Cognition and language seminar Birmingham 24 Oct 2003

Requirements for visual/spatial reasoning

Aaron Sloman
http://www.cs.bham.ac.uk/˜axs

School of Computer Science
The University of Birmingham

These slides are available online as

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

THANKS

To the developers of Linux
and other free, portable, reliable, software systems,
e.g. Latex, Tgif, xdvi, ghostscript, Poplog/Pop-11, etc.
No Microsoft software required

Abstract

Many people have remarked that humans are able to reason using pictures or images, either using external media, e.g. when constructing a geometrical proof on paper, or ’in their heads’, or when looking at a complex object in order to work out how something happens. There has been much empirical research trying to establish whether such ’spatial’ reasoning really happens or not (e.g. mental rotation experiments) and also empirical research on whether spatial aids do or do not help with problem-solving, communication or learning. Usually such empirical research assumes that we understand what the questions mean, and investigates whether something happens or how long it takes, as opposed to what mechanisms or architectural features make it possible.

This talk is more about conceptual questions and design questions. The conceptual question is: What does it mean to say that someone is reasoning spatially or visually as opposed to logically or verbally or in some other way? The design question is what sorts of mechanisms and architectures are capable of doing it: what are the requirements for such systems? For instance it is often thought that actual construction, rotations and transformations of 2-D structures may somehow occur in spatial problem solving.

I shall try to show that the requirements for any such system to work are complex and subtle and involve, among other things, the ability to perceive affordances which are not themselves spatial entities but far more abstract. The talk will discuss some examples and present some half-baked ideas, though I don’t yet have a fully developed model. A preparatory question for people who are interested: Does Betty the hook-making Caledonian crow do spatial reasoning? http://news.bbc.co.uk/1/hi/sci/tech/2178920.stm http://users.ox.ac.uk/ kgroup/tools/tools_main.html

Some of the ideas of the talk are available online here:

http://www.cs.bham.ac.uk/research/cogaff/talks/#talk7

http://www.cs.bham.ac.uk/research/cogaff/talks/#talk19 (PDF and postscrtipt)

Visual/Spatial reasoning

Humans seem to have different sorts of reasoning capabilities.

Logical reasoning is one of them.

- Premisses
  - All fleas are insects
  - Horace is a flea
- Conclusion
  - Therefore Horace is an insect

Another special case is our ability to use diagrams and visual images, even in reasoning about very abstract mathematical problems, e.g. thinking about the complexity of a search strategy.

These do not depend on the syntactic forms and rules that characterise logical reasoning

They involve other modes of representation, with different structures and different types of transformations.

What modes? What transformations?

To answer that we need to a deeper understanding of what vision is.

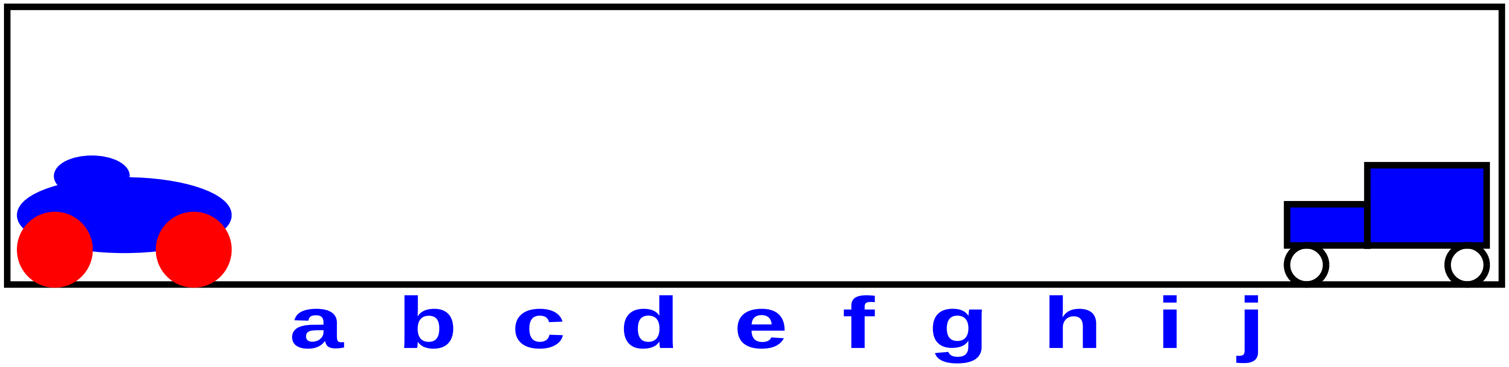
Towards some answers

- Vision is concerned with objects, properties relationsions.
- Vision is concerned with different levels of abstraction.
- Vision uses different ontologies, depending on what is seen.
- Different parts of a cognitive architecture use vision for different purposes – e.g. reactive, deliberative, meta-management/executive purposes.

- Common assumption: vision is concerned with what exists.
- Alternative view: much of vision is concerned with what does not exist but is possible.
- Compare Gibson on “affordances”.

What is seeing possibilities and impossibilities?

Colliding cars



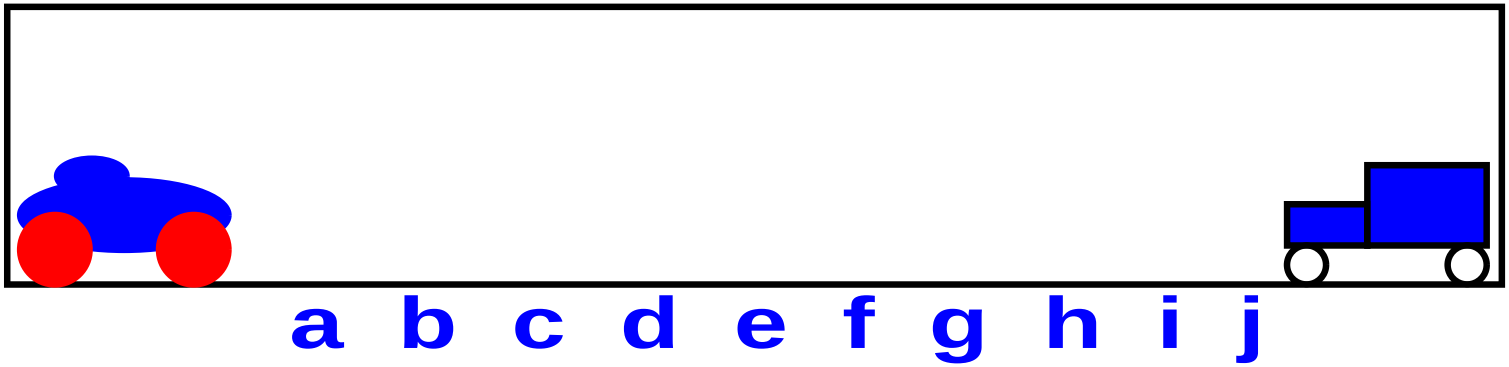
The two vehicles start moving towards one another at the same time.

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?

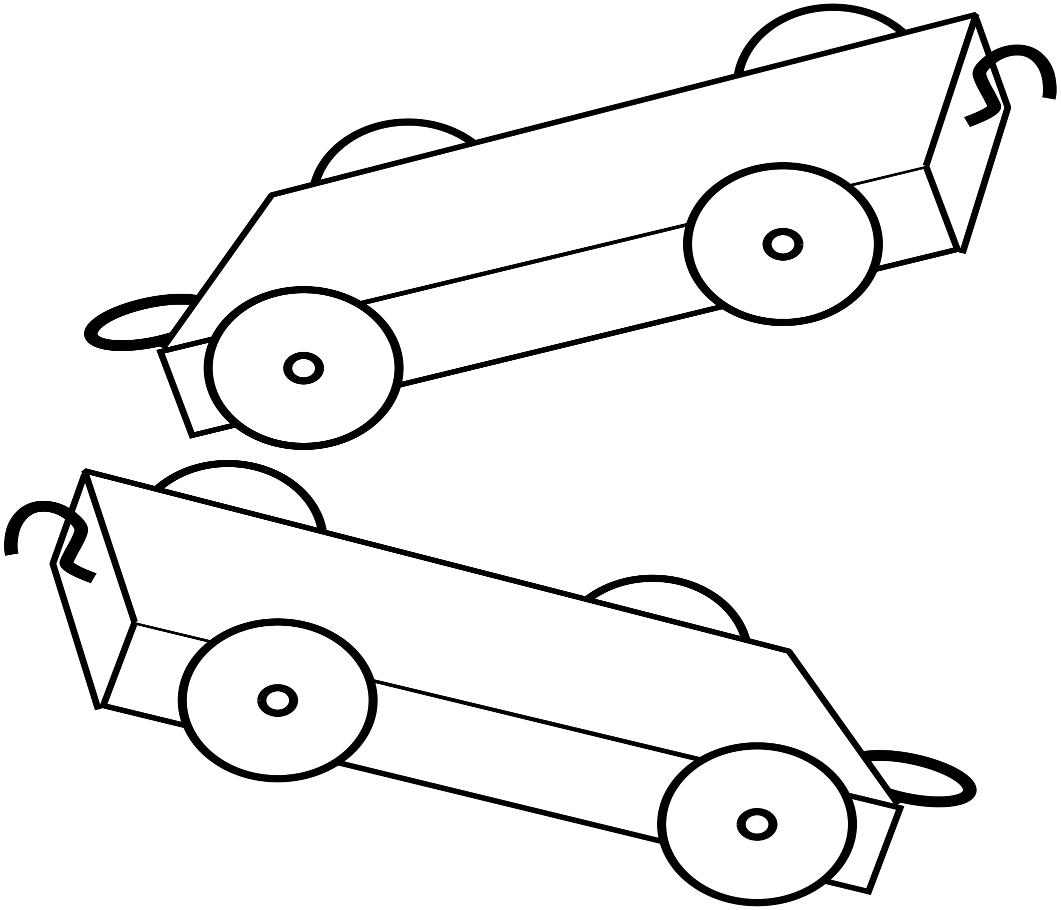
Where do you think a five year old will say they meet?

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

Me: Why?

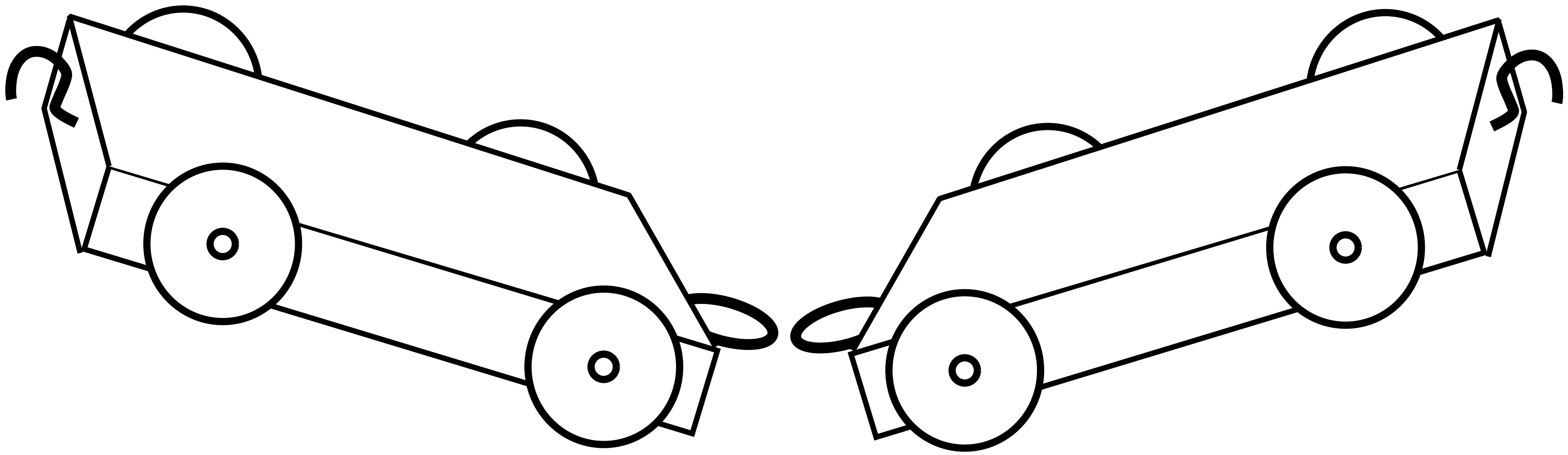
Child: It’s going faster so it will get there sooner.

Building trains



How would you have to move the trucks to join them together?

Why won’t this work?



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

What changes between a child not understanding and understanding?

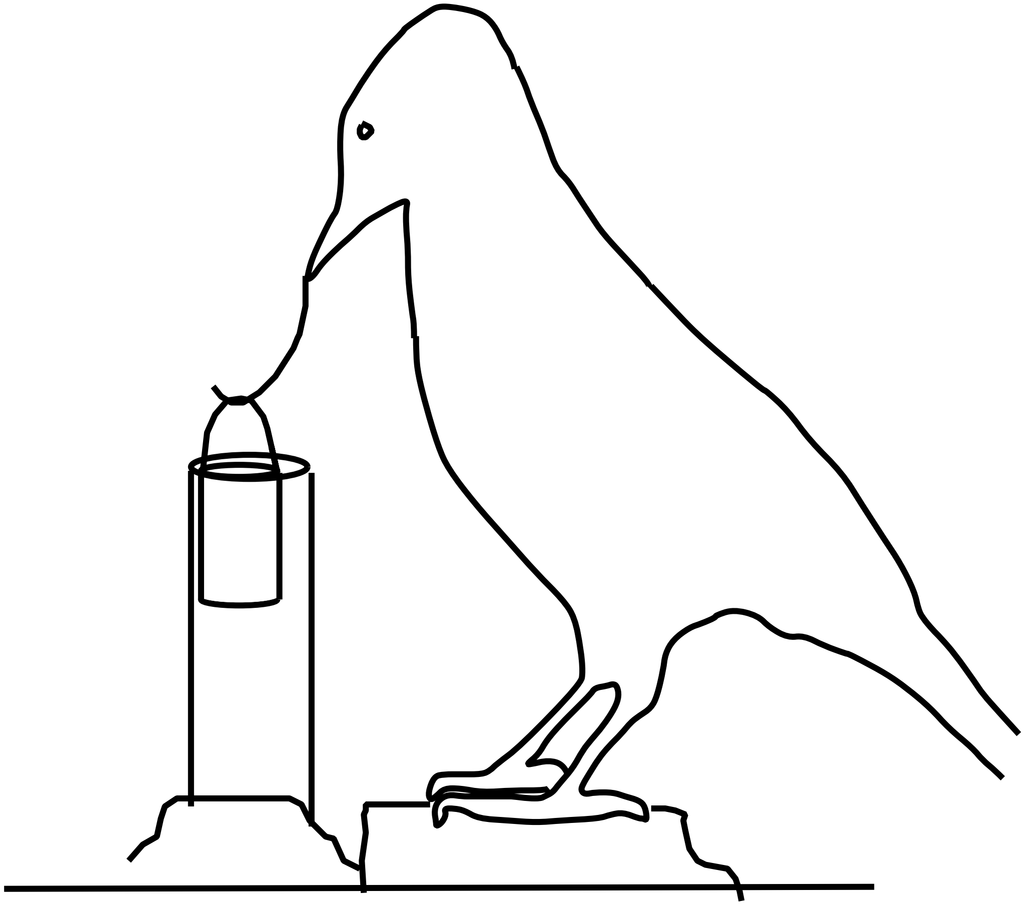
Show video of Josh (aged about 18 months) failing to understand the affordances in rings as opposed to hooks:

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

A few weeks later, he seemed to understand.

WHAT CHANGED?

Betty



Betty the hook-making crow.

See the video here:

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

Visual reasoning in adults

There are no points common to the triangle and the circle.

Suppose the circle and triangle change their size and shape and move around in the surface.

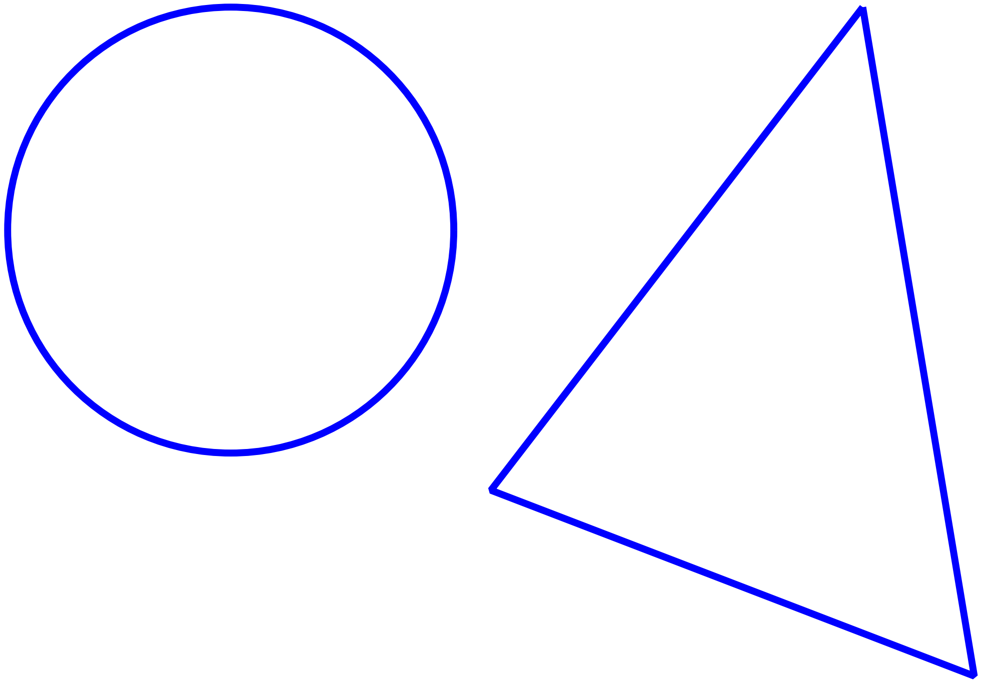
They could come into contact.

Clearly if a vertex touches the circle, or one side becomes a tangent to the circle, there will be one point common to both figures.

If one vertex moves inside the circle and the remainder of the triangle is outside the circle how many points are common to the circle and the triangle?

What are all the possible numbers of common points?

How do humans answer this question?



How many different numbers of contact points can there be?

This requires the ability to see empty space as containing possible paths of motion, and fixed objects as things that it is possible to move, rotate and deform

Visual reasoning in humans

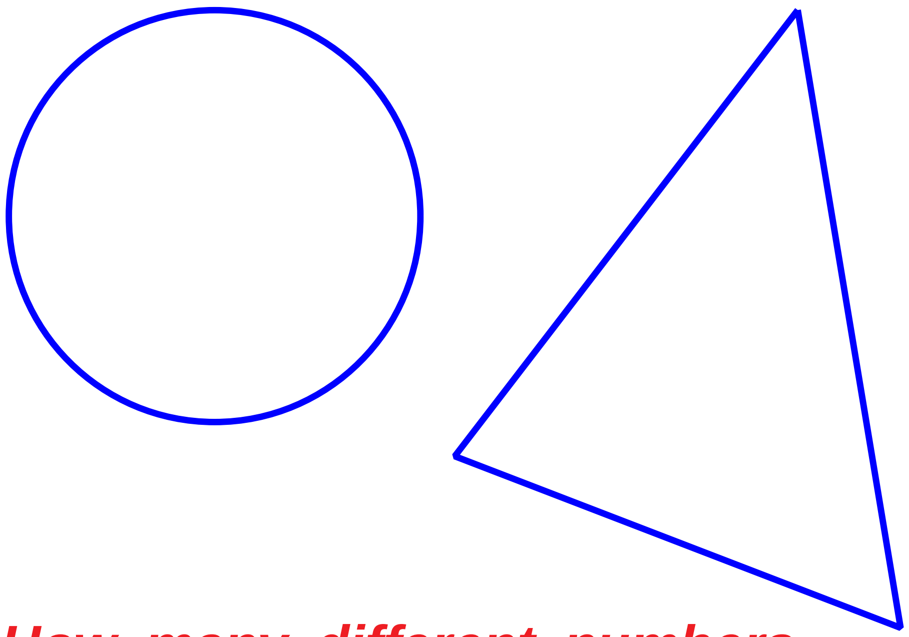
E.g. No points are common to the triangle and the circle.

Suppose the circle and triangle change their size and shape and move around in the surface.

They could come into contact.

If a vertex touches the circle, or one side becomes a tangent to the circle, there will be one point common to both figures. If one vertex moves into the circle and the rest of the triangle is outside the circle how many points are common to the circle and the triangle?

How do humans answer the question on the right?



How many different numbers of contact points can there be?

This requires the ability to see empty space as containing possible paths of motion, and fixed objects as things that it is possible to move, rotate and deform. 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.

What are Affordances?

Affordances are not “objective” properties intrinsic to physical configurations.

Rather, they are relational features dependent on the perceiver’s

- Common or likely goals and needs
- Capabilities for action (physical design + software)
- Constraints and preferences (avoid stress, injury)

Affordances in a complex scene can, as suggested above, be construed as

(1) *sets of sets* of counterfactual conditionals,
(2) *spatially indexed*: different sets are associated with different parts of objects.

But this still leaves open what sorts of mechanisms, architectural configurations of mechanisms and forms of representation can be explain the ability to perceive them.

In particular it is not clear how animal brains represent counterfactual possibilities.

Do they ‘use’ something like modal logics (steedman02), or is there some powerful new form of representation waiting to be discovered? Could they be built out of condition-action rules?

Multi-scale multi-purpose spatial understanding

For mathematicians, space is homogeneous: the same in all places and at all scales, but not for most animals, including most humans:

- Manipulable-scale space
- Reachable, mostly visible scale space
- Domestic-scale space
- Urban-scale space
- Geographical-scale space
- Migratory-scale spaces

What is seen is related to possibilities for action. But the possibilities for action are different at the different scales: hence one source of non-homogeneity.

CONJECTURE

Humans and other animals with spatial skills have to learn about all these different aspects of space, location, structure, motion, time, and causation separately (though some aspects may have been ‘learnt’ by evolution and transmitted genetically.)

The integrated view comes much later by a quite distinct sort of learning process, using rare architectural mechanisms.

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.

Using affordances

A purely reactive system like our simulated sheepdog implicitly detects affordances insofar as what it perceives immediately triggers appropriate actions (internal and external).

A deliberative system, like the crow or a typical human being, has to be able not only to perceive and react to affordances, but also to be able to represent alternative possible reactions and the new affordances that they will generate.

This hypothetical reasoning about what might be or what might have been the case is most sophisticated in human beings, but it seems that some other animals are capable of it, to varying degrees.
(Humans also differ in their deliberative capabilities.)

The collection of affordances available in a typical cluttered spatial environment is huge.

Being able to detect them and select relevant ones and being able to learn about new affordances, are among the great achievements of ordinary human minds

THE CAPABILITIES ARE NOT ALL THERE AT BIRTH

Machines can vary in what they operate on, use, or manipulate:

1. Matter
2. Energy
3. Information (of many kinds)

- Scientists and engineers have built and studied the first two types of machines for centuries. Newton provided the first deep systematisation of this knowledge.
- Until recently we have designed and built only very primitive machines of the third type, and our understanding of those machines is still limited.
- Evolution 'designed' and built a fantastic variety of machines of all three types – with amazing versatility and power long before human scientists and engineers ever began to think about them.
  - Biological organisms are information-processing machines, but vary enormously in their information-processing capabilities.
  - There are myriad biological niches supporting enormously varied designs, with many trade-offs that we do not yet understand (e.g. trade-offs between cheapness and sophistication of individuals).
  - The vast majority of organisms have special-purpose information-processing mechanisms with nothing remotely like the abstractness and generality of TMs
  - A tiny subset of species (including humans) developed more abstract, more general, more powerful systems. Turing's ideas about TMs were derived from his intuitions about this aspect of human minds. But at best that's a small part of a human mind.

Understanding all this requires us to think more about architectures than about algorithms.
See: http://www.cs.bham.ac.uk/research/cogaff/talks/

We don’t yet know how many ways there are to represent and manipulate information

Even if logic can be used to represent anything at all (which is doubtful) it should not be assumed that it is always best form of representation for every task.

E.g. why do we use maps, musical notation, flow-charts?

Note that not all maps and pictures need to preserve metrical properties of what they represent. Example: London tube map.

Exercises

- Consider some of the tasks for which you use maps.
- Investigate how those tasks would change if, instead, you had logical descriptions of everything represented in the map.
- For which tasks is logic better?
- Why do maps have to have symbols on them,
  and why do they need an index of places (gazetteer)?

A little philosophical history

Hume thought that there are only two forms of knowledge:

- analytic (based only on definitions and logical deduction)
- empirical (requiring observation and experiment)

(Hume thought everything else was nonsense,
e.g. theology, and the books should be burnt!)

Immanuel Kant, in his *Critique of Pure Reason* argued against Hume. He claimed

- there is also non-analytic, non-empirical (synthetic apriori) knowledge,
- e.g. in mathematics: we can gain new insights by examining structures and their properties, including the use of diagrams (on paper or in the mind) in doing Euclidean geometry.

Even Frege agreed, partly

Gottlob Frege was one of the greatest logicians of all time. He invented predicate calculus and was the first to clarify the notion of higher order functions (functions of functions) e.g. quantifiers.

Frege disagreed with Kant about arithmetic, since he thought arithmetic could be reduced to logic. I.e.

- He thought the concepts of arithmetic could be defined in terms of purely logical concepts
- He thought that all the truths of arithmetic could be derived purely from the axioms of logic plus the aforementioned definitions.

  (In modern jargon, that would prove that the truths of arithmetic are all ‘analytic’, not synthetic.)

Frege tried to prove that all of arithmetic could be derived from logic

He tried to demonstrate this as follows:

- He tried to show that all arithmetical concepts could be defined in terms of purely logical concepts.
- He tried to show that all arithmetical truths could be proved on the basis only of logical axioms, rules and the definitions.

But the attempt fell foul of Russell’s paradox.

Let S be the set of all sets that are not members of themselves.

Then

S is a member of S $\Longleftrightarrow$ S is NOT a member of S

Therefore:

S is and is not a member of S

Subsequent attempts to fix this this remain controversial.

But Frege thought Kant was right about truths of geometry being non-empirical, and non-analytic, i.e. he too thought they were *synthetic apriori* truths.

(Show pythagoras demo.)

Focus on modes of reasoning not kinds of truths

Disagreement over who was right still continues.

- One problem with all this is that there may be different ways of arriving at the same (or very closely related) results.

  E.g. you may be able to prove using something like Frege’s method or a later variant, that this is a truth of logic

  $7 + 5 = 12$

- But there may be another proof of a very similar result, where the concepts are defined not in terms of logic, but in terms of visually detectable structures.

- So perhaps we should not ask about the status of the propositions themselves, i.e. attempting to distinguish different kinds of truths, but instead talk primarily about different means of proof.
- The things proved by different means may look the same, but actually have subtly different contents.

KANT’S EXAMPLE: 7 + 5 = 12

It is obvious that this equivalence is preserved if you spatially rearrange the blobs within their groups:

```
ooo          o        oooo
ooo   +      o    =   oooo
o          ooo        oooo
```

Or is it?
How can it be obvious?
Can you *see* such a general fact?
How?

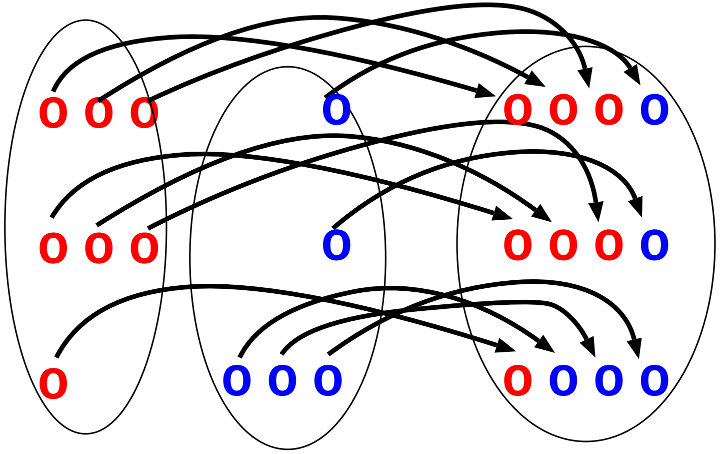
What sort of equivalence are we talking about?

I.e. what does “=” mean here?

Obviously we have to grasp the notion of a “one to one mapping”.

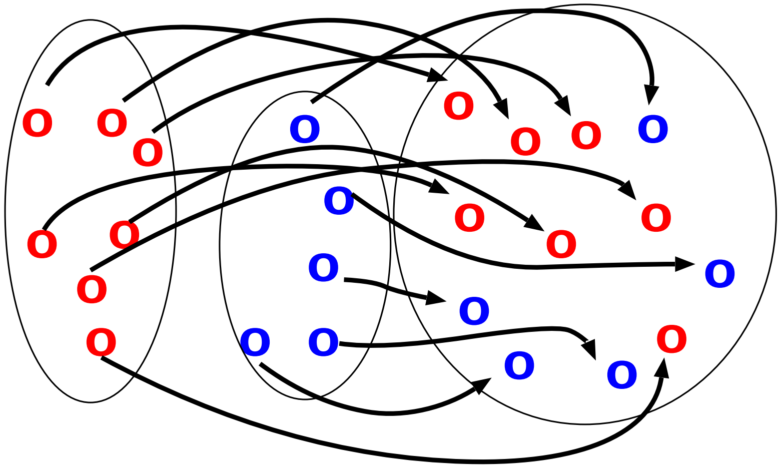
That can be defined logically, but the idea can also be understood by people who do not yet grasp the logical apparatus required to define the notion of a bijection.

SEEING that 7 + 5 = 12



Join up corresponding items with imaginary strings.

Then rearrange the items, leaving the strings attached.



Is it ‘obvious’ that the correspondence defined by the strings will be preserved even if the strings get tangled by the rearrangement?

Is it ‘obvious’ that the same mode of reasong will also work for other additions, e.g.

777 + 555 = 1332

What does a diagrammatic proof prove?

Even if this is a way of discovering a truth,

is it the same truth as is proved starting from purely logical definitions and axioms?

Or do we have several *different* concepts of number?

And different arithmetical truths expressed using the same forms?

How many different meanings are there for:

777 + 555 = 1332

Consider

- numbers of oranges,
- numbers of numbers,
- adding areas,
- adding operations on numbers,

etc.

Both right?

Perhaps Kant and Frege were both right:

- there are some analytic truths of arithmetic and some non-analytic ones.
- they differ in that they use concepts that are understood (defined) in different ways, using different conceptual resources (and different cognitive mechanisms)
- but there is a very strong structural correspondence between them.
- Understanding that relationship would be part of the task of philosophy of mathematics, and of AI.
- Maybe educators also need to understand it, in order to teach mathematics effectively.

Who can do it?

What are the information-processing requirements for

- being able to grasp structural relationships
- being able to visualise transformations
  (e.g. seeing that dots can be rearranged),
- being able to grasp higher level generalisations that are preserved by such transformations
  (e.g. seeing that the one-to-one correspondence is preserved)?

Can a dog or a monkey see such truths?

Can a two year old child?

If not, then why not?

What changes when a child becomes able to see them?

What do other animals (and some humans) lack that prevents them learning to see such things?

(Kant thought that sort of ability was innate.
I don't think he ever considered other animals, or dreamed of the possibility of intelligent machines.
Had he been born 200 years later he would been an AI enthusiast.)

A partial analysis

The ability to see arithmetical truths using a grasp of spatial structures requires at least:

- **The ability see the spatial structures involved in the proof.**
- **The ability to see possibilities for variation in those structures (e.g. rearrangements of components, as in logical reasoning)**
- **The ability to grasp features that are invariant under those rearrangements.**
- **The ability to grasp a collection of structures, possibilities for change, and invariants, involved in a sequence of configurations or maybe even a continuous transformation covering a range of configurations.**

Compare logic:

Everything is discrete and all syntactic composition involves function application

Information processing architectures that are able to support human visual reasoning capabilities will be far more complex than those required for logical reasoning.

Reasoning is part of what we call seeing

That specification of requirements for visual reasoning is very vague, and would not be easy to mechanise in a general way in an AI program.

Those features are involved in much of our ordinary use of seeing, e.g. when we think about possible ways of rearranging furniture in a room, or possible routes from here to there.

This requires grasping structures, seeing possibilities for change, seeing "affordances" etc.

(See the talks on vision here

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

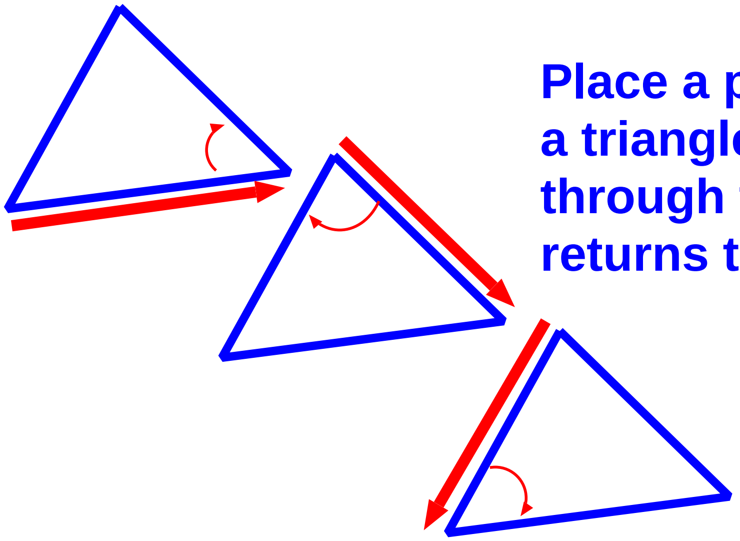
my 'Actual Possibilities' paper in http://www.cs.bham.ac.uk/research/cogaff/)

When we have understood how human (and animal) visual capabilities evolved, what their functions are, what sorts of architectures support them, and what sorts of mechanisms and representations are used in those architectures, then perhaps we shall be in a far better position to understand what our reasoning capabilities are.

Example 2: The angles of a triangle

Why do the angles of a triangle add up to a straight line?

Most people cannot remember the proof they were taught as children. Mary Ensor (former Sussex student) invented this highly memorable proof, while teaching mathematics:



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?

Of course, like the standard proof, this gives wrong results on a sphere. (Visualise a triangle with three right angles.)

But such a proof still teaches one something.

Example 3: List processing

Consider the list-process operation rotate, which when given an integer N and a list L, creates a new list by moving N items in turn from the beginning to the end of the original list. Most programmers could easily define such a procedure. So:

rotate(2, [a b c d e]) = [c d e a b])

What can you say about the following general expression, where L is any list:

rotate(length(L), L) = ???

Does anyone reason about this by starting from logical axioms and definitions and using logical deduction?

That’s what many AI theorem provers would do, e.g. using structural induction over binary trees, etc.

But do they prove the same thing as we *see* to be obviously true?

We abstract away from the tree-structured implementation of lists and reason about them as spatially manipulable linear structures.

This is important for teaching list processing.

... continued

If you can see the operations in your mind’s eye the result is obvious.

But seeing a particular case is one thing, e.g.

rotate(5, [a b c d e]) = [a b c d e])

Seeing the general principle is quite another.

Can you see in your mind’s eye the core properties of a spatio-temporal process that is common to a large (infinite) set of cases, covering all possible lists?

Example 4: Transfinite ordinals

We can visualise, and reason about, things that are impossible to see – including the infinitely thin and infinitely long lines of Euclidean geometry. More recently, mathematicians have discovered transfinite ordinals: infinite discrete structures.

The simplest one is the familiar sequence of natural numbers

TO1: $1, 2, 3, 4, ...$

From that we can easily derive new ones by rearranging them. E.g. take all the even numbers followed by all the odd ones:

TO2: $2, 4, 6, ..., 1, 3, 5, ...$

Questions:

1. In TO1, is it possible to move between the numbers 398 and 300002 without passing the number 1057?
2. Same question for TO2
3. In TO2, is it possible to move between the numbers 398 and 300003 without passing the number 1057?
4. In TO2, going from 999999 to 222, which will you meet first, 777 or 888 ?

How do you think about such questions?

More on transfinite ordinals

Consider the transfinite ordinal obtained by taking each prime number in turn, and for each of them forming a sequence of all their powers, followed by the sequence of the powers of the next prime.

$$2^1, 2^2, 2^3, \ldots\ldots \quad 3^1, 3^2, 3^3 \ldots\ldots \quad 5^1, 5^2, 5^3, \ldots\ldots \quad 7^1, 7^2, 7^3, \ldots\ldots$$

Is the sequence well-ordered? I.e. does every arbitrary sub-sequence of the sequence have an earliest member?

What if we reversed the order of the powers of every second prime?

$$2^1, 2^2, 2^3, \ldots\ldots \quad \ldots\ldots 3^3, 3^2, 3^1, \quad 5^1, 5^2, 5^3, \ldots\ldots \quad \ldots\ldots 7^3, 7^2, 7^1, \ldots\ldots$$

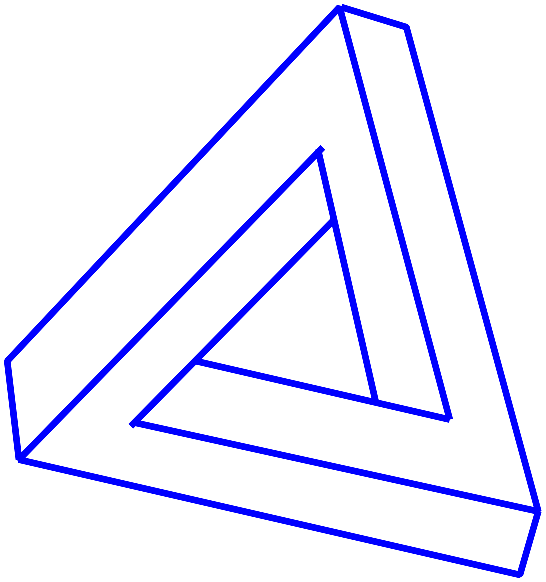
It is possible to produce logical axioms characterising these infinite ordinals, and derive answers to questions using those axioms.

Humans can use both logical and visual modes of thinking about such structures. Something about how our visual systems evolved produced, as a side-effect, the ability to visualise not only naturally occurring but also physically impossible objects.

Two physically impossible objects

Just as logic does not prevent us forming contradictory assertions, the syntax of spatial representations does not prevent us depicting incoherent spatial structures. E.g.

- **The Penrose triangle:**



- **A round square:**

(viewed from the edge)

There are many more examples

There are many uses of human spatial reasoning.

- Knowing where to look for an object thrown over a wall
- Route planning
- Engineering design
- Seeing how to assemble a toy crane from components in a meccano set: including seeing that a particular assembly will NOT work (e.g. because a quadrilateral is not rigid, or because a triangle is).
- The use of spatial concepts in many programming designs, e.g. the notion of a search space.
- Understanding the relationships between spatially defined search strategies and syntactically specified programs, e.g. depth first search uses a stack and breadth first search a queue.
- Many uses in physics, e.g. the notion of a phase space, a trajectory in phase space, an attractor in phase space
- Many examples in control engineering, e.g. the notion of a feedback “loop”

Can we replicate, or even explain, these human capabilities?

Conjecture:

Our visual reasoning capabilities in realms like mathematics are in part a side-effect of the interactions between different components in an animal architecture.
Those components evolved at different times, to serve different sorts of purposes.

For more on these ideas see these papers:

http://www.cs.bham.ac.uk/research/cogaff/sloman.ppsn00.pdf

http://www.cs.bham.ac.uk/research/cogaff/sloman.bmvc01.pdf

http://www.cs.bham.ac.uk/research/cogaff/Sloman.actual.possibilities.pdf

http://www.cs.bham.ac.uk/ axs/misc/draft/esslli.pdf

http://www.cs.bham.ac.uk/research/cogaff/sloman.diagbook.pdf

(or Postscript versions)

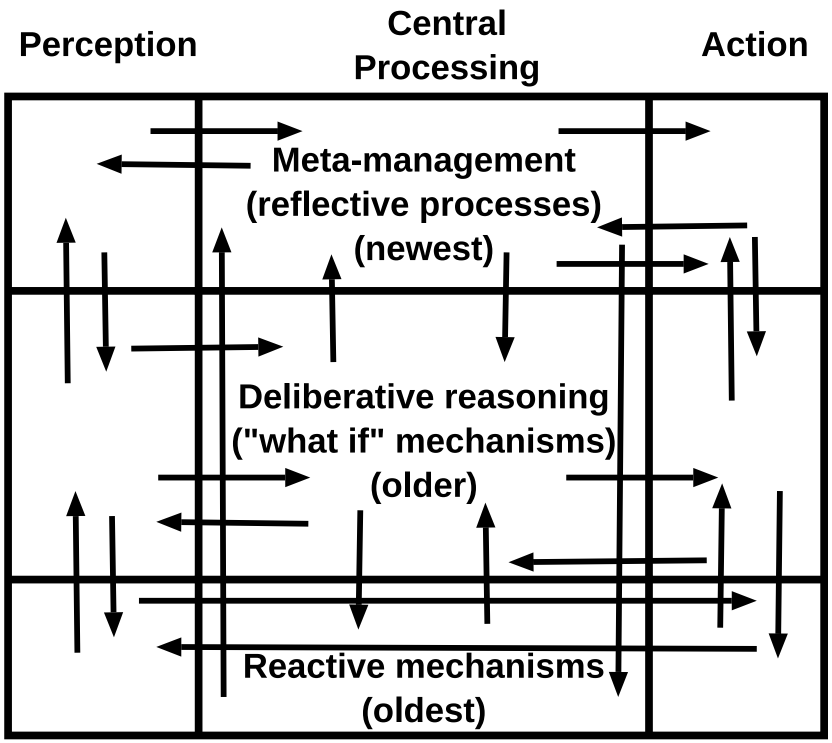
There are also many books and journal articles on diagrammatic or spatial reasoning.

E.g. The latest issue of the AI Journal (Vol 145, Nos 1–2, April 2003) has an article by M. Anderson and R. McCartney

Towards an explanatory theory

The Birmingham Cognition and Affect project has been investigating architectures for “complete” intelligent agents.

It is conjectured that the various architectural components can be located in the nine main regions of the COGAFF architecture schema:



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

...continued

- There are reactive mechanisms involved both in triggering a variety of responses, including internal and external reflexes, and in managing tight feedback control loops (e.g. posture control, grasping accurately)
- There are deliberative mechanisms that require the ability to reason in a discrete fashion about possible futures, possible unobserved entities (the back of that tree), possible explanations (possible pasts) possible sequences of actions (plans for oneself, predictions regarding others).
- There are self-monitoring (meta-management, reflective) capabilities that involve the ability to monitor, categorise, evaluate, and perhaps modify internal states, processes and strategies.

These evolved at different times, are present to different degrees in different organisms, develop at different stages in human beings.

They involve multiple forms of representation, with multiple mechanisms for manipulating those representations, and multiple varieties of semantics.

Interactions between concurrent modules

If there are so many concurrently active components

- doing different tasks,
- using different representations (possibly expressing information derived from the same source, such as a retina)
- including tasks where one module monitors another

then powerful new capabilities (and bugs) can emerge from the interactions.

Investigating all this requires a long hard slog, including attempting to characterise with far greater precision than before the types of visual competences found in humans, and the varieties of forms of representation and information manipulation that can occur.

Maybe logic will have some useful role as part of this.

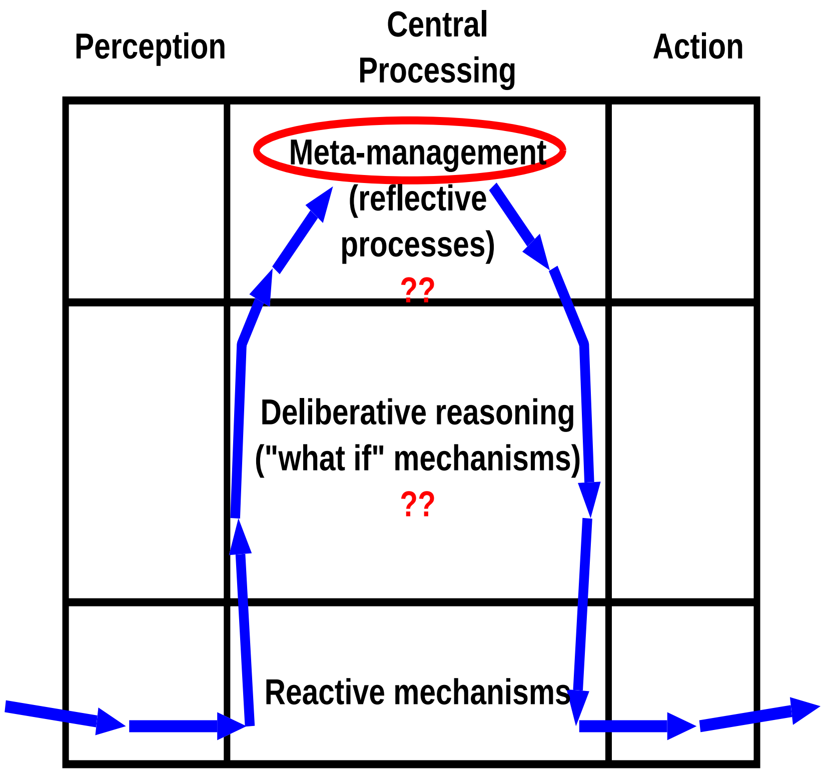
Layered architectures have many variants

With different subdivisions and interpretations of subdivisions, and different patterns of control and information flow.

Different principles of subdivision in layered architectures

- **evolutionary stages**
- **levels of abstraction,**
- **control-hierarchy,**
  (Top-down vs multi-directional control)
- **information flow**
  (e.g. the popular ‘Omega’ Ω model of information flow, described below.)

The “Omega” model of information flow



Perception

Central Processing

Action

Meta-management (reflective processes) ??

Deliberative reasoning ("what if" mechanisms) ??

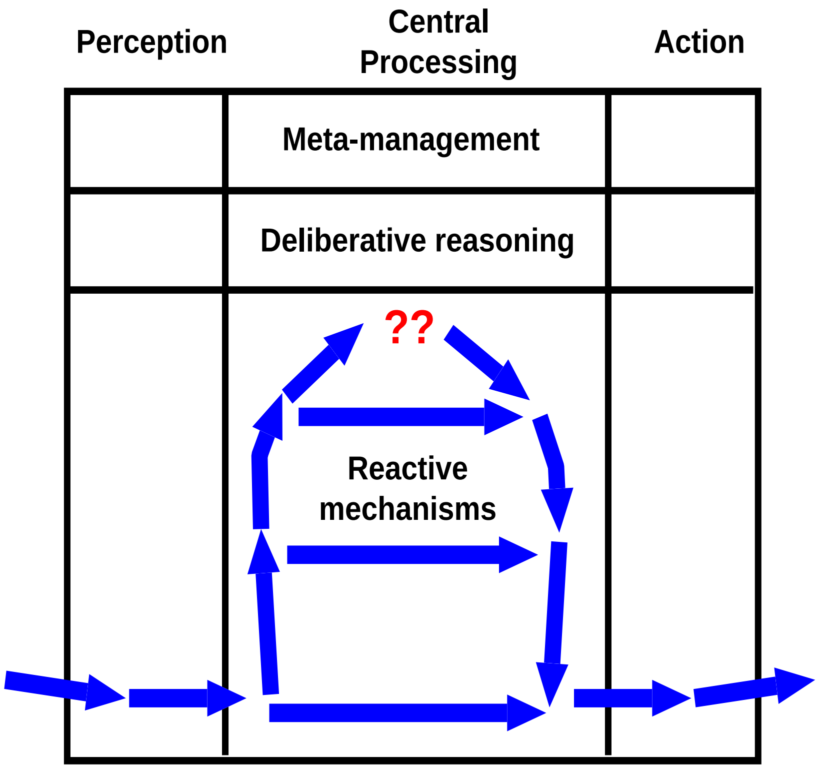
Reactive mechanisms

Rejects layered concurrent perceptual and action towers separate from central tower.

There are many variants, e.g. the “contention scheduling” model. (Shallice, Norman, Cooper, Albus.)

Some authors propose a “will” at the top of the omega.

Another variant (Brooks): Subsumption architectures



Here all the processing is assumed to be reactive, though there are several layers of reactive processing, including adaptive mechanisms.

Supporters deny that animals (even humans) use deliberative mechanisms.

Yet they somehow get to overseas conferences?

What are the functions of vision?

We now try to explain why a tower of perceptual mechanisms may be required for some organisms.

What are the functions of vision?

- **Recognition of objects?**
- **Segmentation/grouping?**
- **Feedback control for many actions.**

What are the functions of vision?

The functions of vision include:

- Segment the image (or scene) and recognize the objects distinguished

  And maybe the processes too, where there’s motion.

- Compute distance to contact in every direction.
- Provide feedback for action
- Provide triggers for action
- Provide a low-level summary of the 2-D features of the image, leaving it to central non-visual processes to draw conclusions
- As above and also to provide a low-level summary of the 3-D geometric and physical features of the scene, leaving it to central non-visual processes to draw conclusions. (Marr)

What’s left out?

SENSING AND ACTING CAN BE ARBITRARILY SOPHISTICATED

- Don’t regard sensors and motors as mere transducers.
- They can have sophisticated information-processing architectures.

  E.g. perception and action can each be hierarchically organised with concurrent interacting sub-systems.

  Think of the differences between

  – perceiving edges, optical flow, texture gradients

  – perceiving chairs, tables, support relations

  – perceiving happiness, surprise, anger, which way someone is looking.

  – perceiving affordances

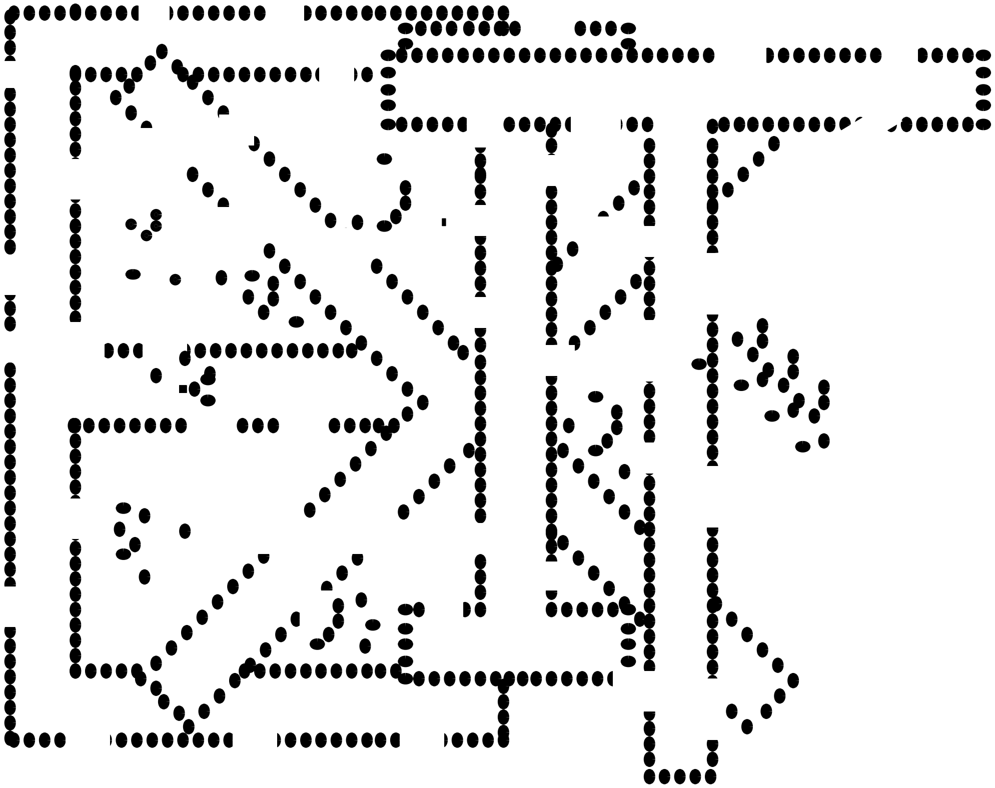
    - Positive affordances and negative affordances
    - Possibilities for and constraints on actions
    - Consequences of possible actions

Perceiving affordances involves going beyond perceiving what exists in the environment to seeing what is possible or impossible, and what the consequences might be.

What do you see?

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.



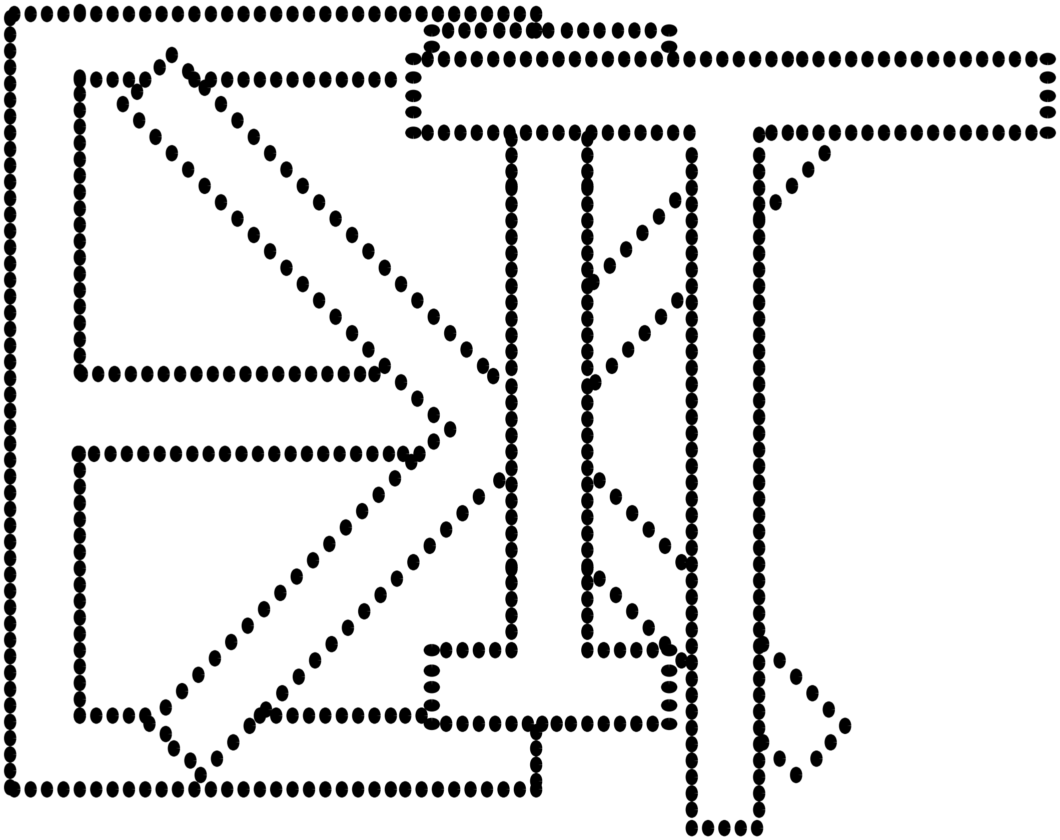
Real seeing is often much harder than the tasks most artificial vision systems can perform at present.

It is also, in its way, much harder than many of the tasks presented by psychologists to subjects in vision laboratories
(selected for suitability for repeatable laboratory experiments)

Some work done in the 1970s: POPEYE

The Popeye project (using Pop2) 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/

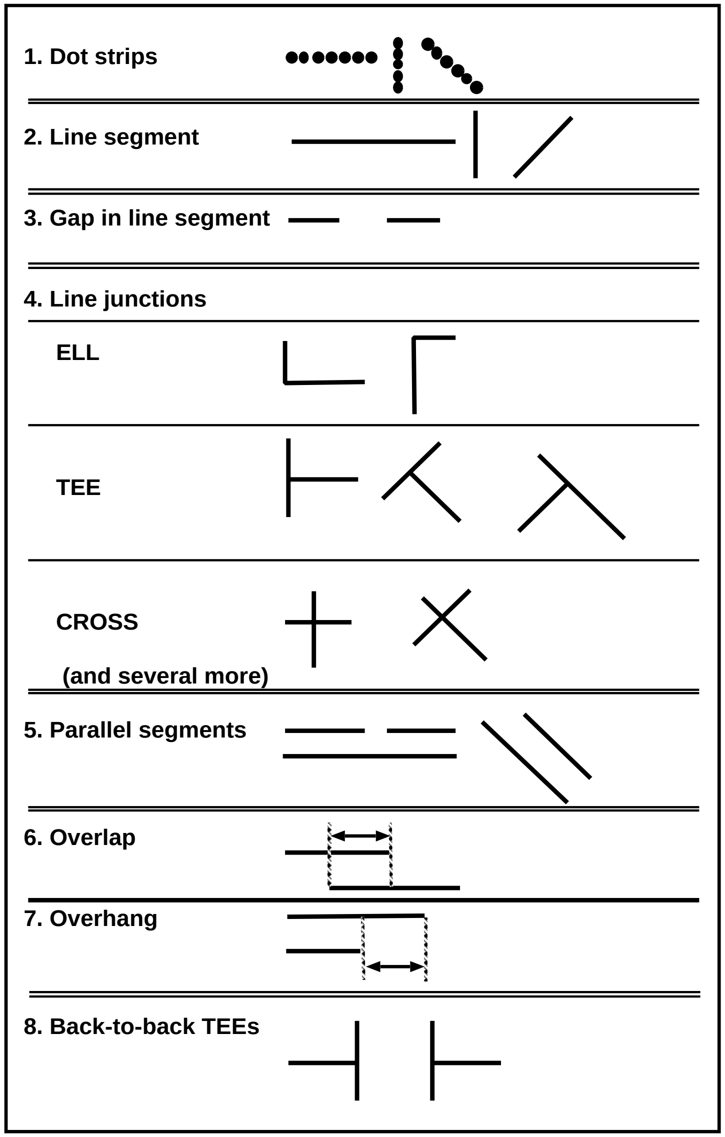
Conjecture:
Assemble fragments at different levels of abstraction

We seem to make use of structures of different sorts,

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

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.

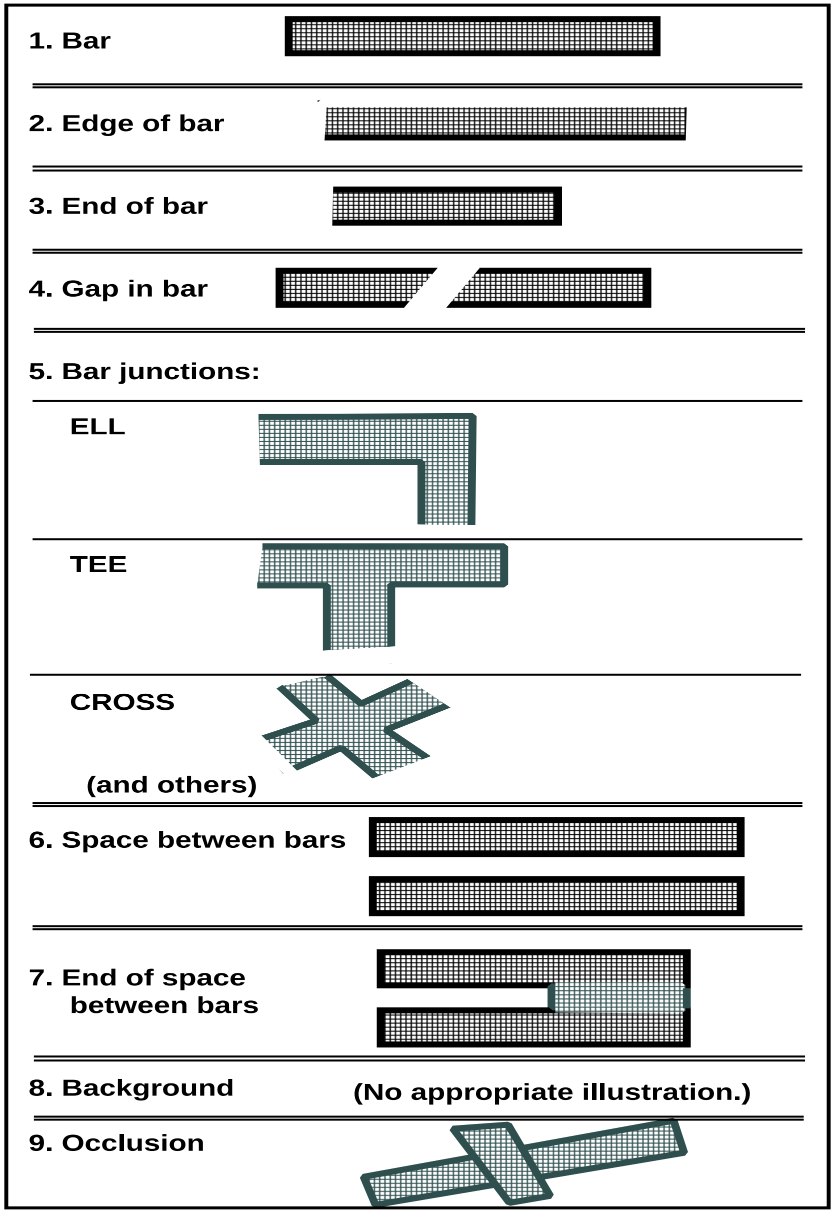
Here we have some of the fragments at the level of configurations of dots, and the next abstraction level, configurations of continuous line segments



Useful fragments at one level 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.



1. Bar

2. Edge of bar

3. End of bar

4. Gap in bar

5. Bar junctions:

ELL

TEE

CROSS

(and others)

6. Space between bars

7. End of space between bars

8. Background (No appropriate illustration.)

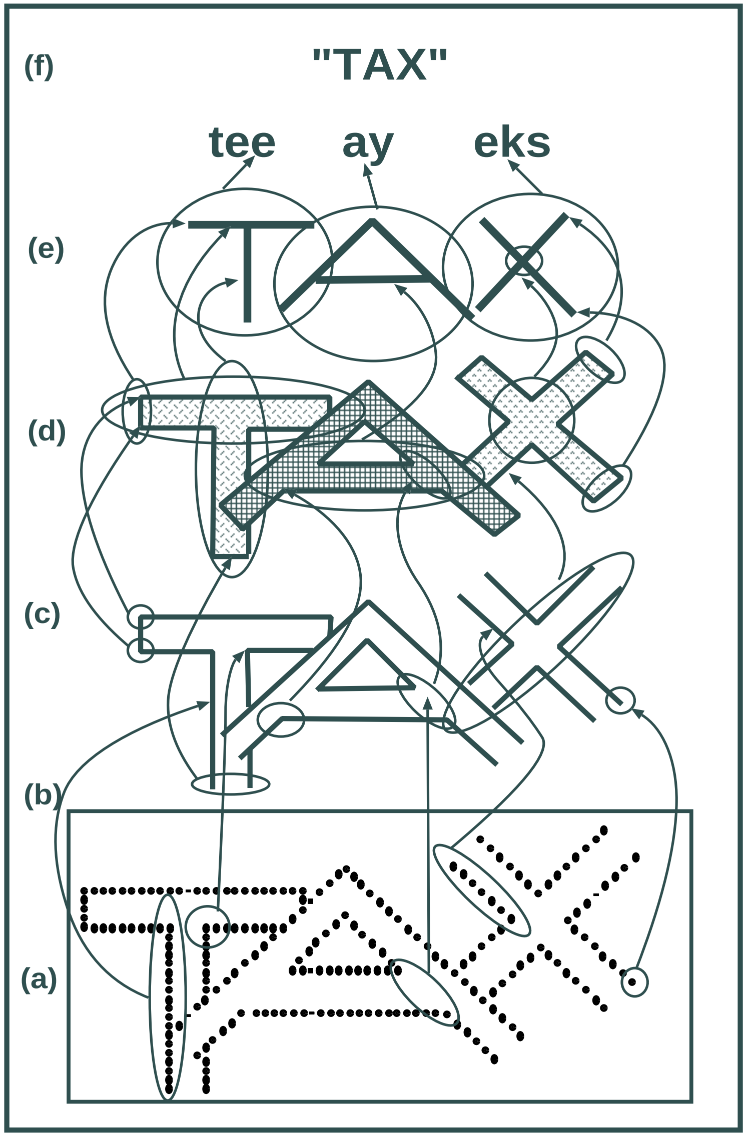
9. Occlusion

Aspects of perception: multiple levels of structure

Layers of interpretation of a 2-D dot pattern.

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 there is also interpretation, i.e. mapping to a new ontology.**
- **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

Putting it all together

CONJECTURE: concurrent perceptual processing occurs at many different levels of abstraction.

- Sub-systems at different levels can 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) will reach a decision before lower levels had finished processing.
- Sometimes a speedy high level decision will be wrong!

  It can be corrected later as information flows upwards.
- Different low level styles (dots, dashes, colour-boundaries) can feed into the same higher level domains: hence different styles of depiction may receive the same high level interpretation.

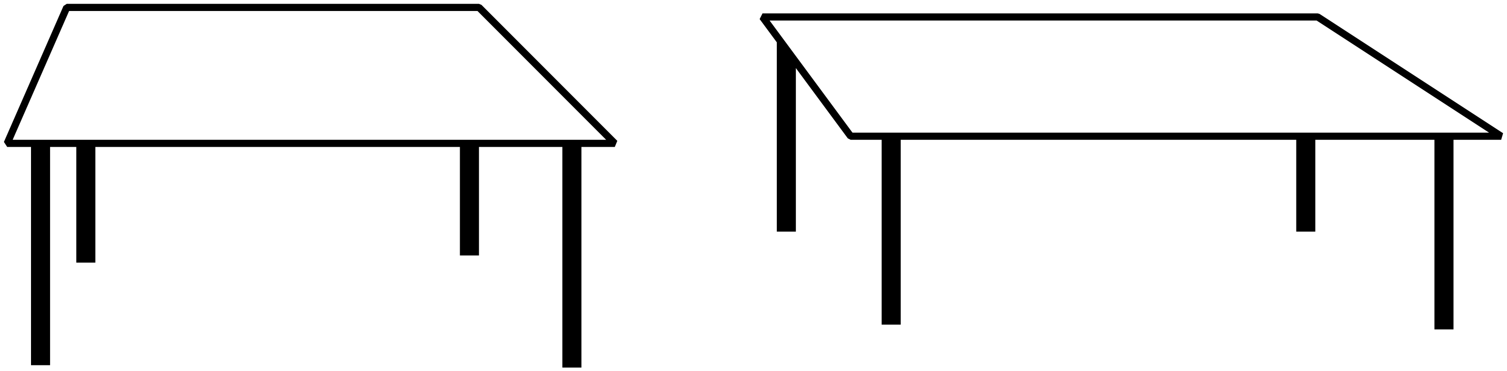
  Perhaps a network of neural nets could learn such things?

  But could they represent appropriate structural information?

Another example: seeing 3-D structure in 2-D optic arrays

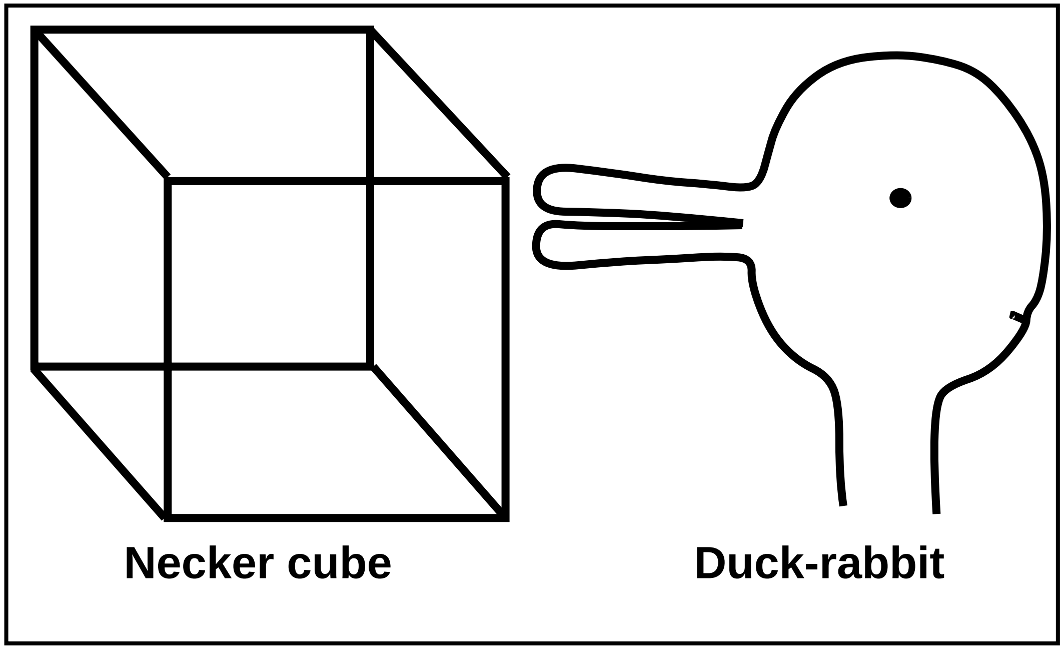
We can interpret 2-D patterns of illumination on our retinas as percepts of a 3-D world.

E.g. seeing these as two views of the same 3-D object



Things get harder when objects have irregular shapes, are flexible, floppy, and constantly moving and changing shape.

Visual ambiguities are not always geometrical ambiguities



How can we see the same 2-D visual input in different ways?

When the cube flips, the perceived change is purely geometrical.

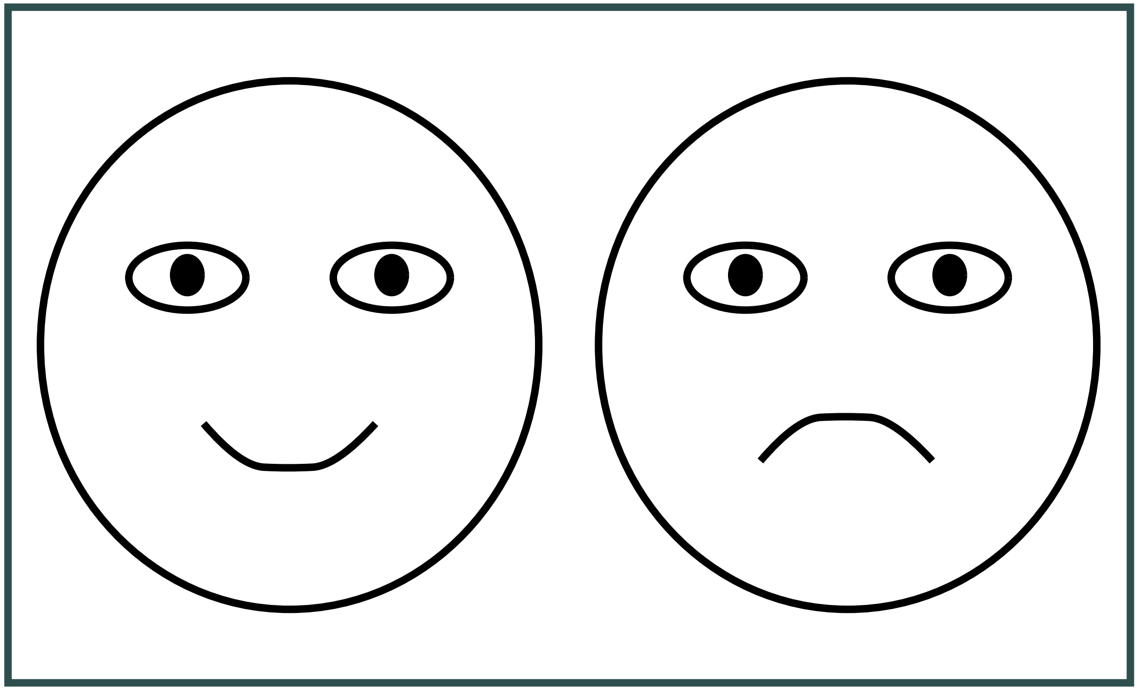
When the duck-rabbit flips, far more subtle things change. (What things?)

There are many more things to explain, including: perceiving motion, seeing how something works, experiencing visual pleasure, etc.

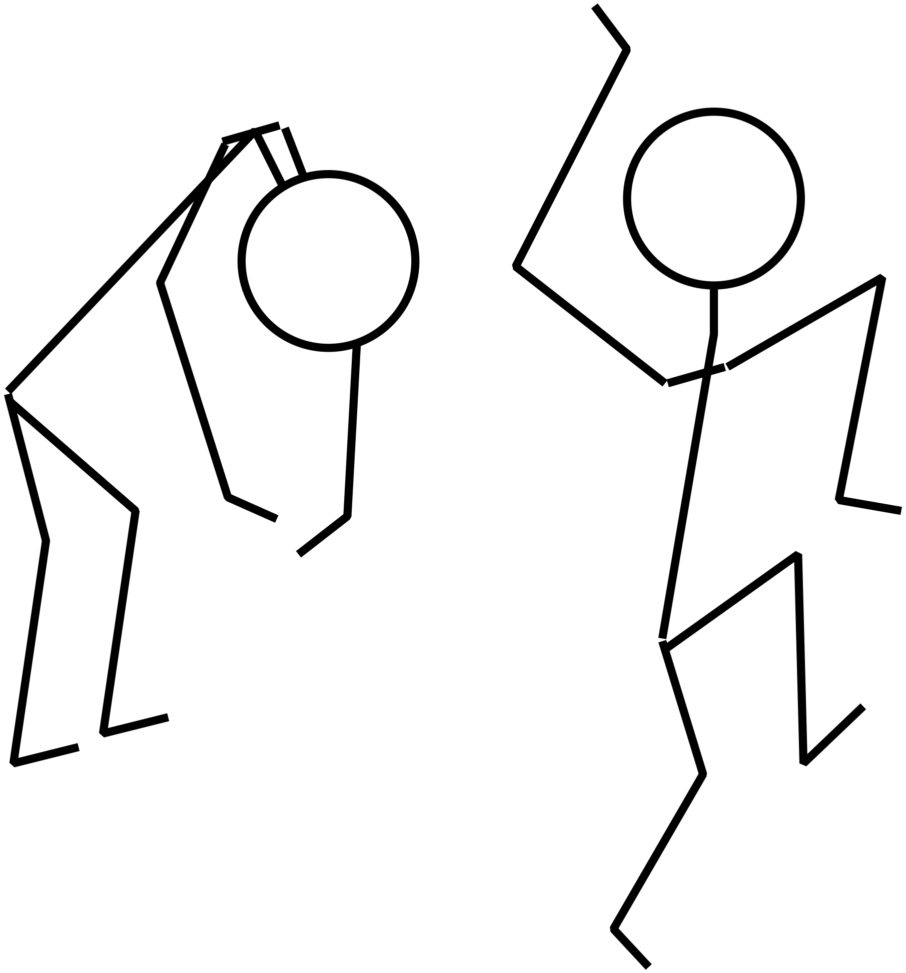
Another kind of non-spatial percept

How can one information-processing system see another as happy or sad?

What does it mean to say that it does?



Seeing emotions in postures



What does it mean to say that affective states are SEEN – not just INFERRED from or associated with the image?

Why would the ability to perceive mental states evolve?

Think about it:

If you are likely to be eaten by X what is more important for you to perceive:

- The shape and motion of X’s body?

  OR

- Whether X is hungry?
- Whether X can see you?

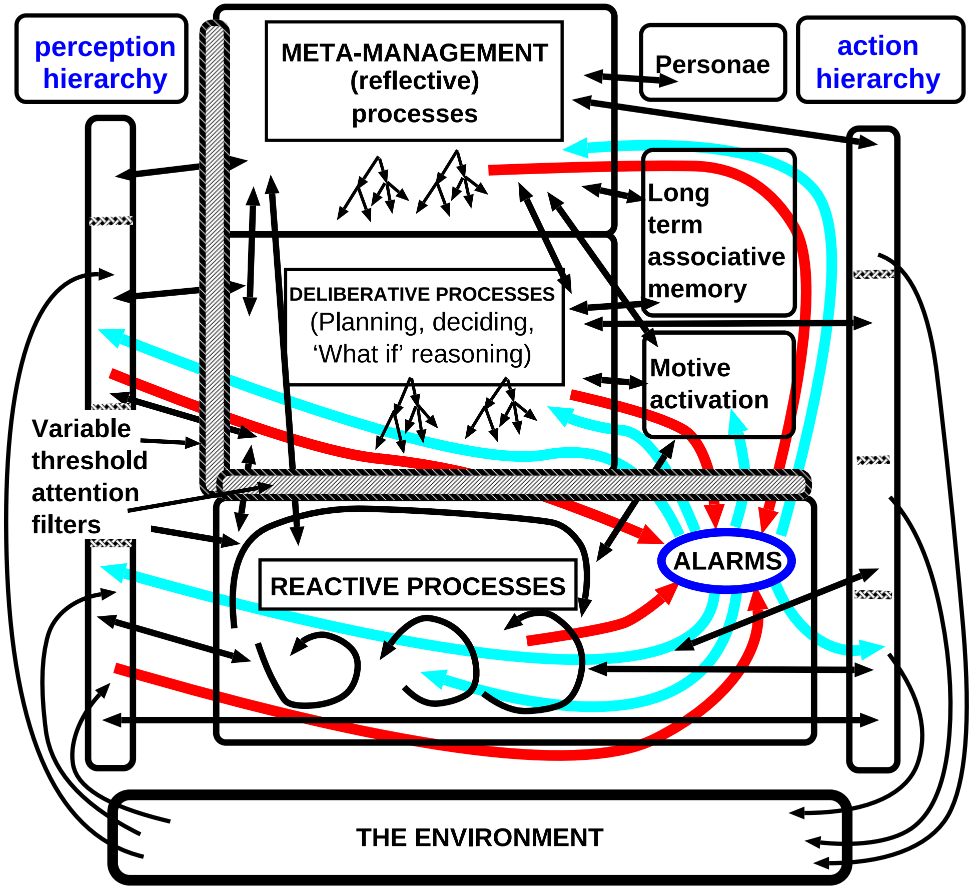
Primitive implicit theories of mind probably evolved long before anyone was able to talk about theories of mind.

Compare other intelligent primates?

(Evolution solved the “other minds” problem before there were any philosophers to notice the problem.)

There was also pressure in some species to evolve self-descriptions at a high level of abstraction: meta-management

- Self monitoring of internal states and processes
- Self evaluation
- Self control (partial)


perception hierarchy
META-MANAGEMENT (reflective) processes
Personae
action hierarchy
Long term associative memory
DELIBERATIVE PROCESSES (Planning, deciding, 'What if' reasoning)
Motive activation
Variable threshold attention filters
ALARMS
REACTIVE PROCESSES
THE ENVIRONMENT

Seeing mental states

- If a meta-management system uses high level descriptions of information-processing states to describe itself, then perceptual systems can evolve to use the same sort of ontology to describe other agents.

- The reverse can happen too: perceptual concepts developed (by evolution or learning) to characterise other agents, might be used to extend the internal ontology for self-description

- Likewise the action systems of agents may evolve high level behaviours to express internal information-processing states e.g. to conspecifics, or rivals, etc.

Summary: Different objects of vision

- Different levels of structure (agglomerations in a domain
  - 2-D structure
  - 3-D structure
- 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)

All of the above are purely factual: the ontology determines what sorts of things can and cannot exist, independently of any perceiver: however there are some kinds of things that are characterised in terms of the relevance to an agent's needs, desires, capabilities, ...

So we need to extend our ontologies to incorporate:

- Ontologies involving possibilities of many kinds inherent in a situation: Affordances

REMEMBER BETTY

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.

Apologies

In fact I believe there are practically no adequate theories yet about how any of this works

There are fragments of theories, but it is not clear that the fragments can be put together to form a coherent whole with the right explanatory properties.

Looking into brains won’t necessarily help any more than looking into computers will tell you about the virtual machines running in them.

SO THIS IS REALLY JUST AN INTRODUCTION TO A LONG TERM RESEARCH PROGRAMME. VOLUNTEERS WELCOME.

There is of course some useful work that has been already done by others.