Some requirements for human-like visual systems, including seeing processes, structures, possibilities, affordances, causation and impossible objects.

Aaron Sloman

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

School of Computer Science, The University of Birmingham

With help from Jackie Chappell and colleagues on the CoSy project

These slides will be made accessible from here:

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

http://www.cs.bham.ac.uk/research/projects/cosy/papers/

Along with other related slide presentations and papers.

WARNING:

My slides have too much detail for presentations. They are intended to make sense if read online.

The problem

- Human researchers have only very recently begun to understand the variety of possible information processing systems.
- In contrast, for millions of years longer than we have been thinking about the problem, evolution has been exploring myriad designs.
- Those designs vary enormously both in their functionality and also in the mechanisms used to achieve that functionality – probably using more types of information-processing mechanism than we have thought of.
- Many people investigating natural information processing systems, especially humans, assume that we know more or less what they do, and the problem is to explain how they do it.
- But perhaps we know only a very restricted subset of what they do, and the main initial problem is to identify exactly what needs to be explained: we need to do a lot more requirements analysis than is usually done.
- For example, it is often assumed as unquestionable that all perception is merely part of a sensori-motor control system, and all learning is learning of sensorimotor contingencies: this ignores the role of ‘exosomatic’ ontologies

  (Compare Plato’s cave-dwellers seeing only shadows on the wall of the cave).
- A piecemeal approach may lead to false explanations: working models of partial functionality may be incapable of being extended to explain the rest.

John McCarthy on The Well Designed Child

Quotes from his unpublished online paper: ‘The well designed child’,

http://www-formal.stanford.edu/jmc/child1.html

McCarthy wrote:

> Evolution solved a different problem than that of starting a baby with no a priori assumptions.
>
> ...
>
> Animal behavior, including human intelligence, evolved to survive and succeed in this complex, partially observable and very slightly controllable world.
> The main features of this world have existed for several billion years and should not have to be learned anew by each person or animal.

Biological facts support McCarthy:

Most animals start life with most of the competences they need – e.g. deer that run with the herd soon after birth. There’s no blooming, buzzing confusion (William James)
So why not humans and other primates, hunting mammals, nest building birds, ...?
Perhaps we have not been asking the right questions about learning.

We need to understand the nature/nurture tradeoffs, much better than we currently do, and that includes understanding what resources, opportunities and selection pressures existed during the evolution of our precursors, and how evolution responded to them.

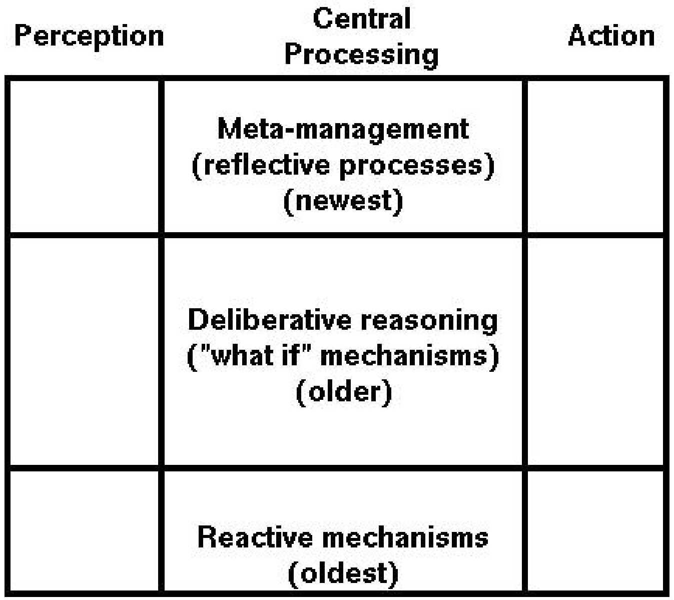
Making progress will require us to agree on terminology for expressing requirements and designs and cooperative exploration of the possibilities for both.

See the papers by Sloman and Chappell listed at end, including

The Altricial-Precocial Spectrum for Robots, in *Proceedings IJCAI’05*, and its sequels.
http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0502

The CogAff Schema (for designs or requirements)

Requirements for subsystems can refer to



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

- **Types of information handled:** (ontology used: processes, events, objects, relations, causes, functions, affordances, meta-semantic states, etc.)
- **Forms of representation:** (transient, persistent, continuous, discrete, Fregean (e.g. logical), spatial, diagrammatic, distributed, dynamical, compiled, interpreted...)
- **Uses of information:** (controlling, modulating, describing, planning, predicting, explaining, executing, teaching, questioning, instructing, communicating...)
- **Types of mechanism:** (many examples have already been explored – there may be lots more ...).
- **Ways of putting things together:** in an architecture or sub-architecture, dynamically, statically, with different forms of communication between sub-systems, and different modes of composition of information (e.g. vectors, graphs, logic, maps, models, ...)

In different organisms or machines, the 'boxes' contain different mechanisms, with different ontologies, functions and connectivity, with or without various forms of learning.
In some the architecture grows itself after birth.

In microbes, insects, etc., all information processing is linked to sensing and acting, and all or most information about the current environment is only in transient states, whereas for more sophisticated organisms, evolution discovered the massive combinatorial advantages of **exosomatic, amodally represented, ontologies,** allowing external, future, past, and hypothetical processes, events and causal relations to be represented.

Perhaps "mirror" neurones – should be called "exosomatic abstraction" neurons?

Can we use brain structure as a guide to architecture?

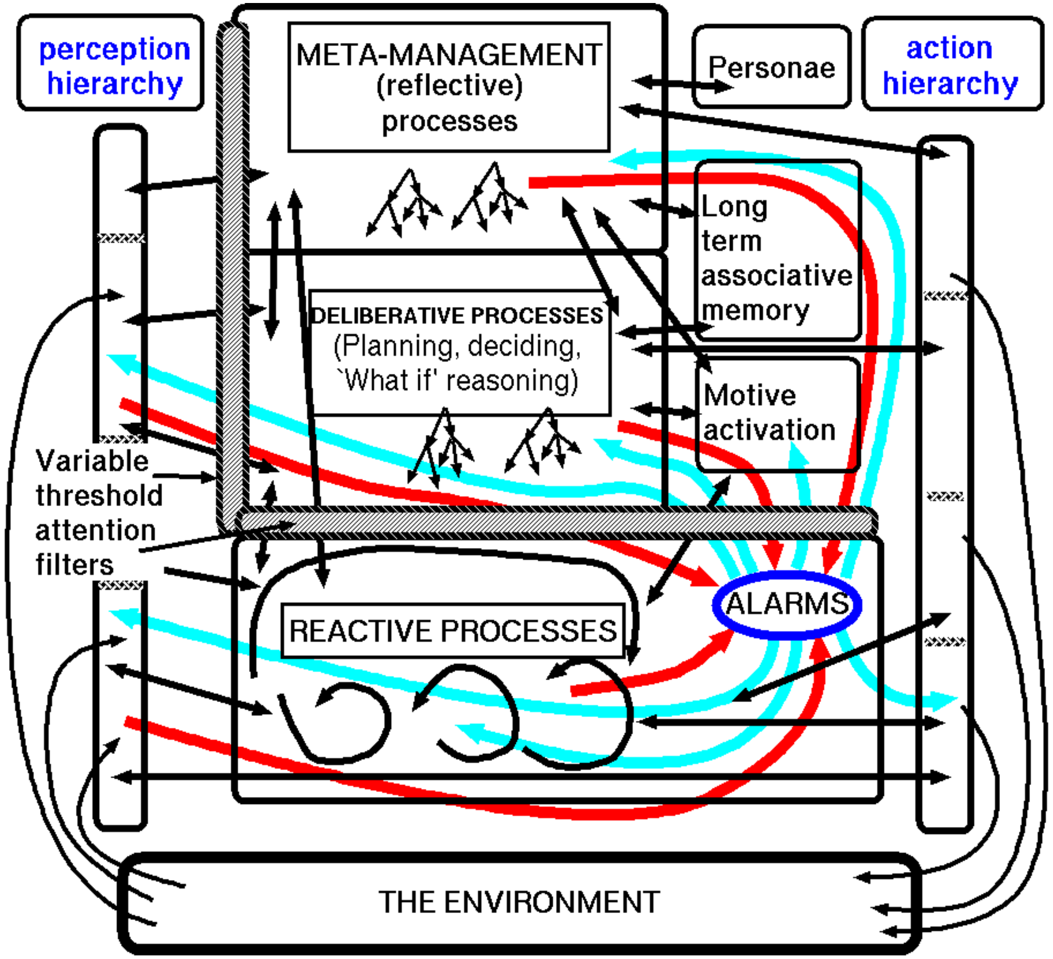
- Some people assume that any accurate information processing architecture must reflect brain structure.
- That could tempt them to assume that an architecture diagram should be labelled with known portions of brains.
- There are two problems with this:
  - it does not allow us to specify an information-processing architecture that is common to an animal with a brain and a machine that uses artificial computational mechanisms.
  - it does not allow for the possibility that high level functions don't map onto separable parts of brains but are implemented in a more abstract way (just as data-structures in a software system may not map onto fixed parts of a computer's physical memory, e.g. if virtual memory and garbage collection mechanisms are used).

Anyhow the attempt to specify an architecture that I talk about makes no assumptions about how the components map onto brain mechanisms. Rather it can be construed as a specification of a large collection of requirements for something to function as a certain kind of thing, e.g. an adult human, an infant human, a nest-building bird, or whatever we are trying to explain.

Of course, that does not mean brain science should be ignored. E.g. see the work of Arnold Trehub *The Cognitive Brain*. (MIT Press 1991 – now online: `http://www.people.umass.edu/trehub/`)

What are the functions of vision in humans and other animals?

Can we describe the functions of vision without producing a theory of the whole architecture and how vision relates to all parts of it?
The Birmingham Cognition and Affect project has many papers describing the CogAff schema for describing a wide variety of architectures for animals and robots, and H-Cogaff, a specific version that summarises some of the requirements for (adult) human-like systems.


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

The diagram summarises a collection of types of functionality in human-like systems.
An architecture for a collection of requirements.

The Cogaff web site is here: `http://www.cs.bham.ac.uk/research/projects/cogaff/`

The role of visual mechanisms in the architecture

The rest of this presentation focuses on aspects of the architecture and the capabilities involved in the architecture that relate to human vision.

The core assumption is that the visual subsystem concurrently sends information to (and may be partly controlled by) many other parts of the architecture that need different kinds of information and process it in different ways, for different purposes: e.g. online visual servoing vs acquiring factual information for future use.

This was described as a labyrinthine model and opposed to the modular models of Fodor, Marr and others, in Sloman 1989.

I suspect that:
everybody grossly underestimates the variety, complexity, and extendability of visual functions – and that probably includes me!

Some of the requirements

- Vision is primarily concerned with information about 3-D processes – of which 3-D structures are a special case
- In many perceived processes things are changing concurrently in different places and at different levels of abstraction.
- Vision requires use of different ontologies for different tasks.
- The visual system includes different sub-architectures that have strong links to different sub-architectures in the rest of the system.
- The different sub-architectures and the different ontologies may not all be available from birth: there are several kinds of extension (epigenetic bootstrapping).
- Some of the functionality requires forms of representation with features normally associated with human languages: rich structural variability and compositional semantics

  We call these “generalised languages” G-languages.

  These must have evolved before language, and must develop before language in humans

  See Sloman and Chappell (2007b)
- The speed at which vision works from very low level retinal stimulation to very high level perception and decision making probably requires mechanisms not yet envisaged in either AI or neuroscience.

Visual and spatial cognition

There is something deep and important about 3-D spatial perception and understanding

CONJECTURE:

The evolution of our ability to perceive and manipulate structured 3-D objects and processes has impacted profoundly on the forms of representation available to us, the ontologies we use in perceiving, thinking about and acting on the environment, and our understanding of causation.

Some of this is shared with other animals, including primates, hunting mammals, and some nest-building birds.

Explaining how this works is a pre-requisite for developing useful human-like robots (though that is not my main goal).

CONJECTURE

Mechanisms for perception of the 3-D environment penetrate deep into the cognitive system, and cognitive mechanisms penetrate deep into the perceptual subsystems.

Similar comments can be made about the relationships between central sub-systems and action sub-systems, which can also have a layered architecture.

Videos

Show some videos

crow (Betty) making a hook to get a bucket out of a tube

baby playing with spoon and yogurt

toddler trying to join up a toy train

Some of the things children try to do, fail to do, then later succeed in doing, provide windows into their minds.

Views on functions of vision

- There are many views of the nature and function(s) of vision, including the following:
  - Vision acquires/produces information about physical objects and their geometric and physical properties, relationships in the environment.

    (Marr and many others.)
  - Much recent work treats vision as a combination of recognition, classification and prediction – the latter sometimes used in tracking

    (often using classifications arbitrarily provided by a teacher, rather than being derived from the perceiver's needs and the environment).
  - Vision controls behaviour (obviously true – but only part of the truth)
  - Behaviour controls perception, including vision. (W.T.Powers)
  - Vision is unconscious inference (Helmholtz)
  - Vision is controlled hallucination (Max Clowes) Pretty close
- I'll try to present phenomena that require a richer deeper theory.

  It will be evident that an adequate theory must use many of the above ideas, and assemble them in new ways with some new details, especially emphasising perception of processes.

  The implications seem to be very important: both for studies of vision and cognition in animals (especially, but not only humans), and for attempts to understand requirements for robots with human-like capabilities.

Some themes in what follows

- Hierarchical structures (and processes)

  part-whole hierarchies vs taxonomic hierarchies vs ‘emergent’ ontological hierarchies

- Visual percepts can use different ontologies:

  geometrical structure, kinds of stuff, causal interactions, mental states, musical or mathematical meaning.

- Amodal exosomatic representations: grasping, manipulating

  Represented not in terms of patterns of sensor and motor signals but in terms of interacting 3-D surfaces and kinds of material.

- Multi-strand relationships, continuous, discrete, logical
- Multi-strand multi-level processes continuous, discrete, logical

  perceived concurrently

- Ontologies and forms of representation
- Labyrinthine vs modular architectures (1989 paper)

  Vision feeds information to many different parts of the architecture.

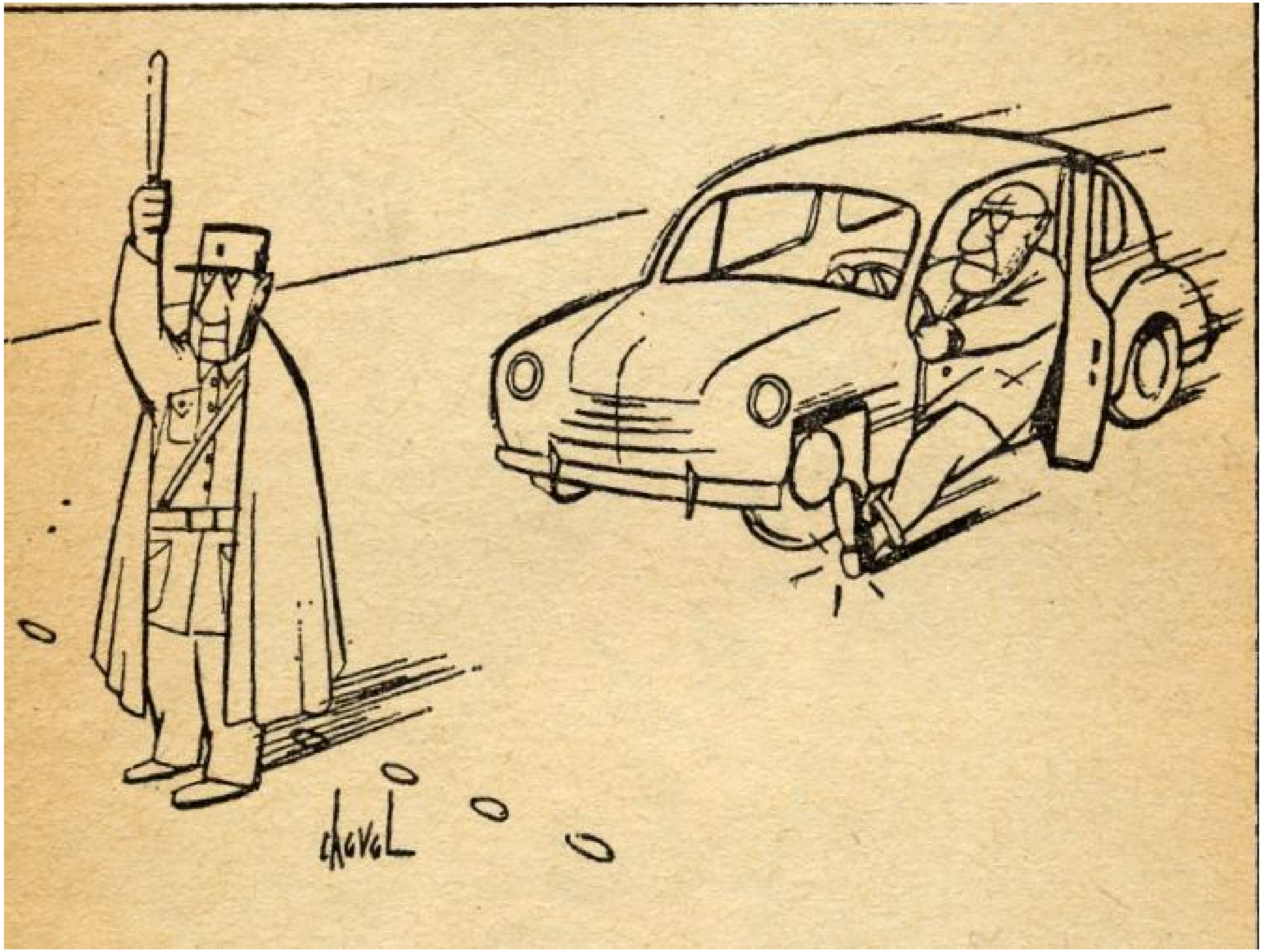
- In order to understand the spaces of possible requirements and designs we must do comparative studies:

  humans and other species
  humans at different stages
  humans from different cultures
  humans with different pathologies
  humans and many kinds of possible machine

Some entertaining examples

The next few slides illustrate that
even when confronted with static images
we may interpret them in terms of
3-D processes extending backwards
or forwards in time and using ontologies
in which causal interactions
are important.

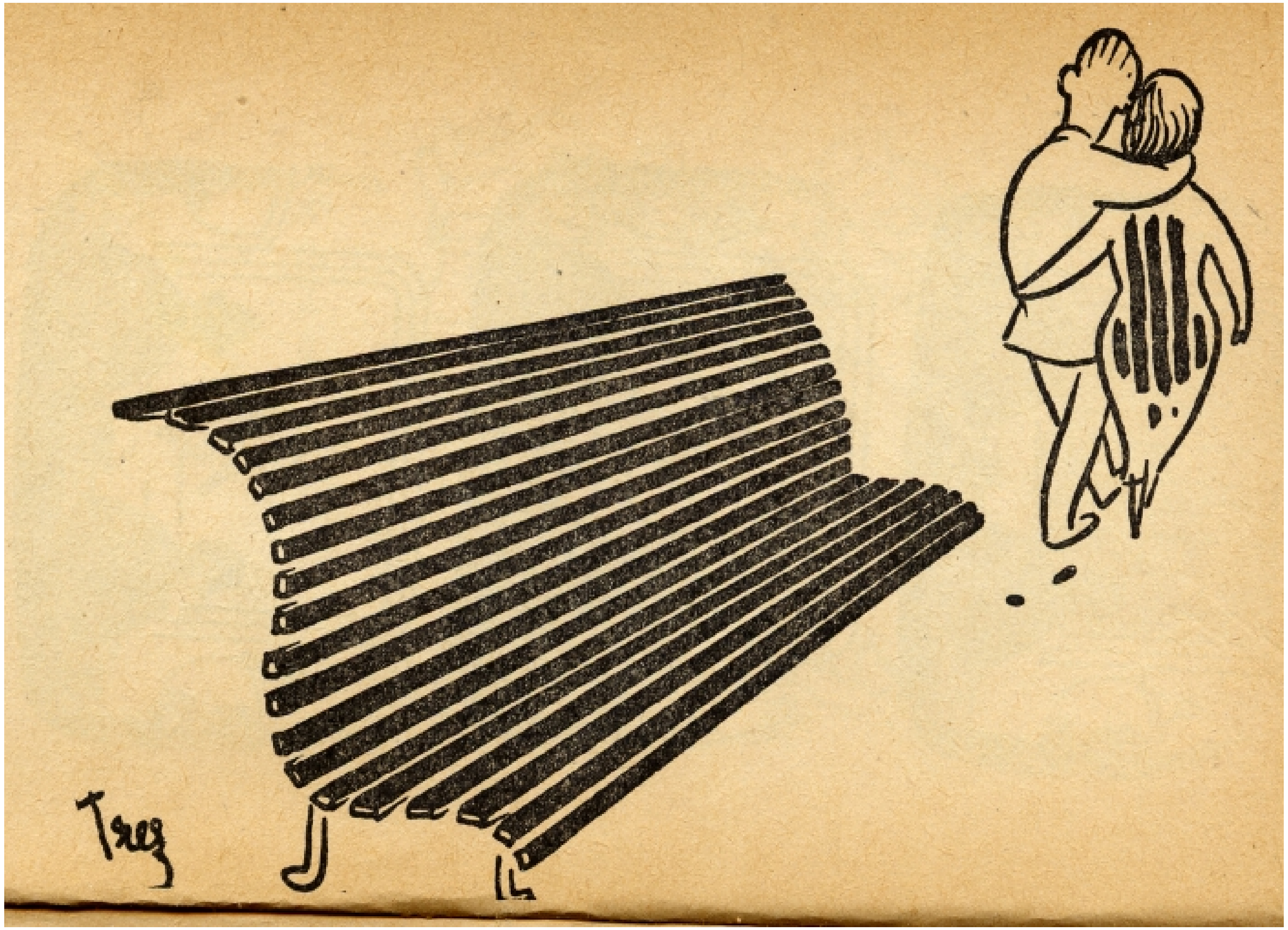
What do you see?



Perhaps you see a process extending to a future time?

And causal connections?

What do you see?

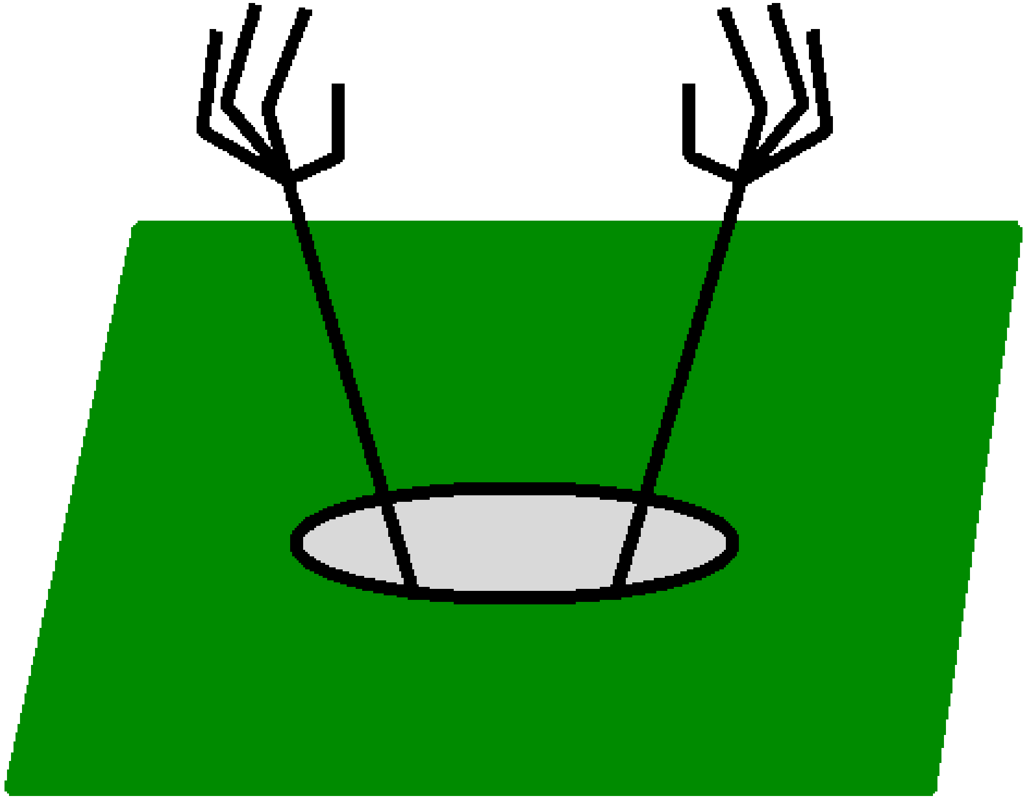


Various objects and relationships of different sorts

Perhaps you see a process starting at an earlier time?

And causal connections?

A Droodle: What do you see?



In many cases what you see is driven by the sensory data interacting with vast amounts of information about sorts of things that can exist in the world.

But droodles demonstrate that in some cases where sensory data do not suffice, a verbal hint can cause almost instantaneous reconstruction of the percept, using contents from an appropriate ontology.

See also `http://www.droodles.com/archive.html`

Verbal hint for the figure: ‘Early worm catches the bird’ or ‘Early bird catches very strong worm’

Example: Multiple perceptual routes

H-CogAff specifies multi-window perception and multi-window action, whereas many architectures assume peephole perception and action.

The visual and action sub-systems have architectural layers (evolved or developed) that handle ontologies at different levels of abstraction (including in some cases mental states of oneself and others), and have multiple connections to different sorts of central sub-systems, as well as to other sensory and motor subsystems.

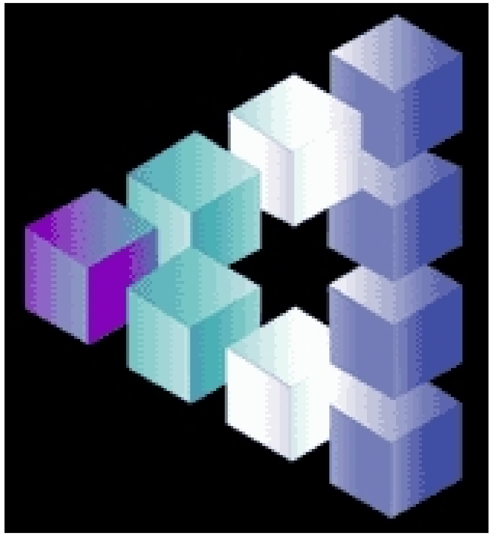
So, instead of one or two routes from vision, we have multiple routes,

e.g. to blinking reflexes, saccade generators, posture control subsystems, visual servoing mechanisms, question answering mechanisms, planning mechanisms, prediction mechanisms, explanation constructors, plan execution mechanisms, learning mechanisms (in several different architectural layers), alarm subsystems, communication mechanisms, social mechanisms.

Similar comments apply to connections with action sub-systems.

High level percepts can be inconsistent

(Picture by Reutersvard – before Penrose)



This tells us important things about the visual system – and some of the contents of visual consciousness.

What you see is not only what exists, but multiple affordances.
Think of all the things you can do with or between the little cubes.
Collections of affordances can be inconsistent: but not models of a scene.
If the picture were huge, you might never discover the impossibility
Compare Escher's pictures, e.g. the Waterfall.

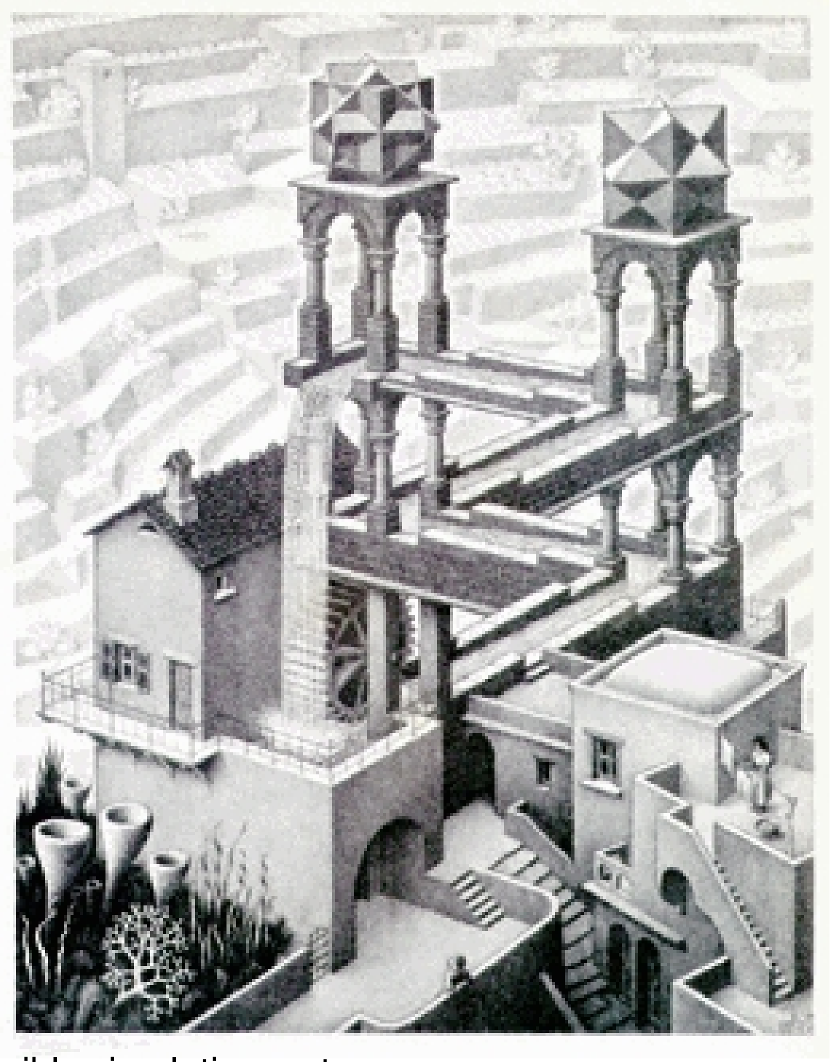
For more on visual processing see

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

Escher's Weird World

Many people have seen this picture by M.C. Escher:

a work of art, a mathematical exercise and a probe into the human visual system.



You probably see a variety of 3-D structures of various shapes and sizes, some in the distance and some nearby, some familiar, like flights of steps and a water wheel, others strange, e.g. some things in the 'garden'.

There are many parts you can imagine grasping, climbing over, leaning against, walking along, picking up, pushing over, etc.: you see both structure and affordances in the scene.

Yet all those internally consistent and intelligible details add up to a multiply contradictory global whole. What we see could not possibly exist.

There are several 'Penrose triangles' for instance, and impossibly circulating water.

Can you see the contradictions? They are not immediately obvious.

A visual percept cannot be a model

Models cannot be inconsistent

However if percepts are made up of fragments combined in a manner that does not correspond to full spatial integration then inconsistencies are possible.

E.g.

- A is bigger than B
- B is bigger than C
- C is bigger than A

or, more plausibly, a large collection of proto-affordances of different sorts, spatially located.

Why might the use of such a fragmented, though spatially related, collection of distinct interpretations of portions of the scene be desirable?

Because the very same scene needs to be perceivable in different ways, depending on current goals, interests, etc.

So it must be possible to switch different items of information in and out of the percept.

E.g. different affordances, different relationships, low level or high level details.

This is one form of attention switching.

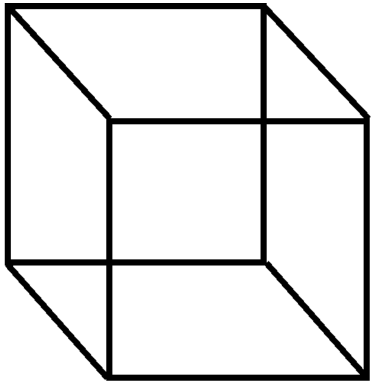
Seeing beyond the retina

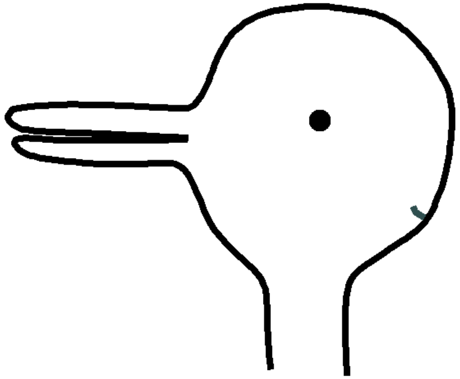
The fact that what we see is not all in our retinal images is shown by ambiguous images: what is seen flips between two different things though what's on the retina does not change.

Some things not in a retinal image are described as seen, not inferred: WHY?

Examples: the depth of the cube, parts of an animal, which way an animal is looking, ...

An ontology is involved in every percept – usually several ontologies.

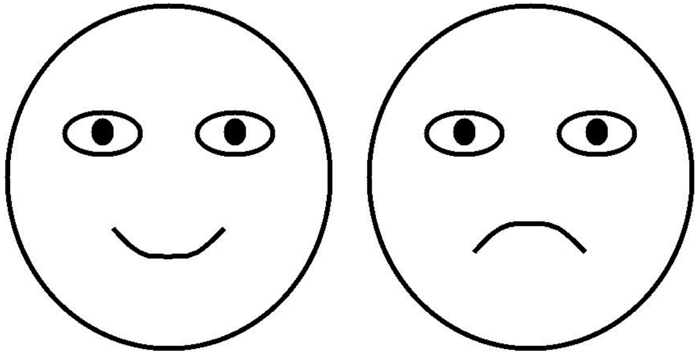




Some ontologies are 2-D only – e.g. line, junction, ...

Some involve 3-D structures and relations that can change while the contents of the optic array do not, e.g. relative distance, 3-D orientation of lines (sloping up and away *vs* down and away), etc.
What sort of ontology is needed to describe the flips in the duck-rabbit?

Some percepts use a meta-semantic or mentalist ontology: including 'happy', 'sad'.



Perception vs inference

In perception the appropriate ontology is deployed to construct an interpretation whose details are in partial registration with the sensory array or some systematic transform of it – and which can change globally as viewpoint and line of sight change.

(NB: Visual percepts are **not** in registration with retinal image or the map in V1, both of which change with every saccade — but with the optic array, as noted in Arnold Trehub *The Cognitive Brain*. `http://www.people.umass.edu/trehub/`)

There are specialised forms of representation for combining spatial topological, metrical and causal properties and relationships in both static structures and in multi-strand processes.

Perceptual processes use dedicated, specialised mechanisms operating on their particular sensory input and the interpretations, as opposed to general purpose inference mechanisms, e.g. using logic, or algebra.

However the perception/inference boundary is fuzzy.

We do not yet understand what these forms of representation are, how they are implemented in brains, what they can and cannot do, etc.

Their properties are very different from Fregean, logical representations.

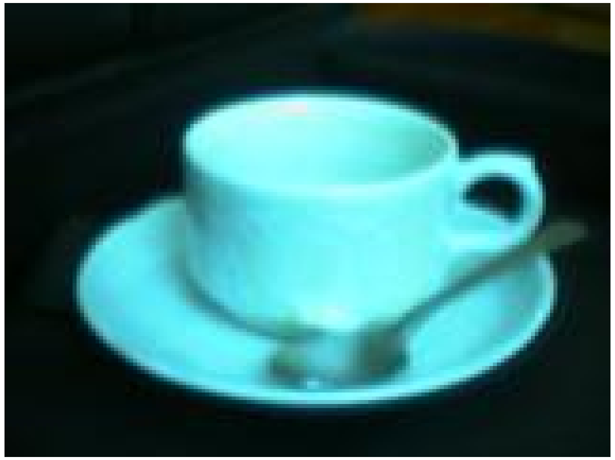
(In Sloman (1971) I called them ‘analogical’ representations’, and showed that analogical representations could be used to reason with.)

Some tasks for a crow-challenging robot?

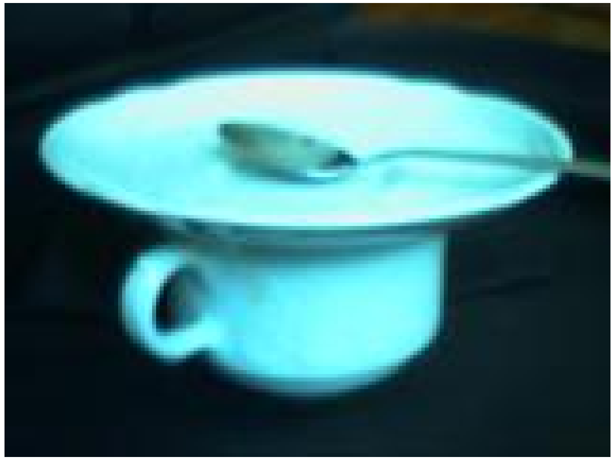
UPDATING THE BLOCKS WORLD

Using a two-finger gripper, what actions can get

from this:



to this:



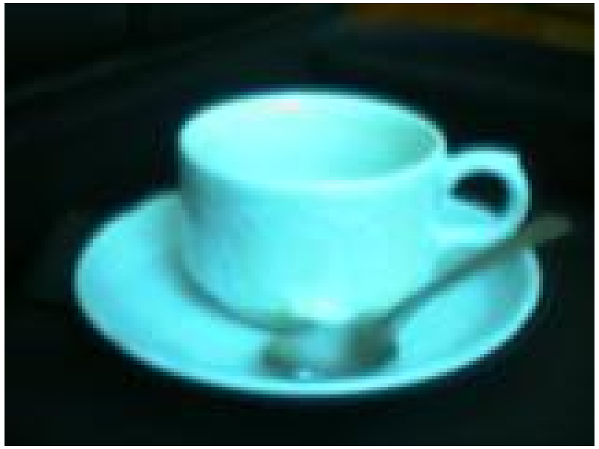
and back again?

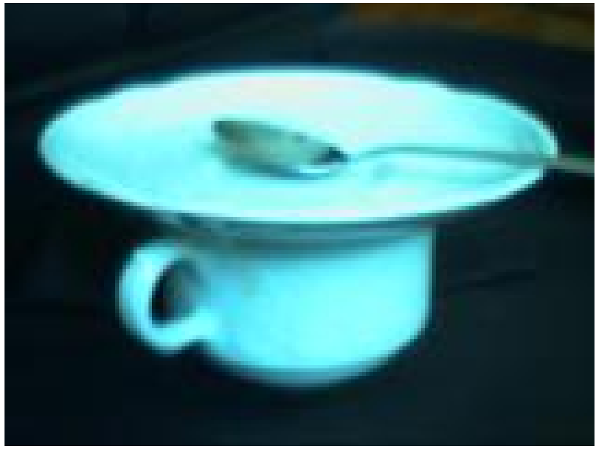
Or with saucer upside down?

Unfortunately even perceiving and representing the initial or final state (e.g. as a collection of surfaces with various possibilities for grasping from different directions) seems to be far beyond the capabilities of current AI vision systems, let alone thinking about possible actions to transform one to the other.

Current robots that can grasp things are very restricted in what they can grasp, and how much they understand what they are doing when they grasp.

Vision is much, much, more than recognition





What competences are required in a visual system to enable a child (or a robot) to get from the first configuration to the second?

- in many different ways,
- with different variations of the first configuration,
- with different variations of the second configuration,
- using the right hand,
- using the left hand,
- using both hands,
- using no hands, only mouth...?

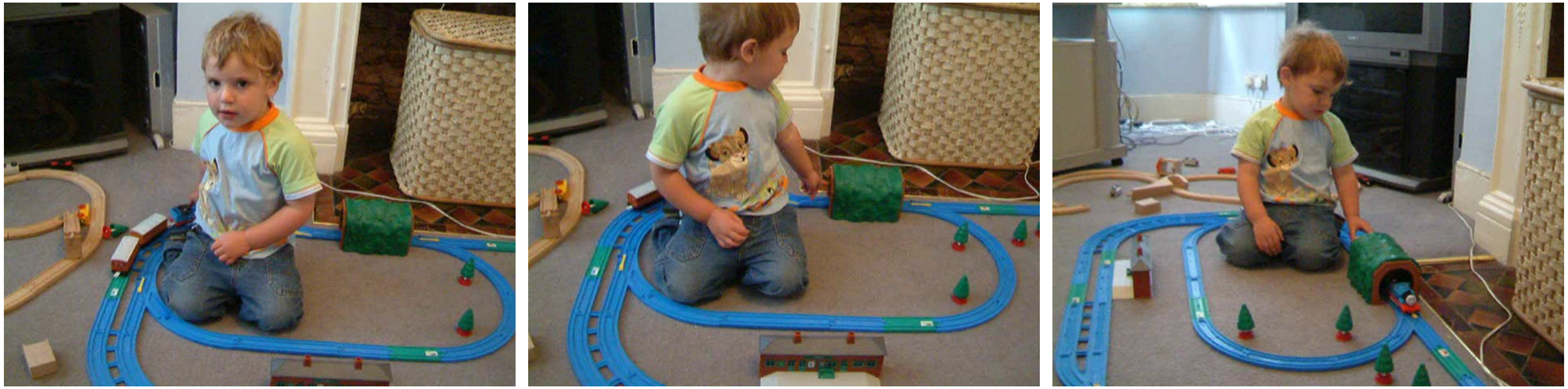
Can you visualise such processes – including interacting curved surfaces?

For more on this see

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

Snapshots from tunnel video

A child playing with his train illustrates many unobvious functions of vision.



- The child clearly knows what's going on in places he cannot see.
- He can point at and talk about something behind him that he cannot see.
- When he turns to continue playing with the train he knows which way to turn and roughly what to expect.
- When the train goes into the tunnel and part of it becomes invisible, he does not see the train as being truncated, and he expects the invisible bit to become visible as he goes on pushing.
- He sees the whole train as one thing while part of it is hidden in the tunnel.
- What is the role of vision in all of this? Frequently sampling the environment?

Not all of this competence is there from birth: at least some of it has to be learnt: what does that involve and what mechanisms make it happen?

The importance of concurrency

Besides emphasising the importance of processes as being the content of what is perceived (i.e. not just static structures), we are also emphasising the importance of concurrency, namely the perception as involving multiple perceived processes, some at the same level of abstraction, some at different levels of abstraction using different ontologies and linked to different parts of the central architecture.

- Perceived concurrency is involved in various human and animal activities involving two or more individuals engaged in fighting, dancing, mating, playing games, performing music, etc.
- Doing this well implies a need to be able (partly by running simulations?) to keep track of the actions of others at the same time as planning and performing one's own actions.
- Conjecture: our architecture evolved to support at least three sorts of concurrency:
  - Perceiving multiple concurrent external processes
  - Representing the same external process at different levels of abstraction
  - Different concurrent actions within the individual, such as

    walking (including posture control),
    working out where to walk,
    discussing philosophy or the view or .... with a companion,

    using different parts of the information-processing architecture simultaneously.

An example old idea that’s still relevant

Around 30 years ago I was working with David Owen and Geoffrey Hinton on a theory of vision that involved multi-level interpretation of static images, as on the next slide.

The theory explained how high level decisions could be reached relatively robustly and quickly, despite considerable complexity and noise at lower levels.

- If high level decisions were derived directly from low level image details the search space would be astronomical.
- By finding intermediate level recognisable structures and using their relationships to trigger high level hypotheses, while higher levels controlled ‘attention’ and some thresholds at lower levels, we allowed the sparsity of high level models to drive both speed and robustness. (U. Neisser called this ‘Analysis by synthesis’ about 40 years ago. Later it was called ‘hierarchical synthesis’. It has probably been reinvented many times.)
- The system degraded gracefully in both speed and accuracy as noise and clutter were added at the lowest level.
- A working implementation of that idea, called ‘POPEYE’ was described in chapter 9 of ‘*The Computer Revolution in Philosophy*’ (http://www.cs.bham.ac.uk/research/cogaff/crp/chap9.html)
- On that view, seeing involved creating multi-level structures concurrently.

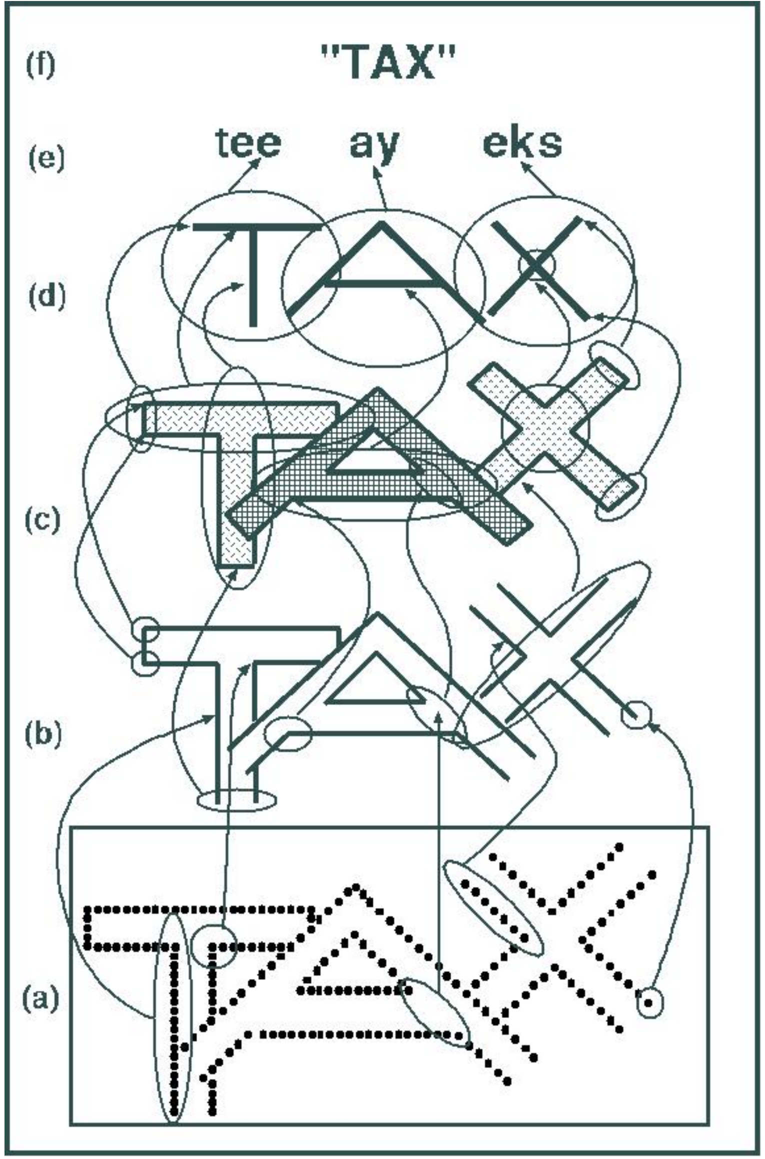
The next slide illustrates this old idea, showing how the Popeye program interpreted pictures made from dots by analysing the picture at different levels of abstraction in parallel, each level involving a different ontology from the others, using a mixture of bottom-up (data-driven) and top-down (model-driven, hypothesis-driven) interpretation, with rich structural relations between details at different levels.

Multiple levels of structure perceived in parallel

Old conjecture (e.g. Neisser 1960s): We process different layers of interpretation in parallel.
Obvious for language. What about vision?

Concurrently processing bottom-up and top- down helps constrain search. There are several ontologies involved, with different classes of structures, and mappings between them – so the different levels are in 'partial registration'.

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



Replace all that with concurrent multi-level processes – using different process-ontologies.

Picture from *The Computer Revolution in Philosophy* (1978 ch 9)
http://www.cs.bham.ac.uk/research/cogaff/crp/chap9.html

From Structures to Processes

We need to replace the idea that

1. seeing involves multi-level structures in partial registration with the optic array using different ontologies,

with the claim that

2. seeing involves multi-level process-simulations in partial registration using different ontologies, with rich (but changing) structural relations between levels.

- Shortly after the work on Popeye was done, David Hogg was a PhD student in the same department working on motion perception.

  D. Hogg. Model-based vision: A program to see a walking person. *Image and Vision Computing*, 1(1):5–20, 1983.

- His well known ‘walking man’ system was an early example of what I am now talking about: his model-based interpretation of a video of a walking man amounted to a simulation of a walker, partly controlled by the changing image data, and partly controlled by the dynamics of the model.
- Despite being his supervisor I did not appreciate the full significance of that work till now.

  I think he also did not see the full significance of what he had done: he described the system as showing how to use a model to interpret individual images, rather than claiming to show how to interpret a sequence of images as representing a process.

  Compare R.Grush in BBS 2004

What we did not do in the Popeye program

- We did not develop a program capable of representing the same multi-level structures, but with the objects in constant motion.
- An experiment to try one day would be producing movies derived from the 'dotty' word representing pictures. The conjecture is that people would not only see moving dots, but also moving lines, moving laminas, moving letters, ... though it is not clear how this would be objectively tested.
- I suspect we could cope with relative motions of parts of letters, e.g. so that the angles between parts of the letters change.

  Compare the work of Gunnar Johansson on movies made by attaching lights to joints on people, and filming them moving in the dark: when the lights start moving a 3-D process is perceived.
  Excellent demo: `http://www.bml.psy.ruhr-uni-bochum.de/Demos/BMLwalker.html`

Changes required for switching the Popeye architecture to a moving scene

- It would be silly to keep all the low level detail indefinitely as new details would be coming in all the time – but perhaps some low level histories are needed for some tasks.
- Different times of preservation would be relevant to different things at different levels in the ontology, e.g. depending on whether they are large or small, static or moving, or of interest relative to some goal.
- It might be useful to add low level motion maps, or even to replace the static low level maps completely. (Compare A.Trehub: *The Cognitive Brain*)

Ontology available to a visual system can vary

We need to find out more about the ontologies used in vision – and other forms of perception, also planning, reasoning, acting, desiring, disliking, etc.

including both somatic (e.g. sensorimotor – modal and amodal) ontologies
and exo-somatic ontologies.

Ontologies used can vary

- between species
- between individuals in a species

  e.g. being able to read different languages, formalisms, music
- between stages in development of an individual

  learning to see new sorts of things
- between different parts of the same individual

NOTE:

An extension to the ontology need not be definable in terms of what was there before.

New born babies are not born with concepts in terms of which all concepts of modern science can be define.

So we need to explain substantive, not just definitional, ontology-extension.

an old problem in philosophy of science.

(Contrast Fodor: *The Language of Thought*)

Example: Perceiving causation

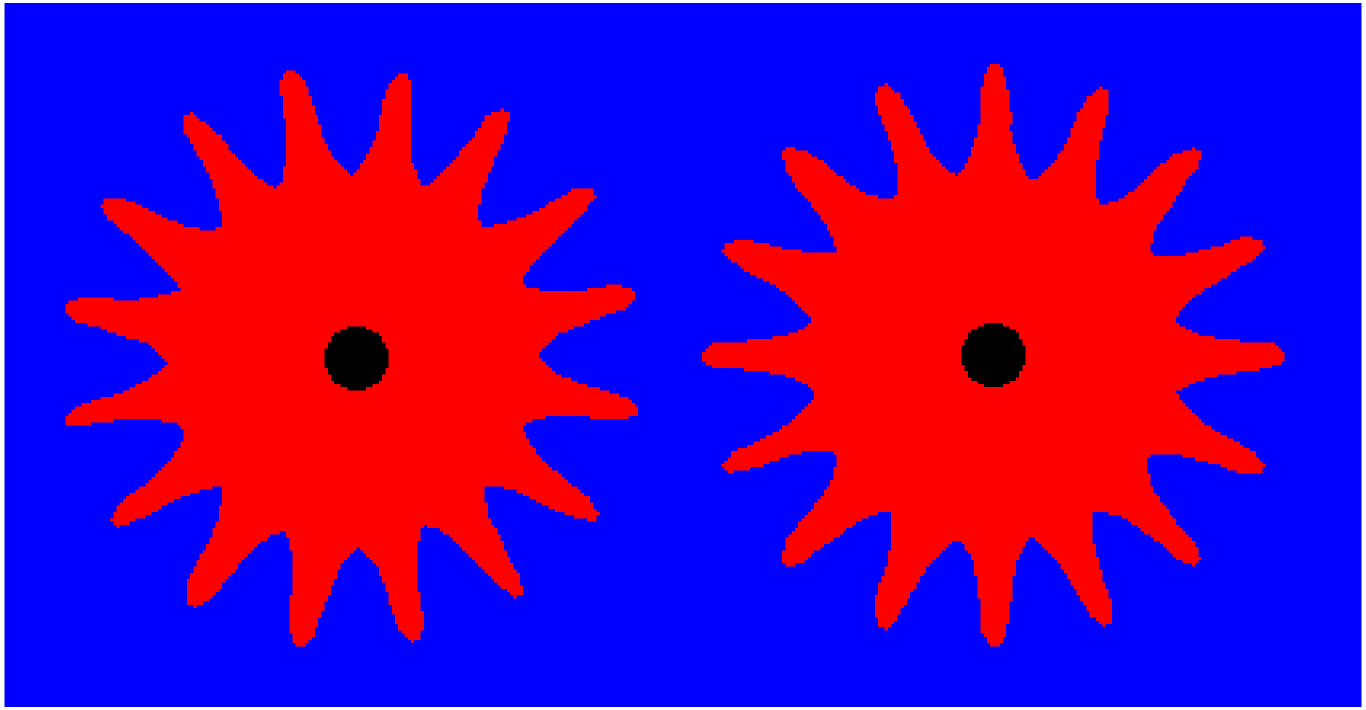
Our ability to perceive moving structures, and our meta-level ability to think about what we perceive, is intimately bound up with perception of causation and affordances.

In some cases the causal relations are inherent in what is seen, whereas in others they involve invisible structures and processes: but the same key idea is used in both cases.

Illustrations follow.

Invisible, Humean, causation – mere correlation

Two gear wheels attached to a box with hidden contents.
Here we do not perceive causation.



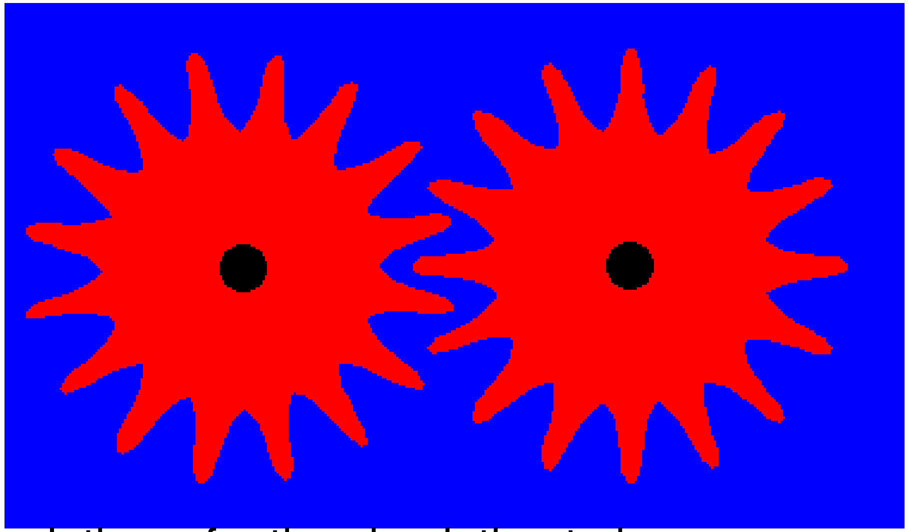
Can you tell by looking what will happen to one wheel if you rotate the other about its central axis?

You can tell by experimenting: you may or may not discover a correlation.

Compare experiments reported by Alison Gopnik in her invited talk at IJCAI'05, Edinburgh July 2005

Visible, intelligible, Kantian, causation

Two more gear wheels:



Here you (and some children) can tell 'by looking' how rotation of one wheel will affect the other.

NB The simulation that you do makes use of not just perceived shape, but also unperceived constraints: rigidity and impenetrability. These constraints need to be part of the perceiver's ontology and integrated into the simulations, for the simulation to be deterministic.

Visible structure does not determine all the constraints: we also have to learn about the nature of materials, to see what is happening, and understand causation.

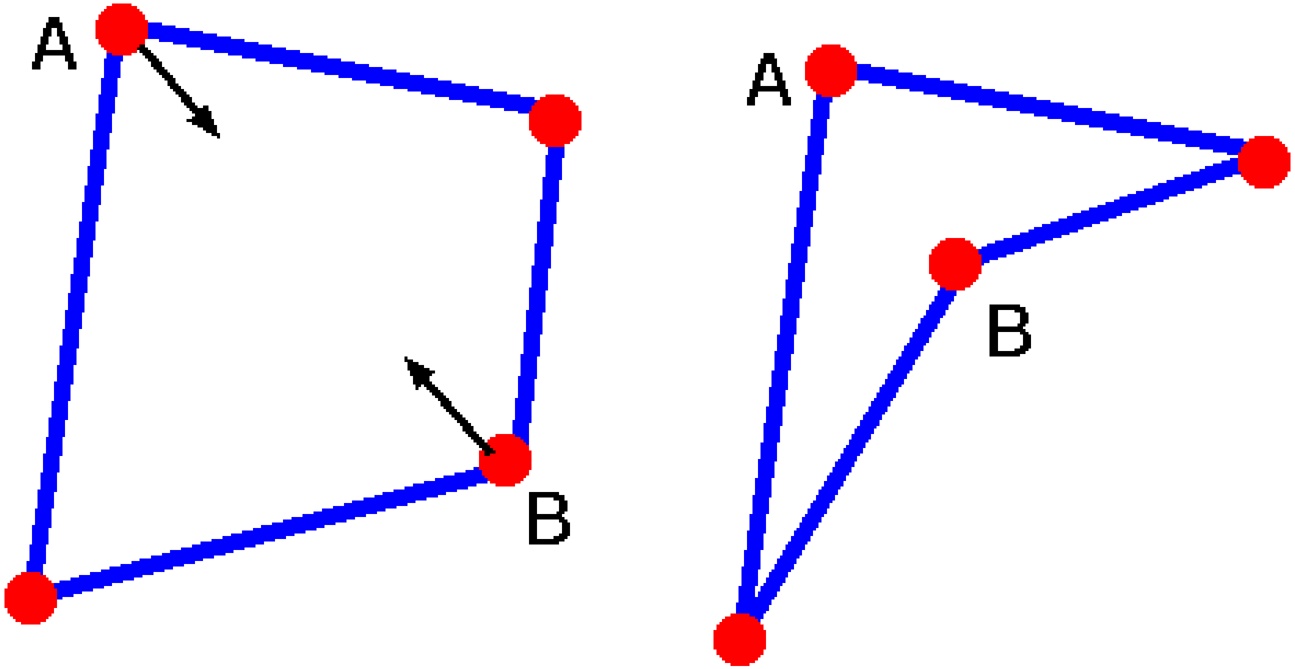
> We need to explain how brains and computers can set up and run simulations involving multiple concurrent changes of relationships, subject to varying constraints determined by context.

These ideas are developed in two online documents

> http://www.cs.bham.ac.uk/research/projects/cosy/papers/#pr0506
> COSY-PR-0506: Two views of child as scientist: Humean and Kantian
>
> http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0601
> COSY-DP-0601 Orthogonal Competences Acquired by Altricial Species (Blanket, string and plywood).

Simulating motion of rigid, flexibly jointed, rods

On the left: what happens if joints A and B move together as indicated by the arrows, while everything moves in the same plane? Will the other two joints move together, move apart, stay where they are. ???



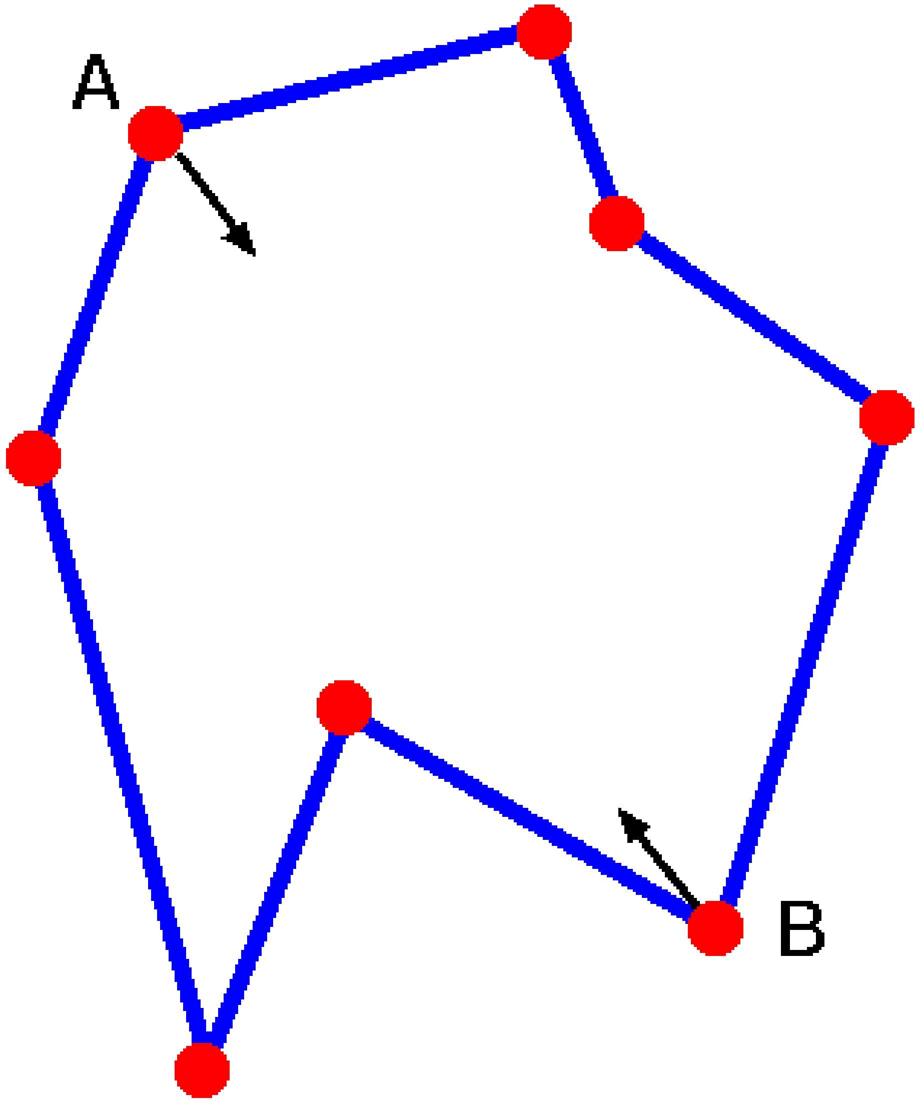
- What happens if one of the moved joints crosses the line joining the other two joints?
- We can change the constraints in our simulations: what can happen if the joints and rods are not constrained to remain in the original plane?

Multiple links: how we break down

Can you tell how the other rods will move, if A and B are moved together and all the rods are rigid, but flexibly jointed?

There are not enough constraints. In this case our causal reasoning merely allows us to think about a range of options, though it is not easy. Unlike simpler linkages, most people will not be able to see whether the continuum of possible processes divides into clearly distinct subsets except (perhaps) by spending a lot of time exploring.

As situations get more complex, human abilities to simulate degrade rapidly: our understanding of Kantian causation tends to be limited to relatively simple, deterministic cases, though we can learn to grasp more complex structures and processes – up to a point.

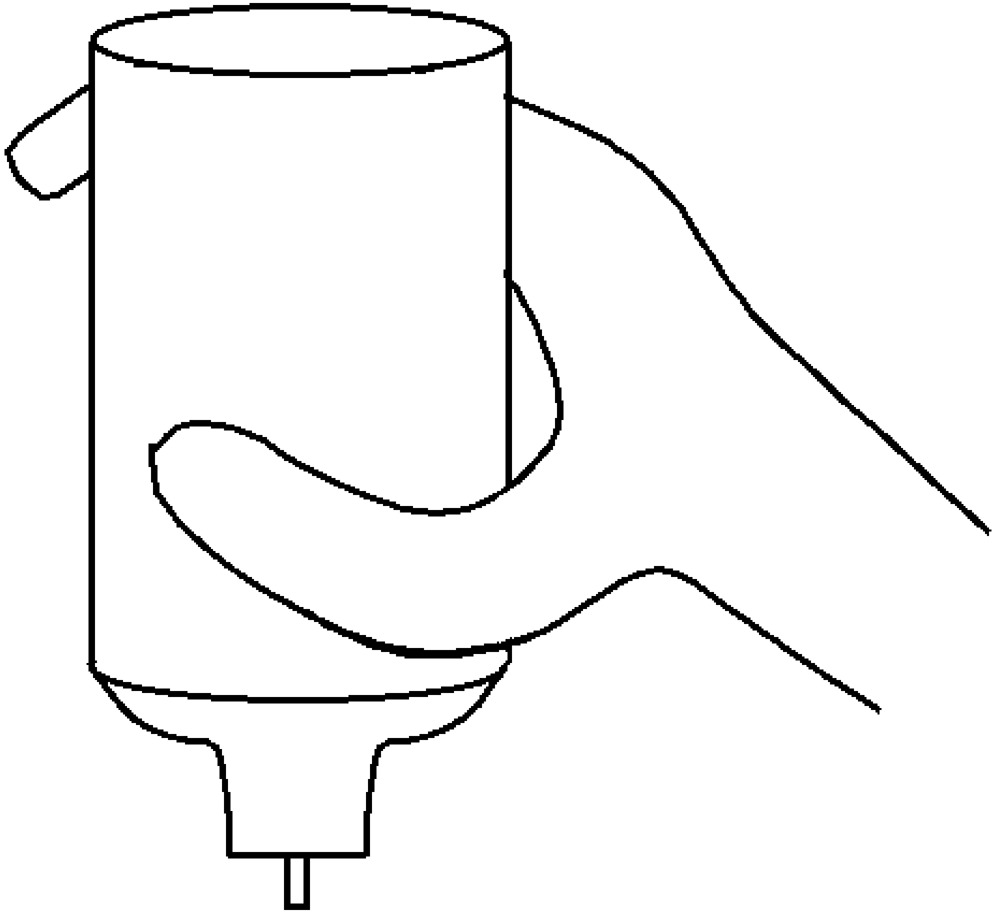


Perhaps intelligent artificial systems will have similar limitations.

Visual reasoning about something unseen

Visual reasoning can be ‘disconnected’ from sensory data.

If you turn the plastic shampoo container upside down to get shampoo out, why is it often better to wait before you squeeze?



In causal reasoning we often use runnable models that go beyond sensory information: part of what is simulated cannot be seen – a Kantian causal learner will constantly seek such models, as opposed to Humean (statistical) causal learners, who merely seek correlations.

Note that the model used here assumes uncompressibility rather than rigidity.

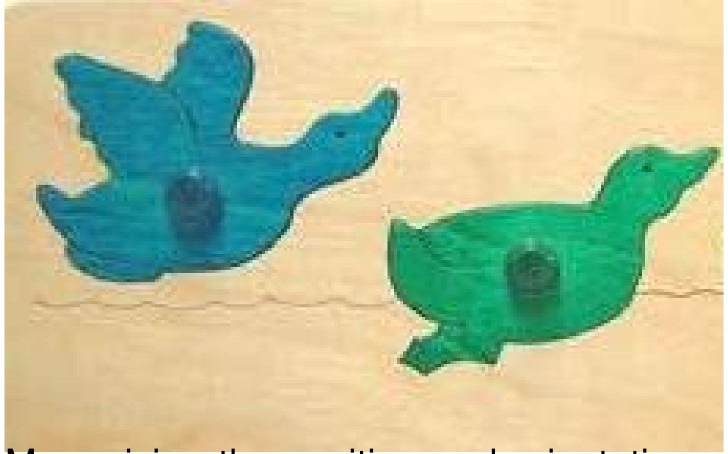
Also, our ability to simulate what is going on explains why as more of the shampoo is used up you have to wait longer before squeezing.

Sometimes we run the wrong simulation if we don’t understand what is going on.

Like the person who suggested that you have to wait for the water from the shower to warm the air in the container.

We cannot do it all from birth

The causal reasoning we find so easy is difficult for infants.



A child learns that it can lift a piece out of its recess, and generates a goal to put it back, either because it sees the task being done by others or because of an implicit assumption of reversibility. At first, even when the child has learnt which piece belongs in which recess there is no understanding of the need to line up the boundaries, so there is futile pressing.

Later the child may succeed by chance, using nearly random movements, but the probability of success with random movements is very low. (Why?)

Memorising the position and orientation with great accuracy will allow toddlers to succeed: but there is no evidence that they have sufficiently precise memories or motor control. Eventually a child understands that unless the boundaries are lined up the puzzle piece cannot be inserted. Likewise she learns how to place shaped cups so that one goes inside another or one stacks rigidly on another.

These changes require the child to build a richer ontology for representing objects, states and processes in the environment, and that ontology is used in a mental simulation capability. HOW?

Stacking cups are easier partly because of symmetry, partly because of sloping sides: both reduce the uniqueness of required actions, so the cups need less precision and are easier to manage.

Learning ontologies is a discontinuous process

- The process of extending competence is not continuous (like growing taller or stronger).
- The child has to learn about new kinds of
  - objects,
  - properties,
  - relations,
  - process structures,
  - constraints,...
- and these are different for
  - rigid objects,
  - flexible objects,
  - stretchable objects,
  - liquids,
  - sand,
  - mud,
  - treacle,
  - plasticine,
  - pieces of string,
  - sheets of paper,
  - construction kit components in Lego, Meccano, Tinkertoy, electronic kits...

I don’t know how many different things of this sort have to be learnt, but it is easy to come up with many significantly different examples.

Much of what is learnt is about kinds of stuff

Human children (and presumably also chimpanzees, nest building-birds and members of other altricial species) learn many things about the environment by playful exploration, using a collection of special-purpose mechanisms developed by evolution for the task.

Part of what they learn concerns the behaviour of various kinds of physical stuff in the environment, including

- kinds of material like:
  - sand, water, mud, straw, leaves, wood, rock,
  - and in our culture also: things like paper, cloth, cotton-wool, plastic, aluminium foil, butter, treacle, velcro, meal, concrete, glue, mortar,
  - various kinds of food (meat, fish, vegetable matter, peanut-butter, etc.)
- kinds of components that can be combined to form larger objects including:
  lego, meccano, tinker-toy, Fischer-technik, and many more,

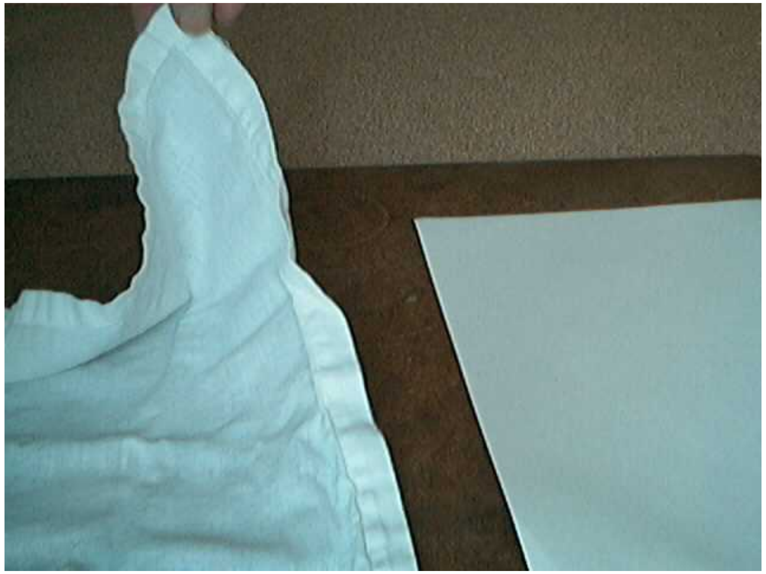
  including, for nest-building birds, twigs, leaves, etc.

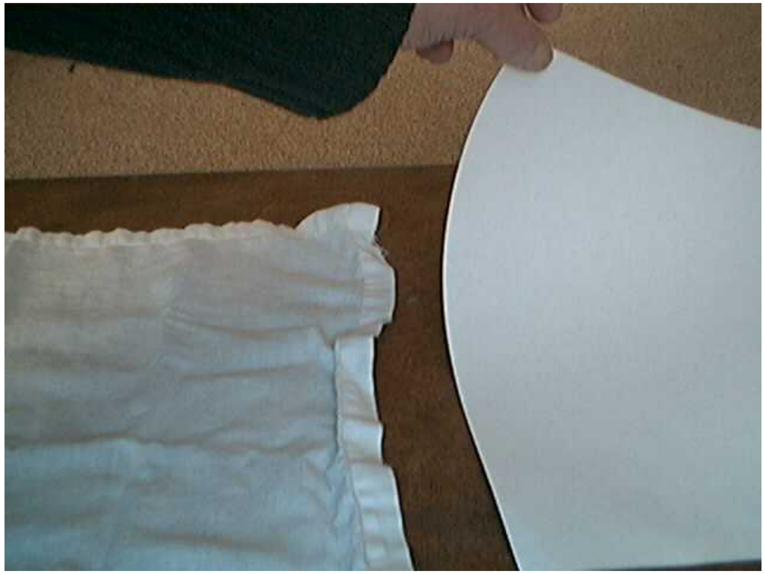
'Behaviour' of such things includes their responses to being folded, crushed, picked up, thrown, twisted, chewed, sucked, pressed together, compressed, stretched, dropped, and also the properties of larger wholes containing them.

The variety of kinds of stuff and kinds of behaviour should not be thought of as a continuum, e.g. something that might be form a vector space parametrised by a collection of real-valued parameters. Rather there are qualitative and structural differences important in many sub-ontologies that have to be learnt separately (even if some precocial species have precompiled subsets).

A few examples follow: you can probably think of many more.

Cloth and Paper





You have probably learnt many subtle things unconsciously about the different sorts of materials you interact with (e.g. sheets of cloth, paper, cardboard, clingfilm, rubber, plywood).

That includes learning ways in which you can and cannot distort their shape.

Lifting a handkerchief by its corner produces very different results from lifting a sheet of printer paper by its corner – and even if I had ironed the handkerchief first (what a waste of time) it would not have behaved like paper.

Most people cannot simulate the precise behaviours of such materials but we can impose constraints on our simulations that enable us to deduce consequences.

In some cases the differences between paper and cloth will not affect the answer to a question, e.g. the example on the slide about folding a sheet of paper, below.

What do you know about cloth and paper?

There are probably many things you know about cloth and (printer) paper that you have never thought about, but implicitly assume in your reasoning about them, including imagining consequences of various sorts of actions.

Common features

- Both have two 2-D surfaces, one on each side.
- Both have bounding edges.
- Both can be made to lie (approximately) flat on a flat surface.
- Both can be smoothly pressed against a cylindrical or conical surface, but not a spherical (concave or convex surface)
- To a first approximation neither is stretchable, in the sense that between any points P1 and P2 there is a maximum distance that can be produced between P1 and P2, if there is no cutting or tearing.
- Both can be cut, torn, folded, crumpled into a ball....

Differences

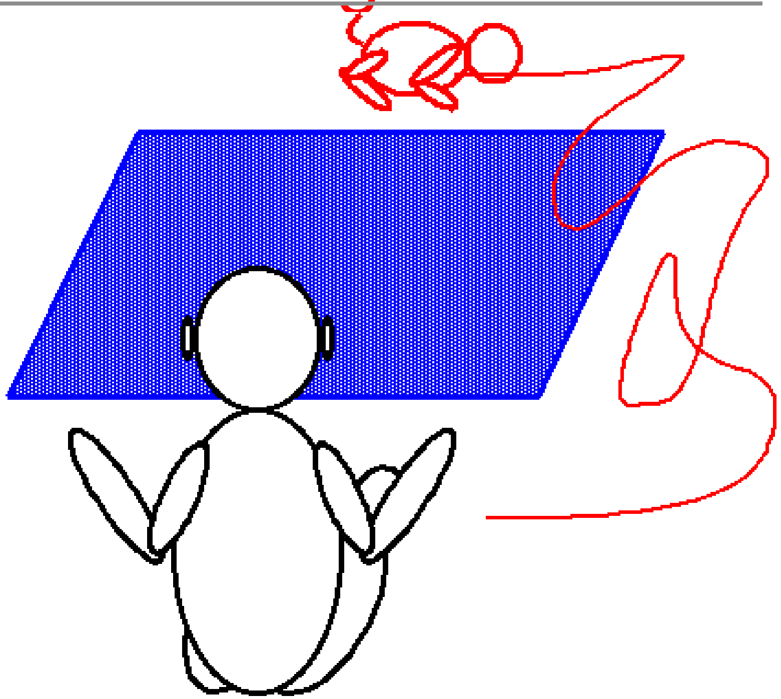
- most cloth can be slightly stretched (though some is very stretchy)
- Paper folded and creased tends to retain its fold, cloth often doesn't (there are exceptions, especially if heat is applied).
- Paper folded and not creased tends to return to its flatter state. It is more elastic.
- Paper folded once can stand upright resting on either a V-shaped edge or a pair of parallel edges.
- Paper is rigid within its plane (three collinear points remain collinear while the paper lies flat).

NOTE: tissue paper is somewhere in between.

Example: Blanket and String

If a toy is beyond a blanket, but a string attached to the toy is close at hand, a very young child whose understanding of causation involving blanket-pulling is still Humean, may try pulling the blanket to get the toy.

At a later stage the child may either have extended the ontology used in its conditional probabilities, or learnt to simulate the process of moving X when X supports Y, and as a result does not try pulling the blanket to get the toy lying just beyond it, but uses the string.



However the ontology of strings is a bag of worms, even before knots turn up.

Pulling the end of a string connected to the toy towards you will not move the toy if the string is too long: it will merely straighten part of the string.

The child needs to learn the requirement to produce a straight portion of string between the toy and the place where the string is grasped, so that the fact that string is inextensible can be used to move its far end by moving its near end (by pulling, though not by pushing).

Try analysing the different strategies that the child may learn in order to cope with a long string, and the perceptual, ontological and representational requirements for learning them.

Ontologies for getting at something

Understanding varieties of causation involved in learning how to get hold of a toy that is out of reach, resting on a blanket, or beyond it.

Some things to learn through play and exploration

Toy on short blanket Grab edge and pull

Toy on long blanket Repeatedly scrunch and pull

Toy on towel Like blanket

Toy on sheet of plywood
Pull if short(!!), otherwise crawl over or round it

Toy on sheet of paper Roll up?
(But not thin tissue paper!)

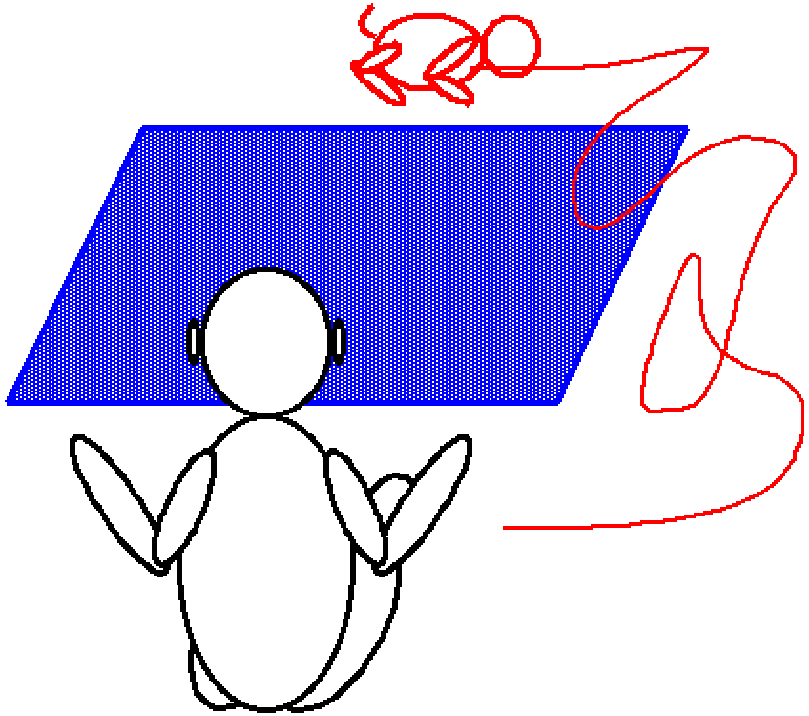
Toy on slab of concrete Crawl over or round

Toy at end of taut string Pull

Toy at end of string with slack Pull repeatedly

String round chair-leg Depends

Elastic string
?????



See this discussion of learning orthogonal recombinable competences
http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0601

It takes a lot of learning to develop all the visual and reasoning competences required for seeing and understanding these affordances – including visualising what would have happened if you had done something different, or if someone else were to move something.
Our spatial and visual competence goes far beyond actually doing.

Creativity in a physical environment

The different kinds of knowledge mentioned above can be combined in many different ways, including novel ways, in understanding what is perceived in the environment and what actions are and are not possible in different circumstances, and what the consequences of those actions will be.

We need to understand architectures and mechanisms for combining such knowledge and competences where appropriate.

Chapter 6 of *The Computer Revolution in Philosophy* attempted to analyse some of the processes about 30 years ago, but only at a high level of abstraction. http://www.cs.bham.ac.uk/research/cogaff/crp/chap6.html

- Sometimes competences are combined in physical action, using new combinations of material, tool, arrangement of parts or actions, to solve a problem; but in some cases it is done in thought (i.e. using deliberative mechanisms), as pointed out by Craik, Popper and many others.
- Precocial species, e.g. spiders, may have very specific 'hard wired' combinations of competence regarding specific kinds of stuff, specific spatial structures and processes; whereas humans some other altricial species are able both to extend knowledge within each of the categories, and to forge new combinations in perceiving novel scenes and performing novel actions — a meta-competence that underlies engineering, science and art.
- Such competence in pre-linguistic children and non-linguistic animals cannot depend on external language, though it may be part of the basis for language, which, with other forms of cultural information-transmission (e.g. toys) enormously enhances and accelerates development.
- In a young child and in many animals the creative recombination of competence is applied in perceiving and using affordances for oneself, whereas humans later learn to see 'vicarious affordances', as discussed previously – essential in parents and carers watching children who may be about to hurt themselves, or may need help, or in seeing opportunities for predators who may attack one's young.

As if this were not complex enough

Humans, though not infants, can combine all that creativity about the physical with creatively deployed knowledge about the mental. How?

- All of the ‘orthogonal’ competences listed above involve semantic competence: the ability to acquire, store, manipulate and use information.
- In humans at least there is also second-order and higher-order semantic competence (meta-semantic and meta-meta-semantic competence), namely the ability to use information about information and information users, e.g. thinking about what another individual, or oneself, can see, knows, wants, intends to do, about their reasoning, learning, planning or decision-making processes, including thinking about what A knows or fears about what B thinks about C.
- Such higher-order semantic competence is crucial to teaching.
- In teaching, learning from teachers, interpreting actions of others, negotiating, cooperating, planning revenge, and other social actions the opportunities for creative combination of previously listed competences with meta-semantic and social competences are enormous.
- E.g. understanding perceived unfamiliar behaviour in another may require combining knowledge (or hypotheses) about what the individual can and cannot see, what he can and cannot do, what he may wish to do, what affordances various physical materials shaped in a specific way can provide, and so on.

  (‘Perhaps he dropped something through that grating and is trying to retrieve it using chewing gum on the end of a twig.’)

How much of this applies to other animals?

- Not all animals can learn these things, even if they share a lot of physical structure with humans.
- So it is likely that there are very specific, very powerful brain mechanisms involved, possibly several different mechanisms that evolved in different combinations — we are not discussing all-or-nothing capabilities.
- Even among humans there may be different combinations, e.g. Archimedes, Shakespeare, Newton, Kant, Mozart, Darwin, Turing. Picasso, Menuhin – in which case there is no such thing as human psychology.
- If the hundreds, or thousands, of different kinds of knowledge acquired in the first few years are stored in different parts of the brain, using different mechanisms, then different sorts of brain damage or deficiency could interfere with different sub-competences. Has anyone looked? (E.g. Williams' Syndrome?)
- Since most of the creative brain mechanisms evolved before human language capabilities and appear in pre-linguistic children, despite involving rich forms of semantic and syntactic competence (using internal representations), it could be that the generative (combinatorial) and extendable aspects of those pre-linguistic competences provided a foundation for the later evolution of linguistic competence.

  G-languages with rich structural variability and compositional semantics needed for internal use in perception, reasoning, planning, goal formation, etc., evolved before external languages for communication evolved. (See the 'primacy' paper.)

Constraints on mechanisms
The problem of speed

Some pictures follow.
View them at 1 or 2 second intervals.

What do you see
and how fast do you see it?

Pictures taken by Jonathan Sloman

The problem of speed

The problem of speed

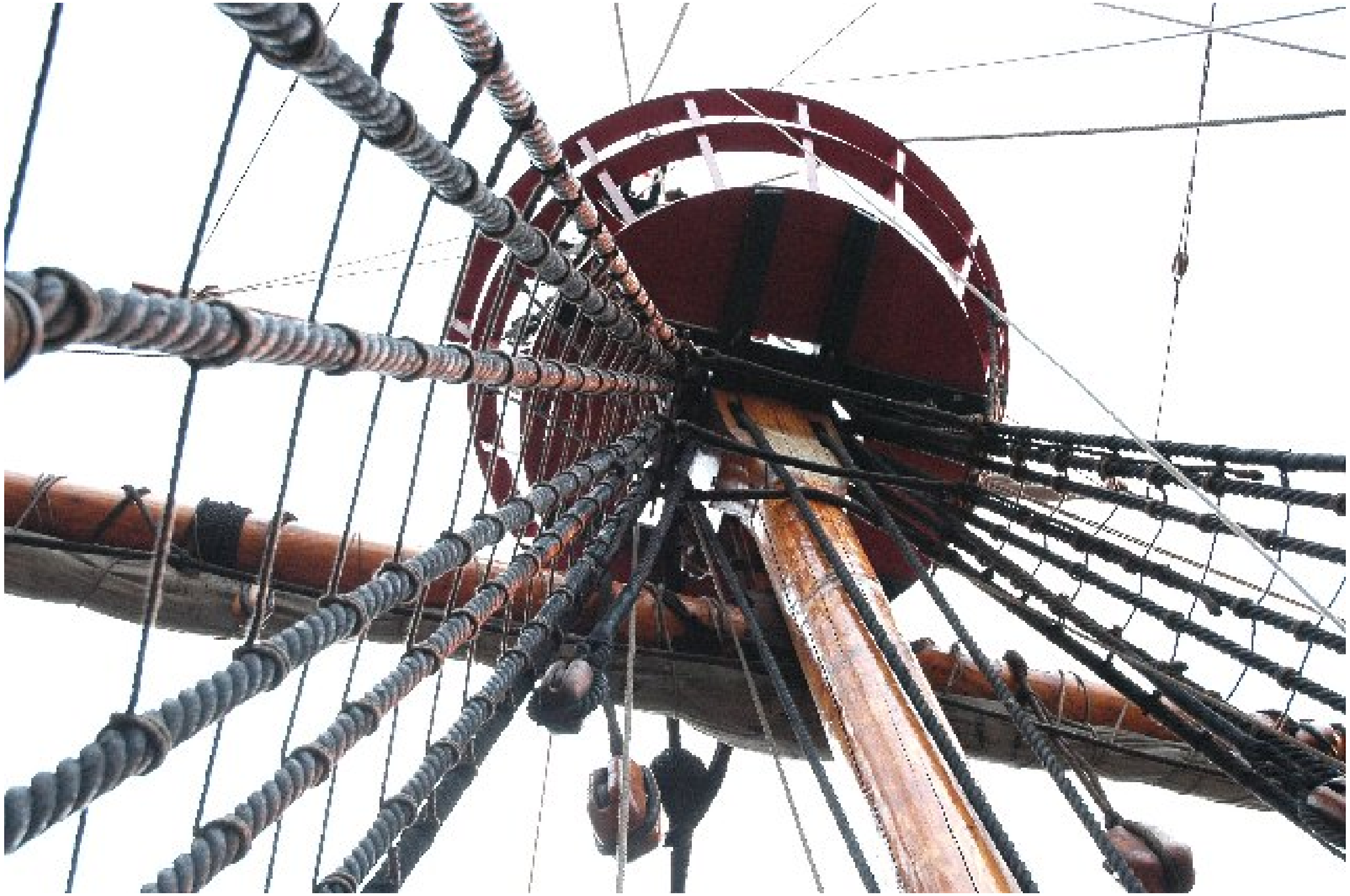
The problem of speed

The problem of speed

The problem of speed

The problem of speed

The problem of speed



In which direction are you looking?

The problem of speed

The problem of speed

The problem of speed

What needs to be explained

The speed with which people can see at least roughly what sort of scene is depicted by each image and what sorts of things are in it implies that our visual mechanisms are capable of finding low level features, using them to cue in features of the image and the scene at various levels of size and abstraction, arriving at percepts involving known types of objects within one or two seconds. Some high level decisions can be made in less than half a second.

This is related to what happens if you go round a corner or come out of an underground station in an unfamiliar town.

I am not claiming that you see everything depicted in all these images, merely that something must be going on to arrive at a 3-D interpretation, including in some cases perceived processes involving pedestrians, bicycles, cars and other vehicles, the state of the weather.

The speed at which this happens, and the variety of types of context in which it can happen, along with the variety of types of information that we can obtain all exceed anything currently possible in computer vision systems.

The inherent ambiguity of all low level image features and the speed with which high level interpretations are formed implies that there is some form of computation going on that AI researchers and vision and neuroscience researchers have not yet identified.

A new kind of dynamical system

Perhaps we need a kind of dynamical system

- composed of multiple smaller multi-stable dynamical systems
- that can be turned on and off as needed,
- some with only discrete attractors, others capable of changing continuously,
- many of them inert or disabled most of the time, but capable of being turned on or off (sometimes very quickly)
- each capable of being influenced by other sub-systems or sensory input or current goals, i.e. turned on, then kicked into new states bottom up or top down,
- constrained in parallel by many other multi-stable sub-systems
- with mechanisms for interpreting configurations of subsystem-states as representing scene structures and affordances, and changing configurations as representing processes
- where the whole system is capable of growing new sub-systems, permanent or temporary, and short-term (for the current environment) or long term when learning to perceive new things.

This contrasts with

- Dynamical systems with a fixed number of variables that change continuously
- Dynamical systems with one global state (atomic state dynamical systems)
- Dynamical systems that can only be in one attractor at a time
- Dynamical systems with a fixed structure (e.g. a fixed size vector or tree).

Sensory modality and mode of representation

- Sensory modality driving a simulation need not determine the nature of the percept: you can see or feel the same shape.
- A unitary percept of a process can be driven by input from diverse sensory modalities – e.g. seeing, hearing, feeling the same things changing.
- The content, i.e. what is represented, does not determine the nature of the medium used to implement the percept, as long as it has a rich enough structure and appropriate mechanisms to create, modify, access and use the contents.
- Perceiving a process has much in common with running a multi-level simulation of that process. Examples of what the simulation might be include:
  - a set of variables with changing values driven by sensory data
  - a database of logical assertions along with insertions and deletions driven by sensory data
  - a hybrid mechanism – logical assertions with equations linking changing variables, as can happen in some spreadsheets,
  - a spatially structured changing model,
  - a stored 'script' for the process with a pointer moving through the script at a rate determined by sensory input,
  - it may use a powerful form of representation that we have not yet thought of though evolution discovered it long ago.
- Whatever form of representation is used, currently known brain mechanisms do not seem to provide the required functionality.

Learning ontologies is a discontinuous process

- The process of extending competence is not continuous (like growing taller or stronger).
- The child has to learn about new kinds of
  - objects,
  - properties,
  - relations,
  - process structures,
  - constraints,...
- and these are different for
  - rigid objects,
  - flexible objects,
  - stretchable objects,
  - liquids,
  - sand,
  - mud,
  - treacle,
  - plasticine,
  - pieces of string,
  - sheets of paper,
  - construction kit components in Lego, Meccano, Tinkertoy, electronic kits...

Contrast definitional ontology-extensions with substantive ontology extensions (as often occur in science).

CONJECTURE

In the first five years

- a child learns to run at least hundreds,
- possibly thousands, of different sorts of simulations,
- using different ontologies

  with different materials, objects, properties, relationships, constraints, causal interactions.
- and throughout this learning, perceptual capabilities are extended by adding new sub-systems to the visual architecture, including new simulation capabilities

Much of this learning is about different kinds of stuff of which things can be made and different kinds of properties distinguishing kinds of stuff, where neither kinds of stuff nor their properties can necessarily be sensed by the animal.

Some more examples are available in

`http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0601`
COSY-DP-0601 Orthogonal Competences Acquired by Altricial Species (Blanket, string and plywood).

Arguments against symbol-grounding and sensori-motor analyses of knowledge can be found in

`http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#grounding`

`http://www.cs.bham.ac.uk/research/projects/cogaff/misc/nature-nurture-cube.html`

Sensory-motor vs action-consequence contingencies

Two evolutionary 'gestalt switches'?

The preceding discussion implies that during biological evolution there was a switch (perhaps more than once) from

> insect-like understanding of the environment in terms of sensory-motor contingencies linking internal motor signals and internal sensor states (subject to prior conditions),

to

> a more 'objective' understanding of the environment in terms of action-consequence contingencies linking changes in the environment to consequences in the environment,

followed by

> a further development that allowed a generative representation of the principles underlying those contingencies, so that novel examples could be predicted and understood, instead of everything having to be based on statistical extrapolation.

To be more precise, it was an addition of a new competence rather than a switch

One of the major drivers for this development could be evolution of body parts other than the mouth that could manipulate objects and be seen to do so.

> However the cognitive developments were not inevitable consequences: e.g. crabs that use their claws to put food in their mouth do not necessarily use the more abstract representation.

Perception of shape is not shape-reconstruction

What sort of 3-D interpretation is required depends on what it is to be used for.

Shape perception in computers is often demonstrated by giving the machine one or more images, from which it constructs a point-by point 3-D model of the visible surfaces of objects in the scene.

This achievement is then demonstrated by projecting images of the scene from new viewpoints.

But there is no evidence that any animal can do that and very few humans (e.g. some artists) can produce accurate pictures of viewed objects using a new viewpoint, whereas many graphics engines do it.

Human/animal understanding of shape, including having information relevant to action and prediction, is very different from having a point by point 3-D model

- The point of perception is not making images: many results must be useful for action – e.g. building nests from twigs, peeling and dismembering food in order to get at edible parts, escaping from a predator, making a tool, using a tool.
- A 'percept' constructed by the perceiver needs to include information about what is happening, what could happen and what obstructions there are to various kinds of happening (positive and negative affordances).
- These happenings are of many different kinds, so different kinds of information must be synthesised from sensory information (influenced by prior knowledge, prior ontologies, prior goals).

How is it possible?

- It has long been known that the problem is too unconstrained to be solvable – every 2-D image is inherently capable of being generated by infinitely many 3-D scenes.
- It has long been conjectured that the environment is constrained in ways that make the problem contingently solvable — where some constraints may be learnt by the individual perceiver and others are derived from the genetically determined structure, functions, and processing mechanisms of perceptual systems: The 'cognitively friendly environment' hypothesis. (Sloman 1978)
- Examples include use of binocular vision (which helps only a little, and only at short distances), motion perception (which can be far more important, whether the motion is in the perceiver or in the perceived objects), and assumptions about the nature of various materials, e.g. how rigid they are, what their surface texture is, the kinds of lighting found in various situations, knowledge of the effects of occluding opaque objects, intervening shrubbery, distortions caused by heat-haze, etc.
- Of course, we and other animals are not perfect perceivers and banking on these constraints can sometimes lead us into error (e.g. the Ames room and other illusions, including some used by animals, such as camouflage) though usually the implicit assumption of cognitive friendliness works well.

How can brains do all this?

What good are examples without any theory of how brains do it?

- Beware: if we theorise on the basis of too few kinds of examples we may come up with inadequate theories: a common problem in AI, philosophy, psychology and neuroscience.
- If all the above is correct, human brains need to be able to run very many different kinds of simulations:

  including processes involving stones, blocks, string, paper, sand, cloth, mud, plasticene, rigid materials, flexible materials, materials that are rigid in two dimensions and flexible in one (e.g. paper), water, sand, mud, cotton wool, plasticine, wire, fibrous materials, viscous liquids, various kinds of meat, various kinds of vegetable matter, brittle materials, stretchable materials, thin films, solid lumps of matter, and many more.
- We are not restricted to simulating what has occurred in our evolutionary history: children can learn to play with and think about toys and devices none of their ancestors ever encountered – e.g. skipping ropes, slinky springs, zip fasteners, velcro, scotch tape, computer games and future inventions too.

If we start building explanatory models based on too few explananda we may fool ourselves into accepting inadequate theories.

So we should seek a ‘generative’ explanation. That’s an old idea, but if the generative explanation is too simple (like current popular theories of learning) it may work on toy examples but fail hopelessly in the tasks summarised here.

How many non-human species?

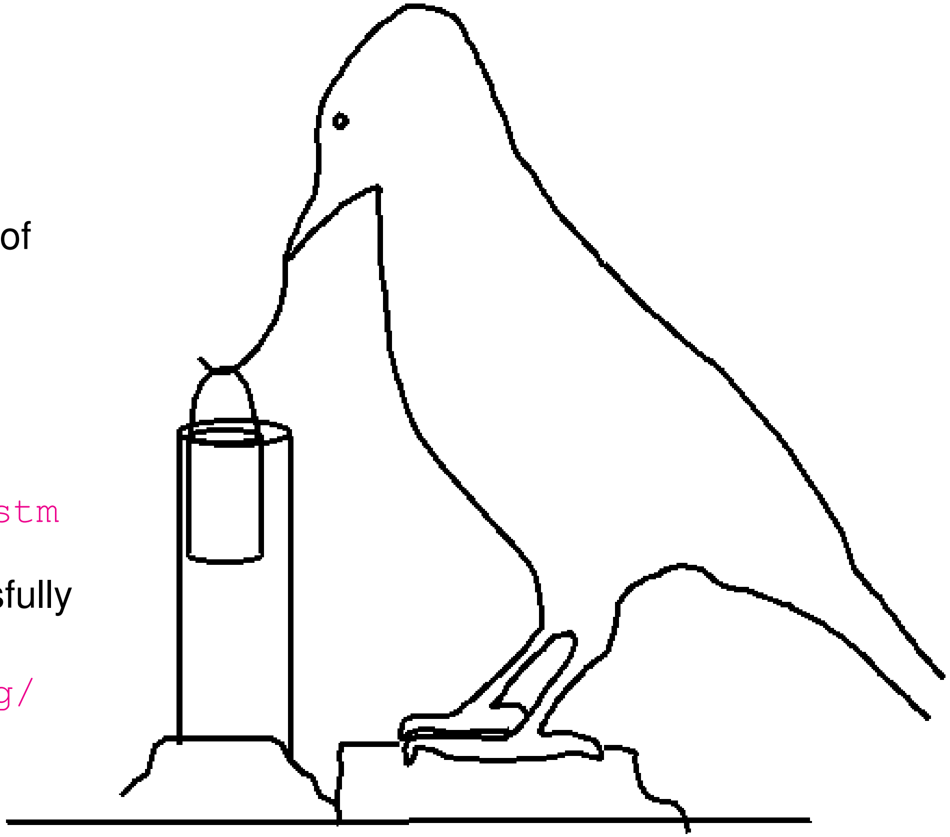
Betty the hook-making New Caledonian crow.

Give to google: betty crow hook:
You'll find a link to the oxford zoology lab, with videos of Betty making hooks in different ways.

She appears to be a Kantian causal reasoner.

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

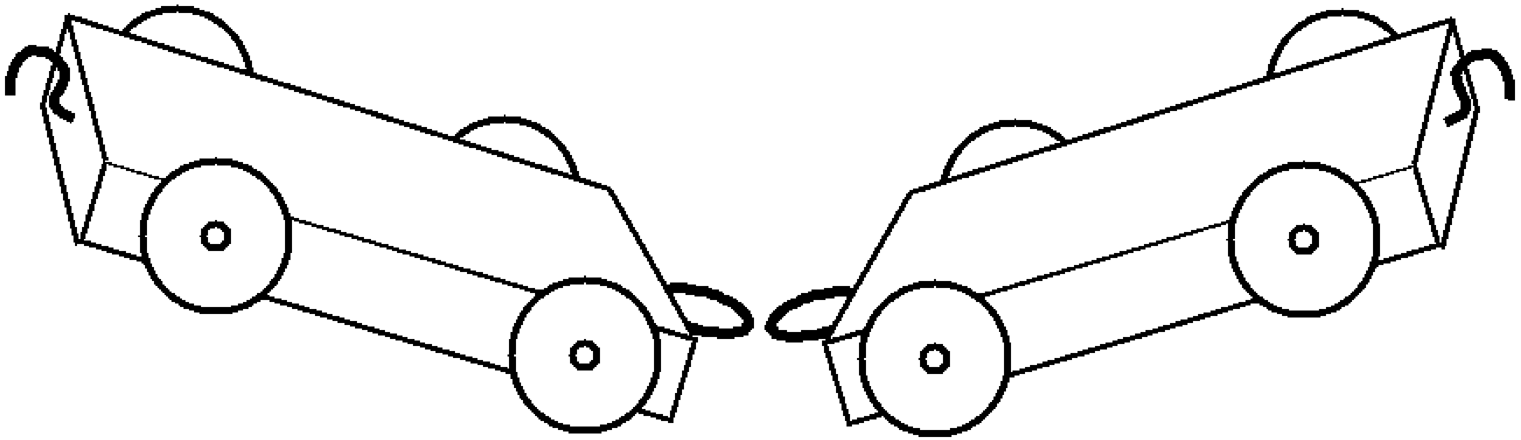
Contrast the 18 month old child attempting unsuccessfully to join two parts of a toy train by bringing two rings together (http://www.cs.bham.ac.uk/~axs/fig/josh34_0096.mpg)



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

She seems to. How?

Hooks defeating a 19 Month old child



See the movie of an 19-month old child failing to work out how to join up the toy train – despite a lot of visual and manipulative competence also shown in the movie.

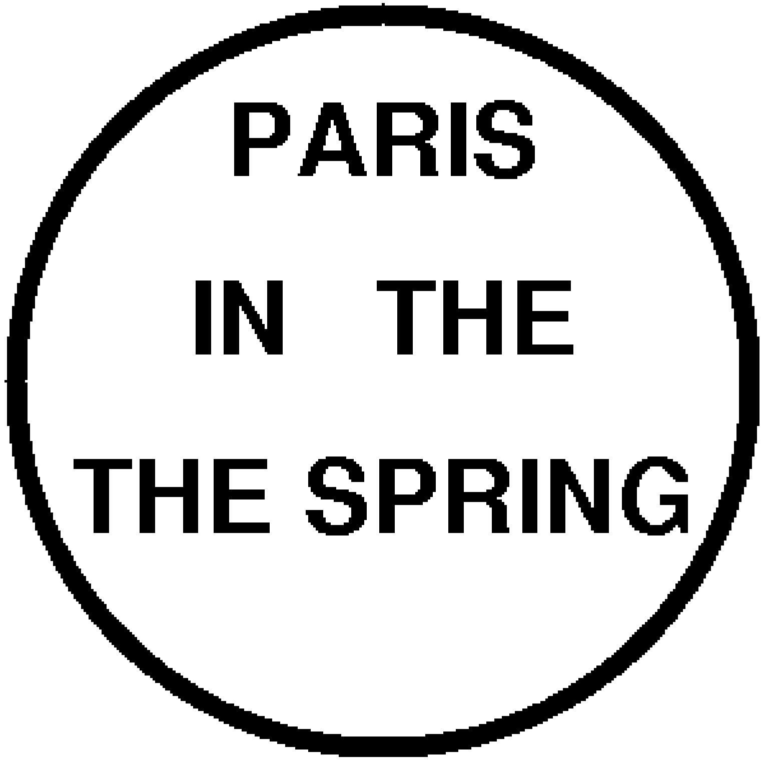
- http://www.jonathans.me.uk/josh/movies/josh34_0096.mpg
  4.2Mbytes
- http://www.jonathans.me.uk/josh/movies/josh34_0096_big.mpg
  11 Mbytes

A few weeks later he had no problem joining up the train.

Was he a Humean causal learner or a Kantian causal learner?

I suspect the latter, but specifying the simulation model developed by a learner who understands hooks and rings will not be easy.

A well known example of controlled hallucination



In this case some people only see an abstraction – a familiar phrase, rather than what is actually visible in the circle.

Similarly when we run simulations we may sometimes hallucinate what we expect to be in the environment rather than what is actually there.

Do you see only a familiar phrase? If so, read on.

A part of you may see what ‘you’ do not see!

Often people who have been shown the example on the previous slide and are convinced, even after insistent questioning, that what they see is just a familiar phrase, can be made to realise their mistake, even with their eyes shut.

- Ask subjects who claim to have seen only ‘PARIS IN THE SPRING’ to shut their eyes.
- Then ask one of these two questions
  - How many words were in the circle?
  - Where was the ‘THE’?
- Some of them realise, even with their eyes shut, that what they were certain they had seen was not what they had actually seen.

This seems to show that, at least for such a person, it is wrong to ask ‘What did he/she see?’, for the answer will be different for different parts of the person.

A part of you may record the layout of the words in the circle even though another part (central to social interactions) decides that it is a familiar phrase on the basis of evidence that is often perfectly adequate, and it does not check for consistency with the low level detail.

In a cognitively ‘friendly environment’ where decisions sometimes have to be taken quickly, this could be a good design, even if it occasionally causes errors.

Learning when to be more thorough can be useful in some environments!

This idea may explain phenomena revealed in experiments on ‘change blindness’ – where experimenters wrongly assume that we know what we see, whereas much perception is subconscious.

Seeing non-existent motion

There are many optical illusions in which things appear to be moving when they are not, including motion after-effects, and patterns used in so-called ‘op-art’.

See `http://www.michaelbach.de/ot/index.html`

Nothing I have said explains any particular phenomenon of illusory motion, but the existence of such things is perhaps less surprising if we think of all visual perception as involving the running of process simulations controlled in part by sensory data, and subject to presumed constraints that may sometimes be inferred wrongly.

If all that powerful apparatus exists ready to be used at very short notice, it may easily be triggered into action by a variety of partial cues: some erroneous interpretations are very likely in that case — but in a ‘cognitively friendly’ environment the result will be fast decisions that are mostly correct.

Seeing intentional actions

Seeing a person or animal or machine doing something may involve a richer ontology than is required for seeing physical things moving under the control of purposeless physical forces.

- If you see a marble rolling down a slope occasionally changing direction or bouncing into the air as a result of surface irregularities or stones in its path, your simulation may include changes of position, speed and direction of motion, all consistent with what you know about physical objects.
- If you see a person walking down a slope occasionally moving to one side and picking things off bushes, you will see not only physical motion, but the execution of an intention, possibly several intentions, e.g. getting to something at the bottom of the slope, collecting biological specimens, and eating berries.
- One of the things a child has to learn to do is interpret perceived motion in terms of inferred goals, plans and processes of plan execution. Thus the simulations run when intentional actions are perceived may include a level of abstraction involving plan execution.

  For a recent discussion see Sharon Wood, 'Representation and purposeful autonomous agents' *Robotics and Autonomous Systems* 51 (2005) 217-228
  `http://www.cogs.susx.ac.uk/users/sharonw/papers/RAS04.pdf`
- When several individuals are involved, there may be several concurrent, interacting, processes with different intentions and plans to simulate. Learning to understand stories beyond the simplest sequential narratives requires learning to do this. (Contrast coping with 'flashbacks'.)

Conjecture

A great deal of our understanding of causality is intimately bound up with our ability to create constrained, deterministic simulations, and to learn about their properties by 'playing' with them.

We are not born with all the specific simulation capabilities we have,
but we, and possibly several other altricial species, are born with mechanisms for developing such simulations — depending on
what is encountered in the environment.

We are born equipped to become Kantian causal reasoners
about more and more aspects of the environment,
though there are always residual unexplained but useful correlations.

Similar remarks can be made about the history of science and technology.

See `http://www.cs.bham.ac.uk/research/projects/cosy/papers/#pr0506`

Disclaimers: No claim is made:

- That the simulations at any level are complete
- That they are accurate (errors, imprecision and fuzziness abound)
- That we are aware of all the simulations we are running
- That only humans can do this
- That all humans can run the same kinds of simulations

  Different kinds of education, different kinds of training, e.g. artistic, athletic, mathematical training, playing with different kinds of toys, etc. can all produce different ontologies, representations and simulation capabilities. Even children with similar competences may get there via different routes along a partially ordered network of trajectories. There are genetic differences too – e.g. 'Williams syndrome' children don't develop normal spatial competences.

- That it is obvious how to implement these ideas in artificial visual systems
- That the theory is compatible with any current theory of learning
- That the theory is compatible with known brain mechanisms

  We may have to search for previously undiscovered mechanisms (including previously unknown types of virtual machines implemented in brains)
  See Trehub's book (*The Cognitive Brain, 1991*) for some relevant ideas.
  There are probably lots of things I should have read but have not.

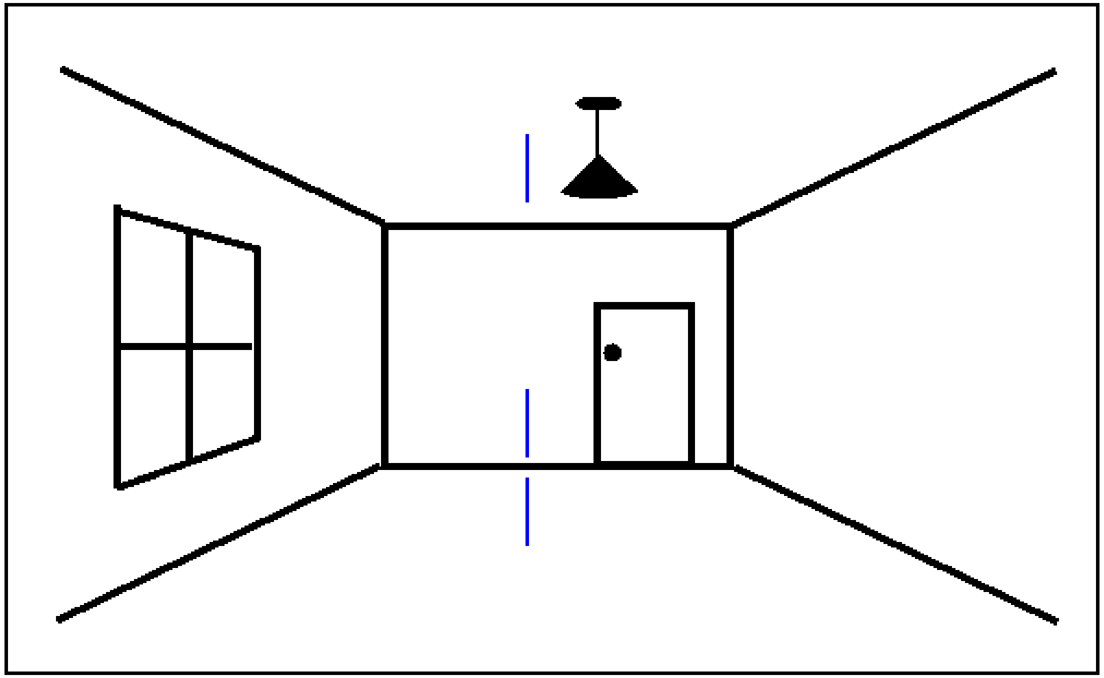
  There is considerable overlap with the BBS paper by R.Grush (2004): The Emulation Theory of Representation.

Isomorphism is not needed

Here's a modified version of a picture from chapter 7 of *The Computer Revolution in Philosophy*, also in the 1971 IJCAI paper.

Objects and relations within a picture need not correspond 1 to 1 with objects and relations within the scene, as is obvious from 2-D pictures of 3-D scenes.

For example: pairs of points in the image that are the same distance apart in the image can represent pairs of points that are different distances apart in 3-D space – e.g. vertically separated points on the walls, and horizontally separated points on the floor and ceiling. (And *vice versa*.)



Some pairs of parallel edges in the scene are represented by parallel picture lines, others by converging picture lines.

The small blue lines can be interpreted in different ways, with different spatial locations, orientations and relationships. On each interpretation the structure of the image remains unchanged, but the structure of the 3-D scene changes.

Inadequate alternative theories

Among the precursors to the theory are several that in different ways are inadequate, despite providing useful steps in the right direction.

- One general kind of inadequate theory assumes that what is perceived can be expressed as a collection of measures, sometimes called ‘state variables’, (e.g. coordinates, orientations, and velocities of objects in the scene) and that what is simulated can be expressed as continuous or discrete changes in a (possibly) large vector of state variables.
- This kind of numerical representation is inadequate because it fails to capture the structure of the environment, e.g. the decomposition into objects with parts, and with different sorts of relationships between objects, between parts within an object, between parts of different objects, etc.

  People who are familiar with a particular collection of mathematical techniques keep trying to apply them everywhere instead of analysing the problems to find out what forms of representation are really required for the tasks in hand.
- Many theories do not do justice to the diversity of functions of vision. E.g. some people seem to think the sole or main function of vision is recognition of instances of object types.
- Most theories of vision do not allow that we see not only what exists but what can and cannot happen in a given situation – affordances.
- Dynamical systems theorists have some of the right ideas but restrict ontologies and forms of representation to mathematics of vector spaces.

Some References

Some of my previous work on this topic.

(1971) Interactions between philosophy and AI: The role of intuition and non-logical reasoning in intelligence, in *Proc 2nd IJCAI* Reprinted twice. Online at http://www.cs.bham.ac.uk/research/cogaff/04.html#200407

(1978) Chapters 7 and 9 of *The Computer Revolution in Philosophy*, online at http://www.cs.bham.ac.uk/research/projects/cogaff/crp/

(1982) Image interpretation: The way ahead?, in *Physical and Biological Processing of Images* Eds. O.J. Braddick and A.C. Sleigh. http://www.cs.bham.ac.uk/research/projects/cogaff/06.html#0604

(1989), On designing a visual system (Towards a Gibsonian computational model of vision), *Journal of Experimental and Theoretical AI* 1, 4, http://www.cs.bham.ac.uk/research/projects/cogaff/81-95.html#7

(2001), Evolvable biologically plausible visual architectures, *Proceedings of British Machine Vision Conference*, Ed. T. Cootes and C. Taylor, BMVA, http://www.cs.bham.ac.uk/research/projects/cogaff/00-02.html#76

(2005a) A (possibly) new theory of vision. (PDF presentation) http://www.cs.bham.ac.uk/research/projects/cosy/papers/#pr0505

(2005b) Two views of child as scientist: Humean and Kantian (PDF presentation) http://www.cs.bham.ac.uk/research/projects/cosy/papers/#pr0506

(2005c) DR.2.1 Requirements study for representations (Interim report from CoSy Robotic project) http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0507

(2006) Orthogonal Recombinable Competences Acquired by Altricial Species (Blankets, string, and plywood), Discussion paper, CoSy robotic project. http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0601

Papers and presentations with Jackie Chappell (biologist).

(2005) The Altricial-Precocial Spectrum for Robots (in IJCAI'2005) http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0502

(2007a) Natural and artificial meta-configured altricial information-processing systems (To appear in IJUC 2007?) http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0609

(2007b) Computational Cognitive Epigenetics (Commentary to appear in BBS 2007?) http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0703

(2007c) Causation and the Altricial/Precocial Distinction (Oxford Workshop June 2007) http://www.cs.bham.ac.uk/research/projects/cogaff/talks/wonac/