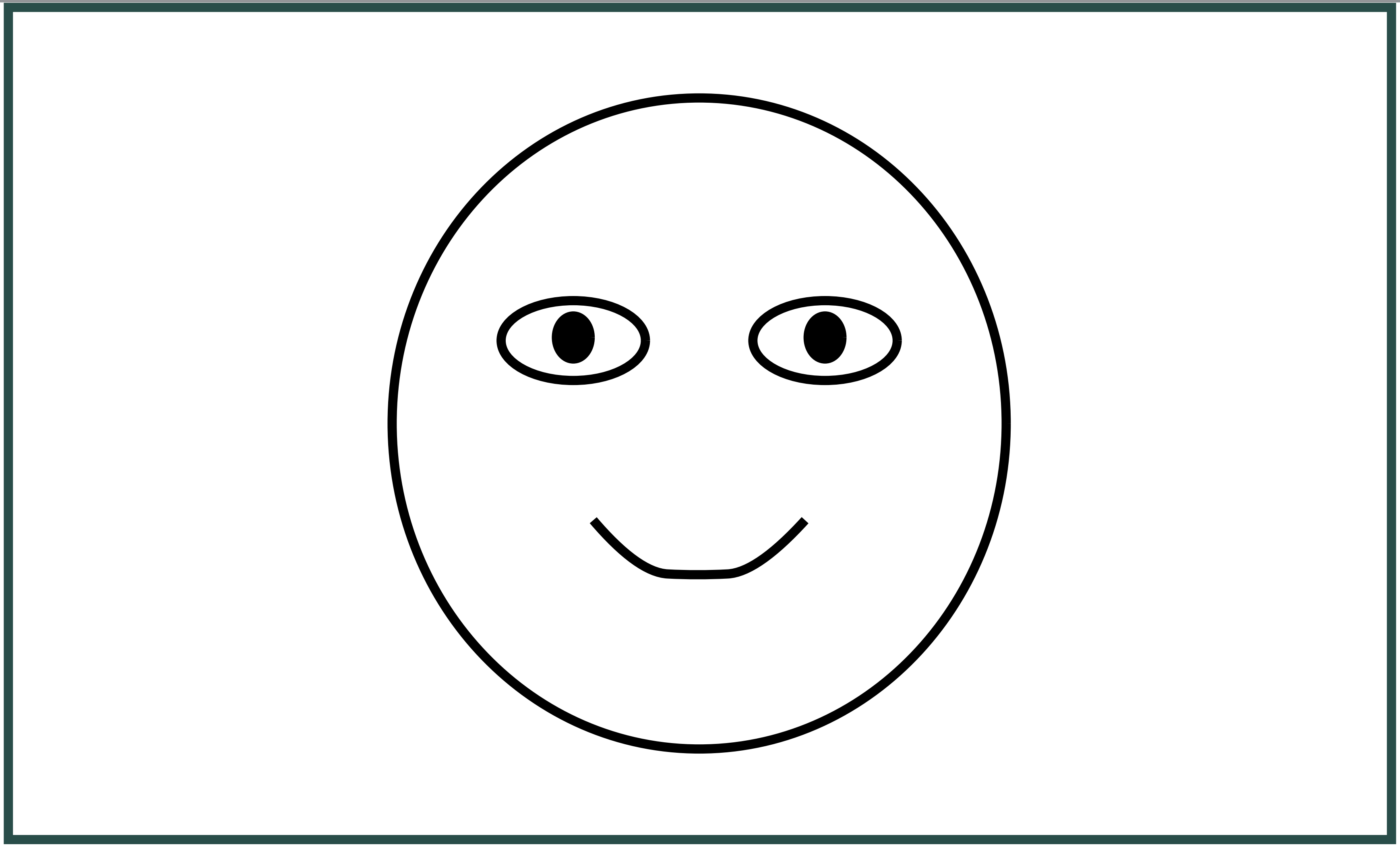


For more on levels in perceptual mechanisms see the talk on visual reasoning and other talks here:
  http://www.cs.bham.ac.uk/˜axs/misc/talks/

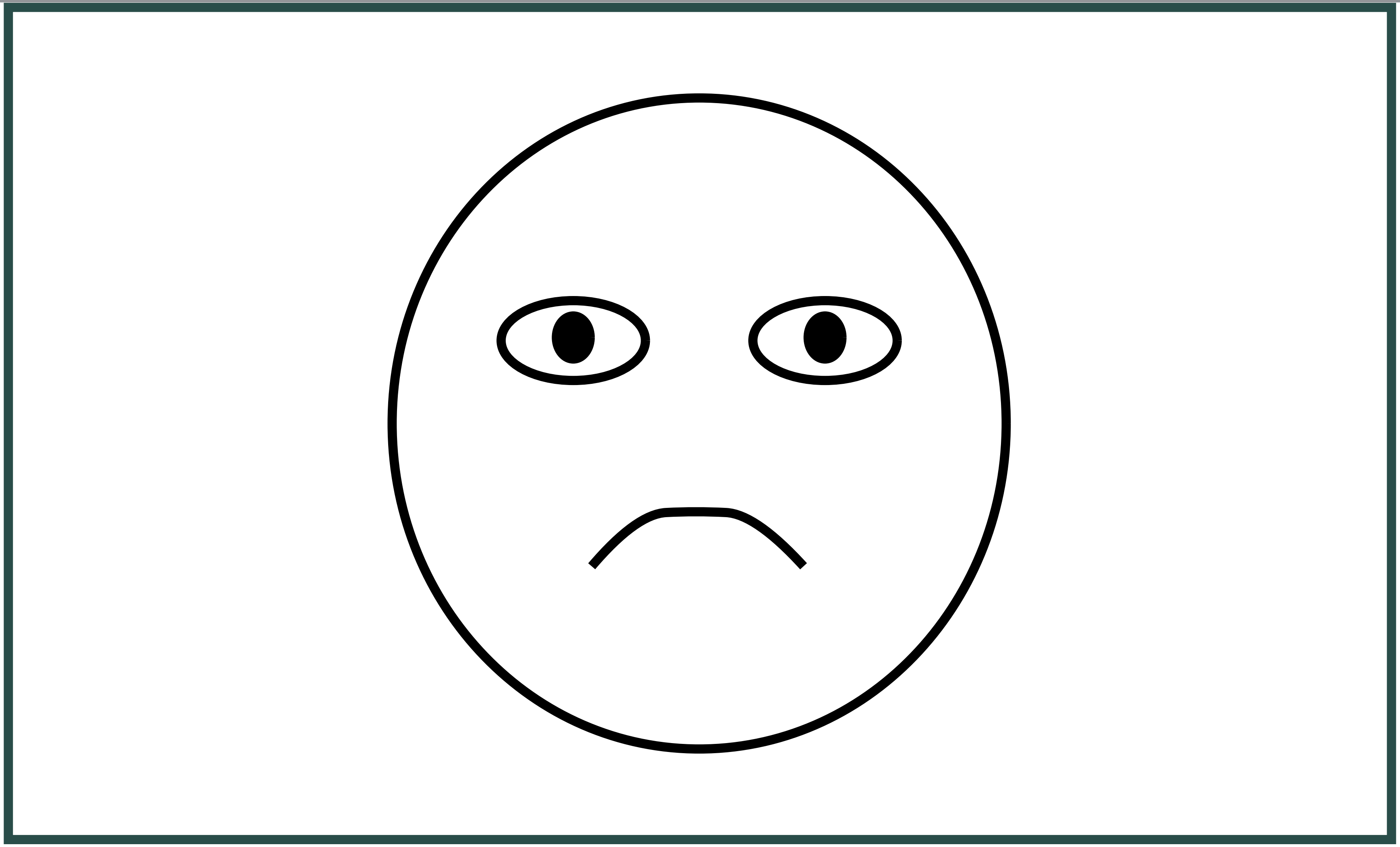
Some people see one pair of eyes as “looking happy” while the other pair “looks sad” or “looks angry”. (A non-geometrical context effect.)

Using the next two slides to flip rapidly between them may make this more evident.

A face



A face



|X| Seeing mental states

What is involved in seeing an “expression”
e.g. happiness, sadness?

It is NOT just a matter of recognising and labelling a pattern.

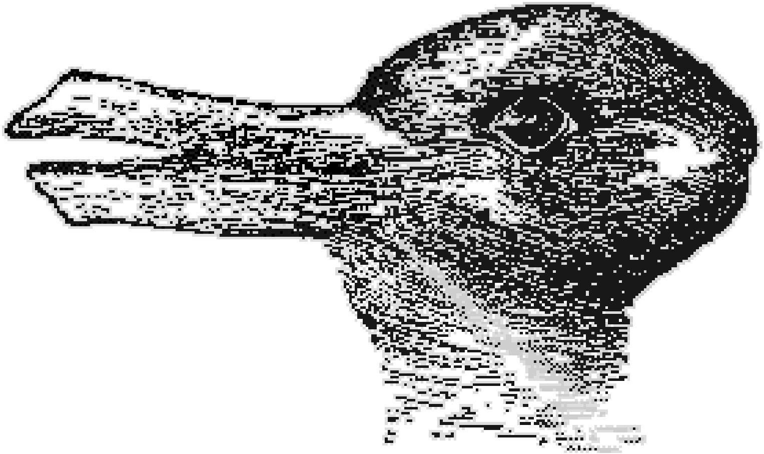
Those visual categories are semantically linked to matters of importance to us as social animals,

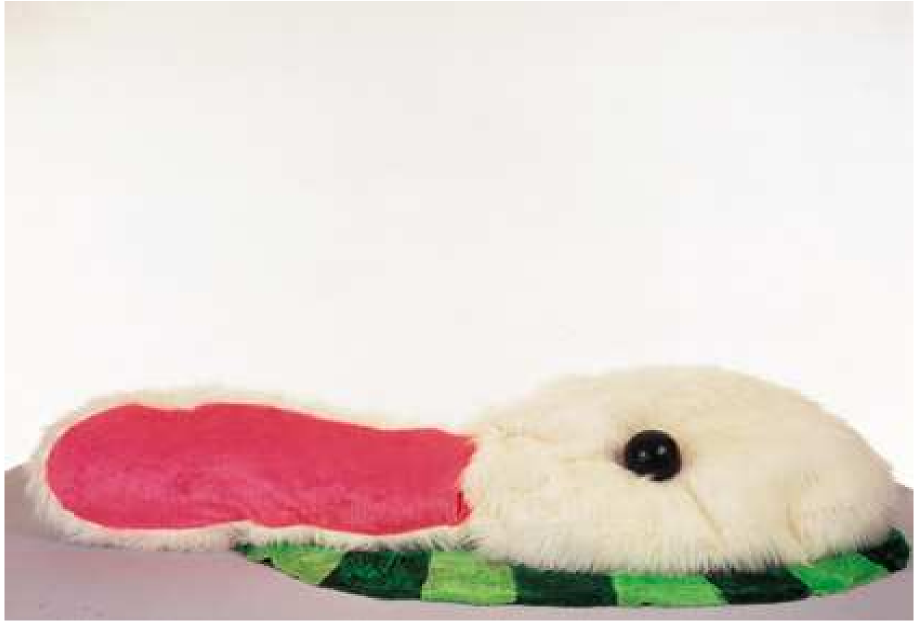
just as the perception of geometric structure
is linked to our needs as agents in complex 3-D world
and our ability to act in that world.

Seeing how someone feels can affect what you should do next: a non-geometric kind of affordance. and it seems to ‘colour’ the whole percept.

An appropriate architecture should explain the ability to have this sort of percept. (See H-Cogaff, later.)

|X| Ducks and Rabbits and other furry things





By Jastrow, and by Kathy Temin (Search on Google)

Another famous example

What changes between the views?

Are there geometric or physical changes?

What else?

Could an infant, or a chimp, or a crow see what we see?

|X| Is it seeing or inferring?

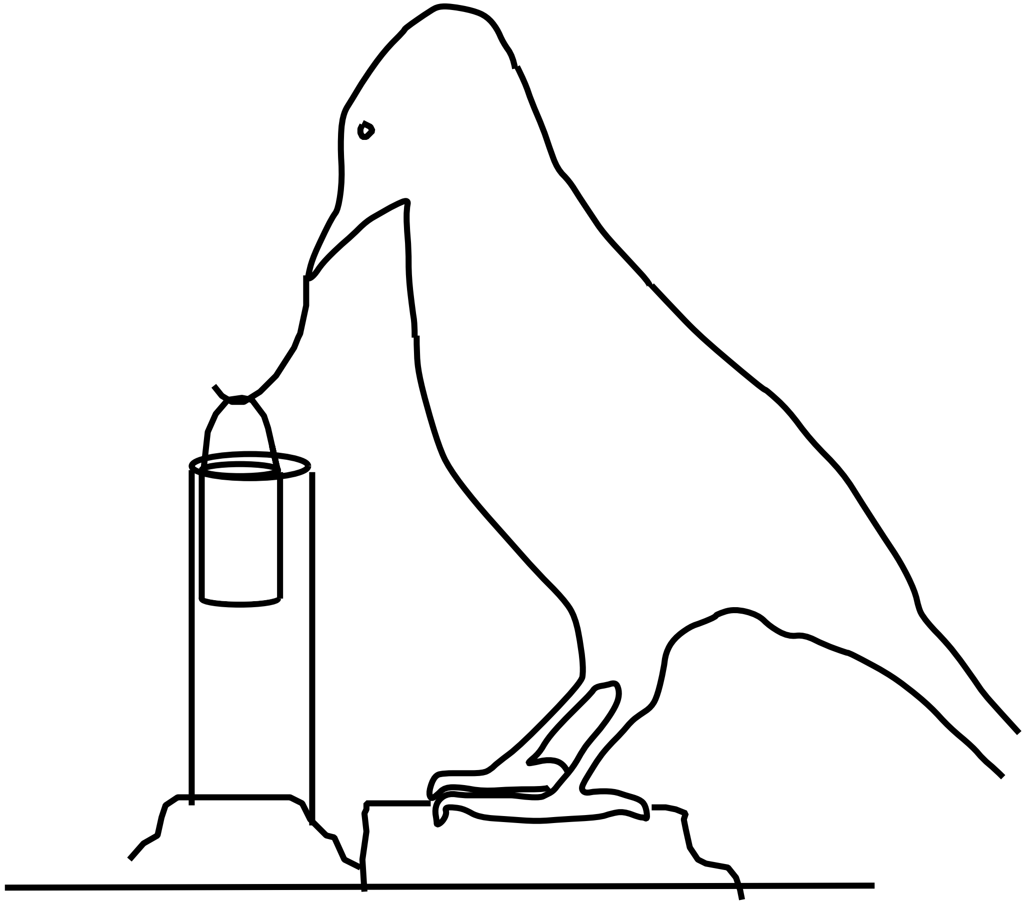
Conjecture:

The explanation of the phenomenology of *seeing* the two interpretations is that:

- **mechanisms specific to visual processing produce both geometric and non-geometric interpretations**
- **the more abstract, non-geometric details, including information about affordances (and the 'reference features' discussed by Pryor and Collins) like the 3-D geometric details, are stored *in registration with* the image features and relationships that give rise to them**

  Compare the 'TAX' picture
- **However, exactly what is done and how it is done remains an open question, to be answered as part of the study of the complete architecture within which the visual system is deeply integrated.**

Meet Betty



Betty the hook-making crow.

See the video here:

http://news.bbc.co.uk/1/hi/sci/tech/2178920.stm

There is far more to perception than detecting what exists in the environment

Betty the crow had to perceive not only the things that were before her at the start:

- The large transparent tube
- The bucket of food in the tube
- The piece of wire

She also had to see the possibility of things that did not exist but might exist, e.g.

- The possibility of the bucket moving up the tube,
- The the possibility of the wire being bent and holding its shape
- The possibility of various steps in the process of bending the wire
- The possibility of using the bent wire (which does not yet exist) to lift the bucket of food.

These are all cases of the perception of affordances, whose importance was noted by the psychologist J.J.Gibson.

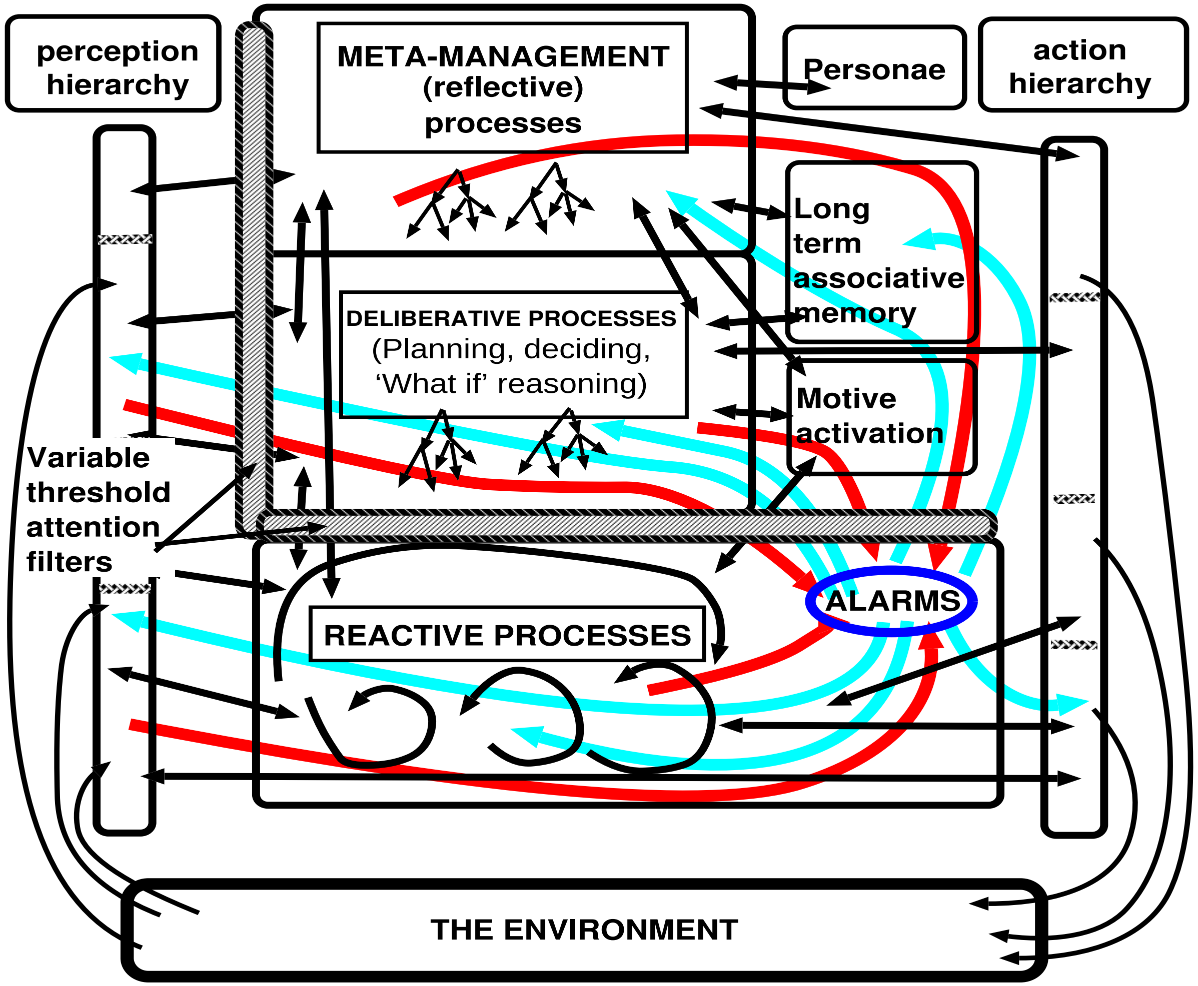
Affordances are the *possibilities for* and *constraints on* action and change in a situation.

Affordances in an environment depend on the goals and action capabilities of the organism (or robot) perceiving the environment.

The “Human-like” sub-schema H-Cogaff

Our conjectured architecture for human-like systems:

The H-Cogaff (Human-Cogaff) architecture is a special case of the CogAff architecture scheme/framework described in various papers and presentations



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

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

|X| The “Human-like” sub-schema H-Cogaff

- The reactive, deliberative and meta-management layers evolved at different times, requiring discontinuous changes in the design, and providing significantly new capabilities.
- An attention filter with dynamically varying threshold may be used to protect resource-limited higher level functions.
  (Luc Beaudoin’s PhD thesis 1994)
- Some aspects of the alarm system apparently correspond to the brain’s limbic system.
- Frontal lobes apparently implement some meta-management functions.

See the Cogaff papers:
http://www.cs.bham.ac.uk/research/cogaff/

Objects of vision

- Different levels of structure (agglomeration and segmentation within a domain)
  - 2-D structure
  - 3-D structure
  - Motion structures (e.g. gestures, actions, events, etc.)
- Different levels of ontology: mappings from one domain of structure to another.
- Ontologies involving causal and functional relations.
- Ontologies involving mental states (information processing states in others)
- Ontologies involving possibilities of many kinds inherent in a situation: Affordances

Multi-window perception and action

If multiple levels and types of perceptual processing go on in parallel, we can talk about

"multi-window perception",

as opposed to

"peephole" perception.

Likewise, there can be multi-window action or peephole action.

Some implications

Within this framework we can explain

- research findings on different visual pathways (and predict more)
- blindsight (damage to some meta-management access routes prevents self-knowledge about some visual processes)
- varieties of emotions (at least three distinct types related to the three layers: primary, secondary and tertiary emotions)
- many varieties of learning and development
- the discovery by philosophers of 'qualia'
- some of the evolutionary trade-offs in developing these systems (Higher levels can be very expensive, and require a food pyramid)

and probably much more

Warning to experimenters

Of the many forms of concurrent perception in different parts of the architecture, we are *aware of* only those aspects accessible to the meta-management processes.

So we cannot report verbally on, or otherwise voluntarily indicate the presence of, the others, including:

- some of the perceptions in the reactive sub-system which influence reactive behaviours
- some of the intermediate stages in visual processing which produce percepts that meta-management can access: e.g. we may be unaware of intermediate stages in producing percepts of objects in the environment, even where we are aware of results of *later* stages

So we cannot assume that asking subjects questions in experiments, or getting them to press buttons or turn dials to indicate what they see is a reliable way to find out everything they can see.

I.e. this theory implies that there are forms of "blindsight" in normal humans: it is not just a product of brain damage.

But much of this is still far too vague

There is a huge amount of work still waiting to be done

Including working out in great detail:

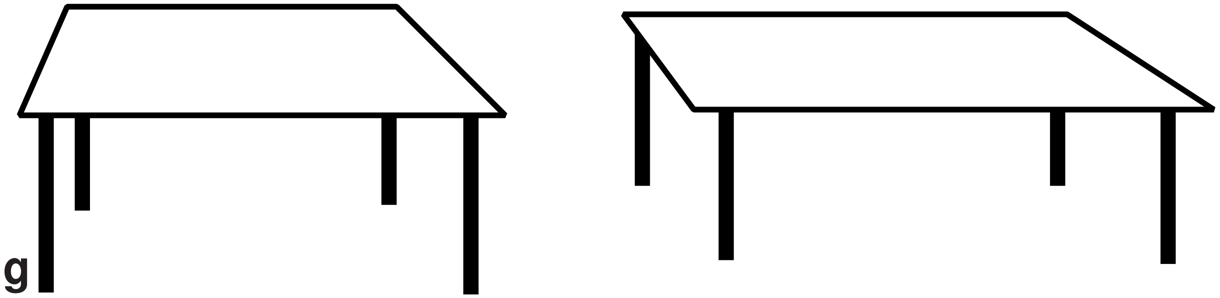
- what sorts of visual capabilities are possible
  (in humans and other animals)
- and how they relate to niche features (PPSN2000 paper),
- and then investigating ways of explaining and implementing them.

This will very likely require us to discover:

- new forms of representation,
- new information processing mechanisms for manipulating them,
- new architectures to incorporate and make use of those mechanisms.
- new characterisations of what such such architectures can use vision for (e.g. seeing *possibilities* and *impossibilities*)

|X| Seeing tables

What sorts of affordances does a table provide?



– Obstruction
– Support
– Pulling, lifting, pushing, in various ways depending where you hold it and how.
– Easy availability of a collection of tools or papers, etc., in easy reach
– Social cohesion during meals
– Types of construction and repair methods
– ....

(See my ‘Actual possibilities’ paper at the CogAff web site.)

Some of the affordances are conditional: e.g. you can pull the table if you (a) move closer and (b) grasp a leg or the edge.

How do we (and other animals) represent collections of possibilities and constraints on possibilities? How do we use our grasp of such possibilities and constraints to work out what to do?

Do we, or chimps, use modal logics?

Seeing possibilities and doing mathematics

Visual mathematical reasoning requires the ability to see not only structures but also

- **Possibilities for change**
- **Constraints on possibilities for change**
- **At various levels of abstraction**
- **E.g. metrical change, topological change, structural change**

The more complex a structure is the more possibilities for (small) changes it supports.

For more examples see talk 7 here:

http://www.cs.bham.ac.uk/~axs/misc/talks/

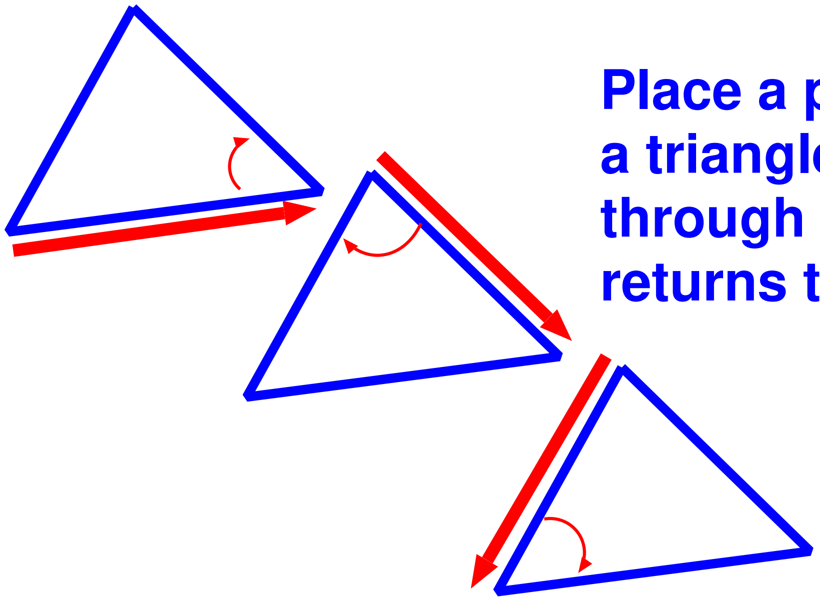
(Talk 7: Seminar slides on visual/spatial reasoning.)

Compare L.Wittgenstein's discussion of "seeing as" in his *Philosophical Investigations*, Part II, section (xi), 1953.

Seeing mathematical relations

There is a long history of people claiming that visual capabilities can be used for reasoning in everyday life and in mathematics.

E.g. how do you prove that the angles of a triangle add up to half a circle, i.e. 180 degrees.



Place a pencil along an edge of a triangle and rotate it in turn through the three angles until it returns to the original edge.

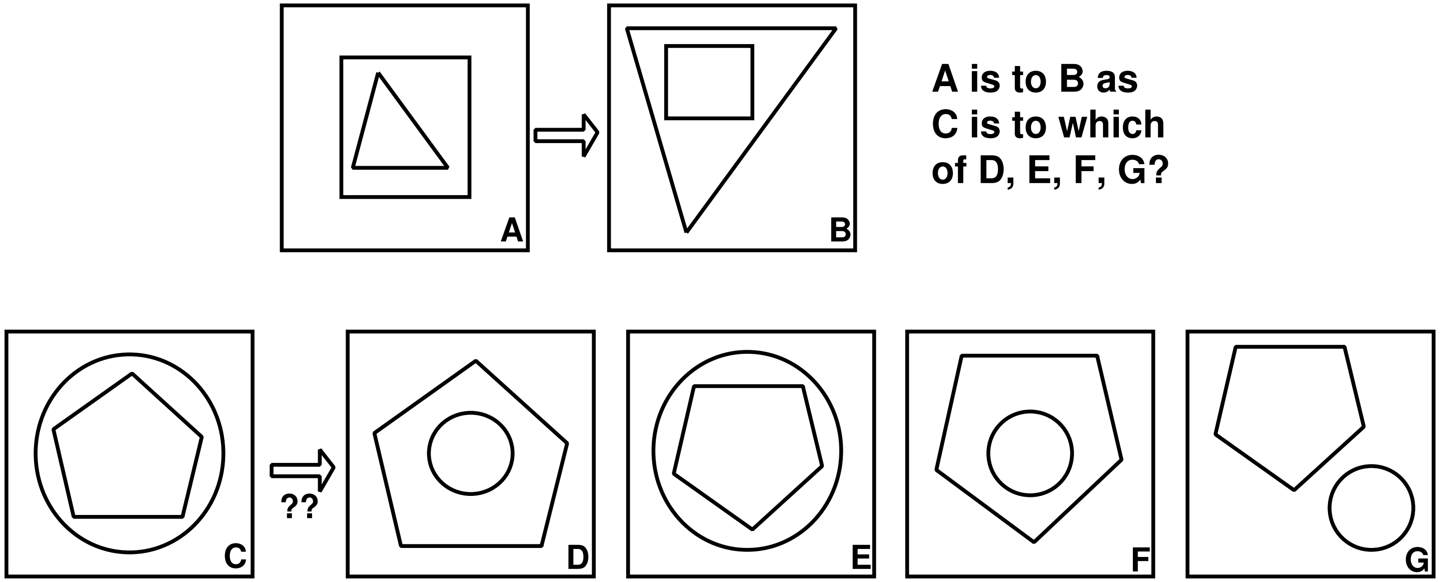
How much has it rotated?

It is not necessary to use an actual triangle and pencil: the process can be visualised.

What difference does it make if you visualise *external* rotations of the pencil? (There are two cases, giving different results.)

Seeing relationships between relationships

Standard intelligence test problems require one to see structures and to grasp not only relationships between parts of the structures, but also relationships between relationships (or, put another way, transformations of relationships). How do we see those?



An amazing program by T.G. Evans did this kind of thing in 1968:

‘A heuristic program to solve geometric analogy programs,’ in
Semantic Information Processing, Ed., M.L. Minsky, MIT Press, pp. 271–353
(Could you program that sort of capability?)

More examples of visual reasoning

The ability to reason visually is part of everyday life

- How far should I lean over the table in order to be able to reach the salt cellar on the far side?
- Where should I stand in relation to the window in order to be able to see the left edge of the building opposite?
- How should I rotate that chair in order to get it through that door?
- Along which branch should I climb in order to be able to swing onto the next tree?
- Is the vase safely out of reach of that child?
- How should I cut these sheepskins in order to be able to assemble a jacket from the pieces?
- How can I design a mechanical loom, or a machine to make wind grind corn?

There are many activities that used visual reasoning long before the development of mathematics as we know it, but which may have used mechanisms that later made mathematical reasoning possible.

Even non-visual mathematical reasoning using algebraic and logical formulae requires us to be able to "see" structural relations in formulae, and to notice possibilities for syntactic transformations in those structures: more visual affordances.

Towards a taxonomy of uses of vision

Has anyone attempted a systematic overview of the uses and capabilities of human and animal vision, including capabilities that are common to all and those that result from specialised training?

The vast majority of visual affordances, and visual reasoning capabilities are not yet understood. (Contrast segmentation, recognition, distance estimation, tracking, ...)

Consider what it is to see a horizontal plane surface:

- **Seeing it as having a uniform or changing texture or colour.**
- **Seeing it as separating the space above and below it.**
- **Seeing it as infinitely thin, or as indefinitely extendable.**
- **Seeing different parts as being at different distances from you.**
- **Seeing empty spaces as *possible locations* for a variety of shapes: lines, circles, pictures of faces, text, musical notation...**
- **Seeing parts of the surface as possible paths or trajectories.**
- **Seeing the possibility of a variety of processes in the plane: changing shape or texture, movement, pulsating objects, oscillations, etc.**
- **Seeing that the surface itself can move or rotate or bend in space.**

Can we explain all this?

CONJECTURES:

- Animal visual architectures evolved several layers of analysis and interpretation.
- These operate concurrently, feeding information into different central layers which require different kinds of information represented in different ways (different affordances).
- Different aspects of human vision are related to differences in the functionality and sophistication of the central systems that they feed into.
- Likewise, there are likely to be different sorts of ‘top-down’ influences on visual processing, coming from different parts of the central architecture, with different requirements.
- In humans these include the ability to visualise what is not there and changes in what is there.
- Animals that have internal self-monitoring capabilities need conceptual apparatus for that task which can also be used in categorising mental states of other agents.

|X| Forms of representation in visual systems

Forms of representation studied so far for vision include

- 2-D rectangular arrays,
- concentric rings of receptive fields of varying size,
- weights or activations in neural nets,
- Fourier transforms,
- histograms and probability distributions,
- structural descriptions (parse trees),
- various symbolic representations of map structures,
- semantic nets,
- logical databases,
- control signals, ... and more

Biological vision probably uses forms of representation not yet thought of.

A hard problem: how to represent "affordances", and more generally information about possible changes and constraints on changes in a visible portion of the world.

See KR1996 paper on "Actual possibilities" at http://www.cs.bham.ac.uk/research/cogaff/

Generalising Gibson's notion of 'affordances'

Instead of thinking about visual 'affordances' for an organism, we think about the affordances for various components of an organism.

E.g. the following need different information from the environment, probably represented differently:

- posture control in two-legged walking
- control of visual saccades
- selection of routes
- building a shelter

The last task might include all the others!

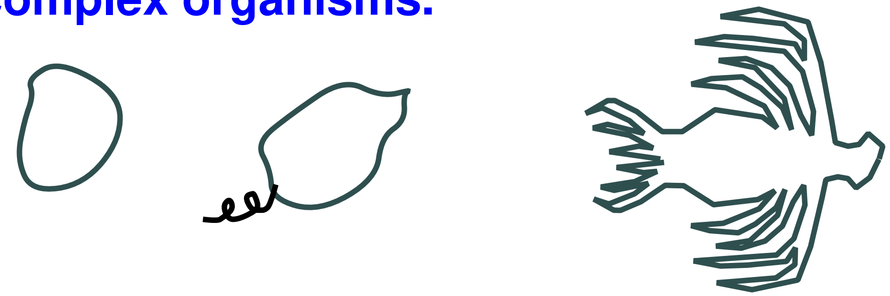
An architectural framework incorporating multiple mechanisms allows us to think about multiple visual pathways and multiple forms of learning and development.

We can then ask deeper questions about evolution: because we can formulate options with a deeper understanding of the space of designs and their trade-offs: e.g. trade-offs between species evolution and individual learning as means of acquiring information.

A biological perspective

Once upon a time there were only inorganic things: atoms, molecules, rocks, planets, stars, etc. These merely reacted to *resultants* of all the physical forces acting on them.

Later, there were simple organisms. And then more and more complex organisms.



These organisms had the ability to reproduce. But more interesting was their ability to *initiate* action, and to *select* responses, instead of simply being pushed around by resultants.

That achievement required the ability to acquire, process, and use *information.*

|X| The ability to act or to select requires information

E.g. information about

- density gradients of nutrients in the primaeval soup
- the presence of noxious entities
- where the gap is in a barrier
- precise locations of branches in a tree as you fly through
- how much of your nest you have built so far
- which part should be extended next
- where a potential mate is
- something that might eat you
- the grass on the other side of the hill
- what that thing over there is likely to do next
- how to achieve or avoid various states
- how you thought about that last problem
- whether your thinking is making progress

and much, much more... (has anyone attempted a taxonomy?)

|X| Resist the urge to ask for a definition of "information"

Compare "energy" – the concept has grown much since the time of Newton. Did he understand what energy is?

Instead of defining "information" we need to analyse the following:

– the variety of types of information there are,
– the kinds of forms they can take,
– the means of acquiring information,
– the means of manipulating information,
– the means of storing information,
– the means of communicating information,
– the purposes for which information can be used,
– the variety of ways of using information.

As we learn more about such things, our concept of "information" grows deeper and richer.

Like many deep concepts in science, it is *implicitly* defined by its role in our theories and our designs for working systems.

|X| Things you can do with information

A partial analysis to illustrate the above:

- You can react immediately (it can trigger immediate action, either external or internal)
- You can do segmenting, clustering labelling of components within a complex information structure (i.e. do parsing)
- You can try to derive new information from it (e.g. what caused this? what else is there? what might happen next? can I benefit from this?)
- You can store it for future use (and possibly modify it later)
- You can consider alternative next actions, or make plans
- If you can interpret it as as containing instructions, you can obey them, e.g. carrying out a plan.
- You can observe the process of doing all the above and derive new information from it (self-monitoring, meta-management).
- You can communicate it to others (or to yourself later)
- You can check it for consistency, either internal or external

... using different forms of representation for different purposes.

The various kinds and uses of information processing did not all evolve at the same time

Not all of them occur in all animals (microbes, insects, fishes, reptiles, birds, mammals, etc.)

A particular collection of sensory transducers (visual, auditory, tactile) can provide many different kinds of information at the same time, e.g. the text on the page, the window beyond the page, the state of the weather visible through the window, all in one visual field.

- Some information is very localised and simple (here’s a dot, there’s some motion).
- Other information may be far more holistic (e.g. recognising a scene as involving a forest glade).
- Some may be very abstract (the weather looks fine; it looks as if a fight is about to break out in that crowd).
- Some mechanisms involve only generally applicable knowledge about the geometry and topology of static and moving shapes.
- Others require specific knowledge about things that are relevant only in a particular part of the world, or a particular type of activity. E.g. seeing text, hunting fast moving prey, seeing geological formations, looking at exposed brains.

Diverse mechanisms of varying sophistication

Extracting information from the basic sensory data may require very diverse perceptual mechanisms with varying types of sophistication.

- Some information can be extracted very simply (using spatial or temporal local change detectors, or mechanisms for constructing histograms of features, such as colour, texture, optic flow).
- Other information may need *relationships* to be discovered between features, e.g. collinearity, lying on a circular arc, parallelism, closure, lying on the intersection of the continuations of two linear segments or two curved segments (where the continuations are also curved).
- Sometimes this requires *searching* for coherent interpretations.
- Some relationships hold only between abstract entities not the image data: e.g. two people seen to be *looking in the same direction*.
- Extracting some of the information requires matching with known models ("That's a triangle, a face, a tree").
- Some learning tasks require noticing new repeated structures within the information structures (e.g. noticing repeated occurrence of polygons with circles at two adjacent corners).

|X| Virtual vs physical machines

In computer science, software engineering and AI we have learnt the importance of virtual machines, e.g. the Lisp, Prolog, Java virtual machines, chess virtual machines, neural net simulations, etc.

Mechanisms that operate on complex information structures are typically virtual machines (parsers, structure matchers, network constructors, search engines, planners, interpreters, etc.) rather than physical machines, though virtual machines are *implemented* in physical machines.

This implies that if we are to explore the full range of architectures for intelligent systems, including architectures for visual systems, we need to be familiar with a wide range of techniques for constructing virtual machines of various sorts.

This has implications for the sorts of education that should be provided broad-minded AI students.

For more on the relation between virtual machines and physical machines (a hard philosophical problem) see the slides for my IJCAI tutorial with Matthias Scheutz:

http://www.cs.bham.ac.uk/˜axs/ijcai01/

Temporal and causal differences in virtual machines for vision

Some perceptual information is used "online", e.g.

- Posture control
- Control of a hand moving to pick up a pencil, or a pin, or to pick a berry in a thorny plant.
- Use of vision when parking your car
- Reading text aloud, or sight-reading a musical score as you play.

Some is stored for future use, in various modes, e.g.

- Recognising the person who punched you a week ago
- Remembering where you put a pencil
- Learning a new discrimination (e.g. learning to distinguish a pair of identical twins)
- Formulating generalisations
  (Xs are found inside Ys, Doing A to X, causes X to do B)
- Storing a plan that is found to be useful.
- Many perceptual-motor skills produced by training

Often online control can use *continuous* variation, whereas much stored information concerns *discrete* categories and relationships.

|X| The evolution of information processing architectures and mechanisms

Evolution "discovered" and used many things long before human engineers and scientists asked the questions: long before they even existed.

Paleontology shows the development of physiology and provides some weak evidence about behavioural capabilities.

But there is very little direct evidence regarding previous forms of information processing: virtual machines leave no fossils.

Archaeologists speculate wildly and (in my view) irresponsibly.

We can be more disciplined!

The variety of forms of information processing now found in nature gives many clues, and we can test theories in working models.

Some of the forms are evolutionarily very old. Others relatively new. (E.g. the ability to learn to read, design machinery, or do mathematics.)

WE NEED TO LEARN HOW TO ASK GOOD (DEEP) QUESTIONS.

Different information processing architectures

The different tasks require different kinds of mechanisms, often operating on different forms of representation and different forms of long and short term storage.

Sometimes they require different sub-mechanisms working together (perceiving, learning, using prior knowledge, deciding what to do, constructing plans, executing plans, etc.)

But there must always be an ARCHITECTURE combining all the mechanisms and processes they produce.

Some of the more sophisticated mechanisms and architectures evolved only relatively recently, and are in very few species (e.g. deliberative capabilities – see below)

We need to understand how they differ from, how they are built on, and how they interact with the much older, more wide-spread mechanisms.

The same organism, e.g. a human being, may include both very old and very new mechanisms, in many sub-systems.

Some differences are very subtle

Physiological and other similarities between visual systems of different mammals, e.g. lions and sheep, may mislead us.

There may be subtle, unobvious, but very important differences, e.g. where one organism has a mostly genetically determined information processing architecture, whereas another builds much of its architecture using a boot-strapping process after birth.

The results may be very different in the capabilities they support.

E.g. a grazing mammal and a hunting mammal have very different visual requirements. Likewise compare birds that just peck grains on the ground and nest at ground-level (chickens) with hunting birds that build tree-top nests (magpies).

Biologists distinguish:

– precocial species (e.g. deer, chickens)
– altricial species (e.g. lions, eagles).

Precocial species are born or hatched more physiologically developed and more behaviourally competent: why?

A clue: look at the different (adult) niches

We need to analyse and contrast the visual requirements of adults:

- grazing mammals (e.g. deer)
- hunting mammals (e.g. lions).

How do their visual tasks differ?

What are the implications of the requirement to be able to stalk, to chase, and then jump and bite the neck of a fast moving animal?

Will a deer and a lion see the same things if they look at the same terrain?

Compare the grasp of spatial structure and motion required for use of a hand with opposing thumb, in picking berries or moving small insects from tree branch to mouth.

Contrast that with the visual requirements of a bird that pecks at such berries or insects.

Contrast using *your own* hand to pick berries with watching *how another person* does it: the tasks are different in subtle ways.

Conjecture: bootstrapping in altricial species

In precocial species evolution can pre-program the visual capabilities required, and they are available to the young almost immediately.

In altricial species (e.g. hunting mammals, nest building birds, apes and monkeys) the activities of adults require a far more sophisticated visual grasp of structure and motion (and links to tactile perception).

Specifications for mechanisms that have all the latter information may be too complex to encode in genes.

But it may be possible to encode a bootstrapping system that causes the required mechanisms to be developed while the architecture grows, using a variety of exploratory actions including infant play.

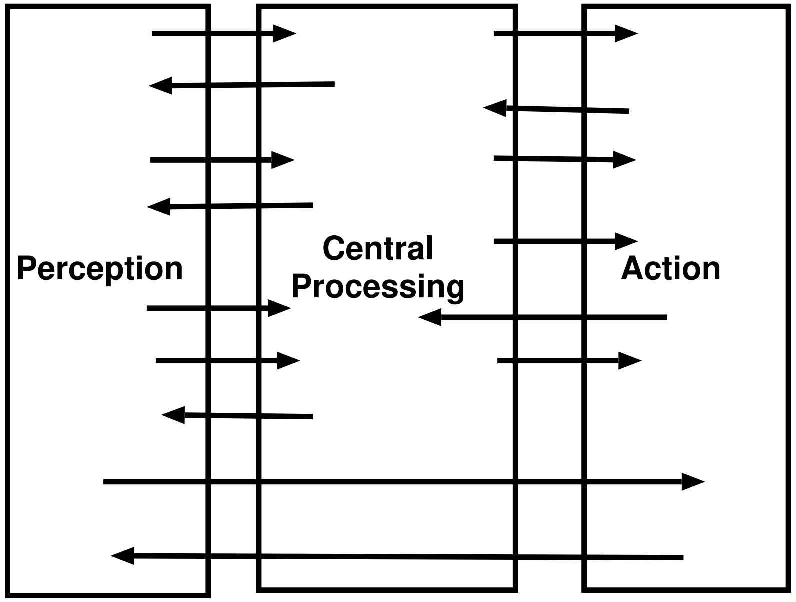
HOW IS THAT ACHIEVED ????

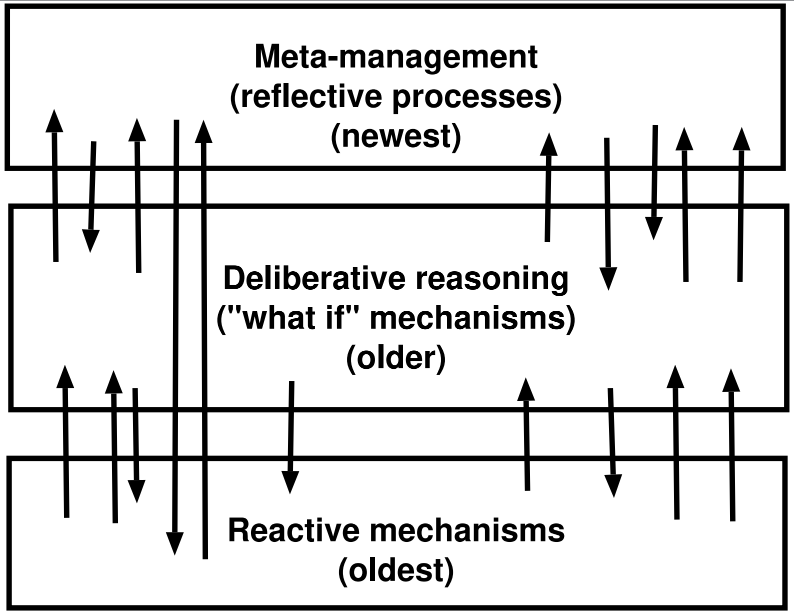
Could it all be just calibration, e.g. using play, etc., to specify quantitative parameters, within a fixed architecture?

I doubt it, but that's an open question.

Human learning capabilities, e.g. learning to speak or read, seem to arise from more general bootstrapping mechanisms.

Towards an architecture schema





Two coarse divisions within information processing architectures – 'towers' and 'layers':

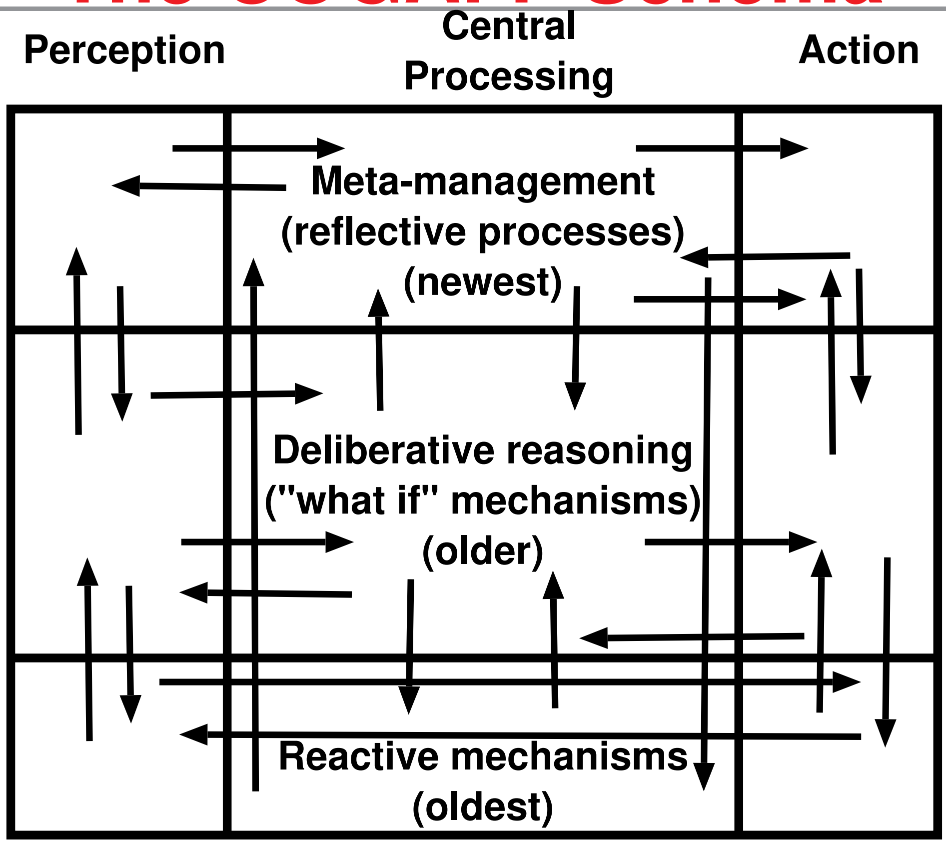
(a) Nilsson's (1998) "triple tower" model

(b) Layered architectures: e.g. reactive, deliberative and meta-management layers.

(a) and (b) express different (orthogonal) functional divisions.

These divisions can be combined, as follows ....

Superimposing the divisions: The COGAFF Schema

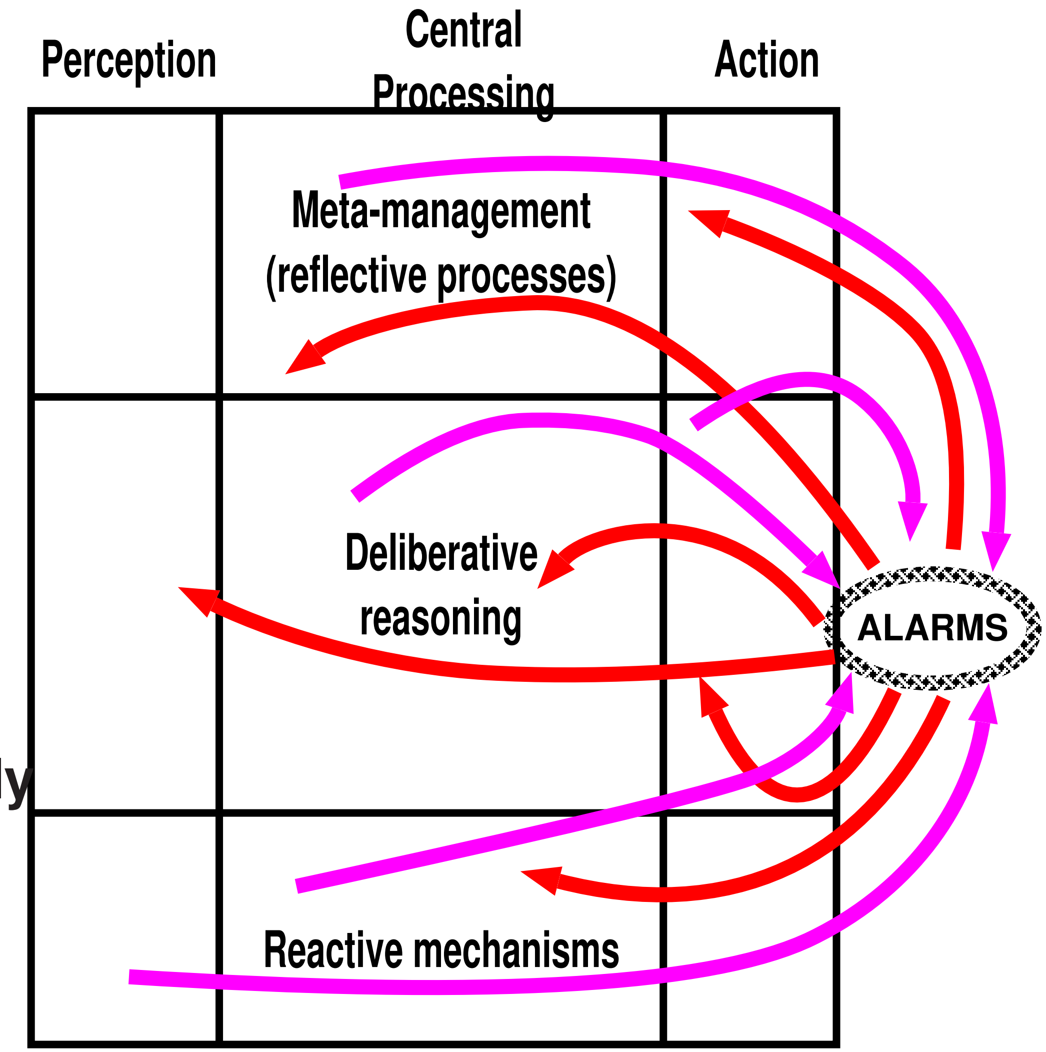


Boxes indicate possible functional roles for mechanisms: only some possible information flow routes are shown (cycles are possible within boxes, but not shown).

COGAFF extended – with “alarm mechanisms”

Alarm mechanisms deal with the need for rapid reactions using fast pattern recognition based on information from many sources, internal and external.

An alarm mechanism is likely to be fast and stupid, i.e. error-prone, though it may be trainable.



Characterising the layers

The differences between the layers are complex and subtle.

Some of the differences are discussed in other slide presentations here
http://www.cs.bham.ac.uk/~axs/misc/talks/

Further discussion is in the papers in the Cogaff directory
http://www.cs.bham.ac.uk/research/cogaff/

It may turn out that there are better ways of dividing up levels of functionality, or that more sub-divisions should be made – e.g. between analog and discrete reactive mechanisms, between reactive mechanisms with and without chained internal responses, between deliberative mechanisms with and without various kinds of learning, or with various kinds of formalisms, and between many sorts of specialised “alarm” mechanisms.

The COGAFF schema is still a draft, likely to evolve

|X| Architectural change in an individual

Learning can introduce new architectural components, e.g. the ability to read music, the ability to write programs.

Development of skill (speed and fluency) through practice can introduce new connections between modules, e.g. links from higher-level perceptual layers to specialist reactive modules.

For instance, learning to read fluently, or developing sophisticated athletic skills.

Highly trained skills can introduce new “layer-crossing” pathways, e.g. visual pathways: rapid recognition of a category originally developed for deliberation can, after training, trigger fast reactions.

Cogaff is a schema not an architecture: a sort of ‘grammar’ for architectures

Different organisms, different artificial systems, may have

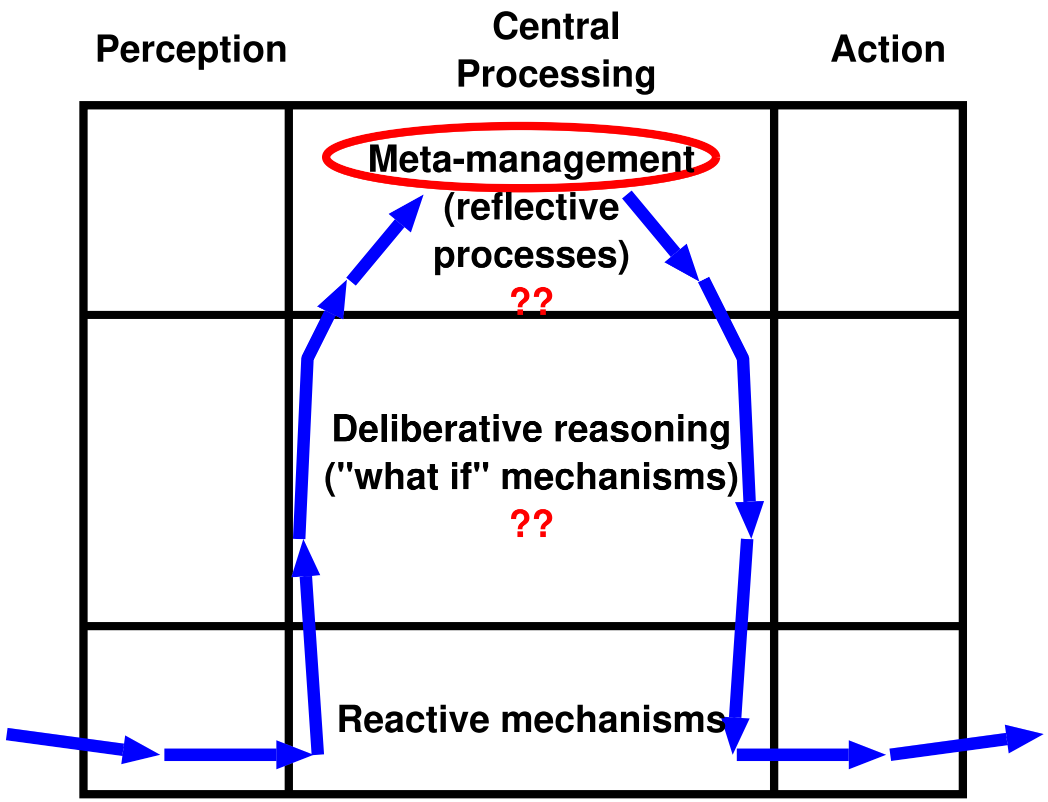
- different components of the schema
- different components in the boxes
- different connections between components

E.g. some animals, and some robots have only the reactive layer (e.g. insects, microbes).

The reactive layer can include mechanisms of varying degrees and types of sophistication, some analog, some digital, with varying amounts of concurrency.

Other layers can also differ between species.

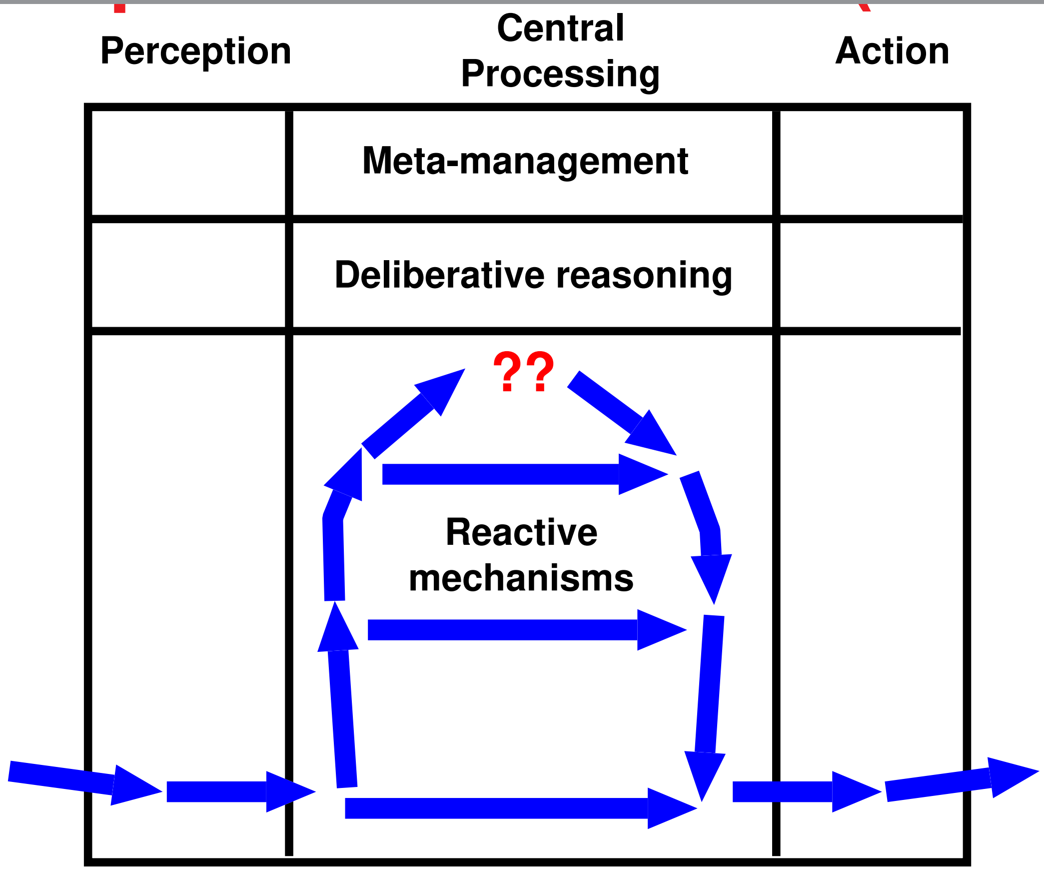
An example sub-category: Omega architectures



This is just a pipeline, with "peephole" perception and action, as opposed to "multi-window" perception and action.

E.g. Cooper and Shallice: Contention scheduling, Albus 1981.

Another sub-category: Subsumption architectures (R.Brooks)



This could be useful for certain relatively primitive sorts of organisms and robots. (E.g. Insects, fish, crabs?)

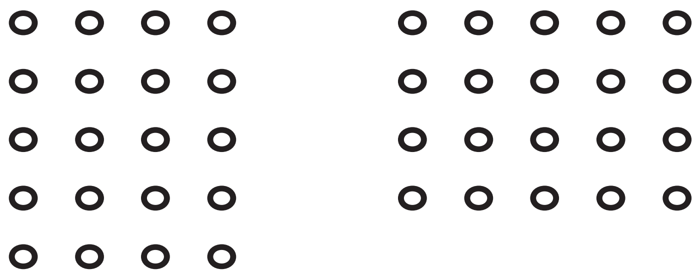
|X| There are discontinuities in design space

E.g. in humans the deliberative and meta-management layers appear to have unique mechanisms and forms of representation, not found in other animals.

Can a chimp (or bonobo) think about the relation between mind and body?

Or learn about predicate calculus and modal logic?

Or see the structural correspondence between these two?



(I don’t know the answers - apes can do amazing things.)

It is not clear that all discontinuities are results of sequences of very small changes. Darwinian evolution might sometimes (rarely!) produce large useful changes. (DNA is a discrete structure.)

For more on all this see

http://www.cs.bham.ac.uk/research/cogaff/
(papers)

http://www.cs.bham.ac.uk/~axs/misc/talks
(slides for several talks, including this one)

http://www.cs.bham.ac.uk/research/poplog/freepoplog.html
(software tools for exploring hybrid architectures)