Towards A Theory of Vision:

Requirements for a robot with human child-like or crow-like visual and learning capabilities.

Aaron Sloman

http://www.cs.bham.ac.uk/~axs
School of Computer Science, The University of Birmingham

With help from colleagues on the CoSy project

http://www.cs.bham.ac.uk/research/projects/cosy/
(Maria Staudte at DFKI kindly commented on an early draft)

And others, including Jackie Chappell, Biosciences, Birmingham.

These slides will be accessible from from symposium web site. See also

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

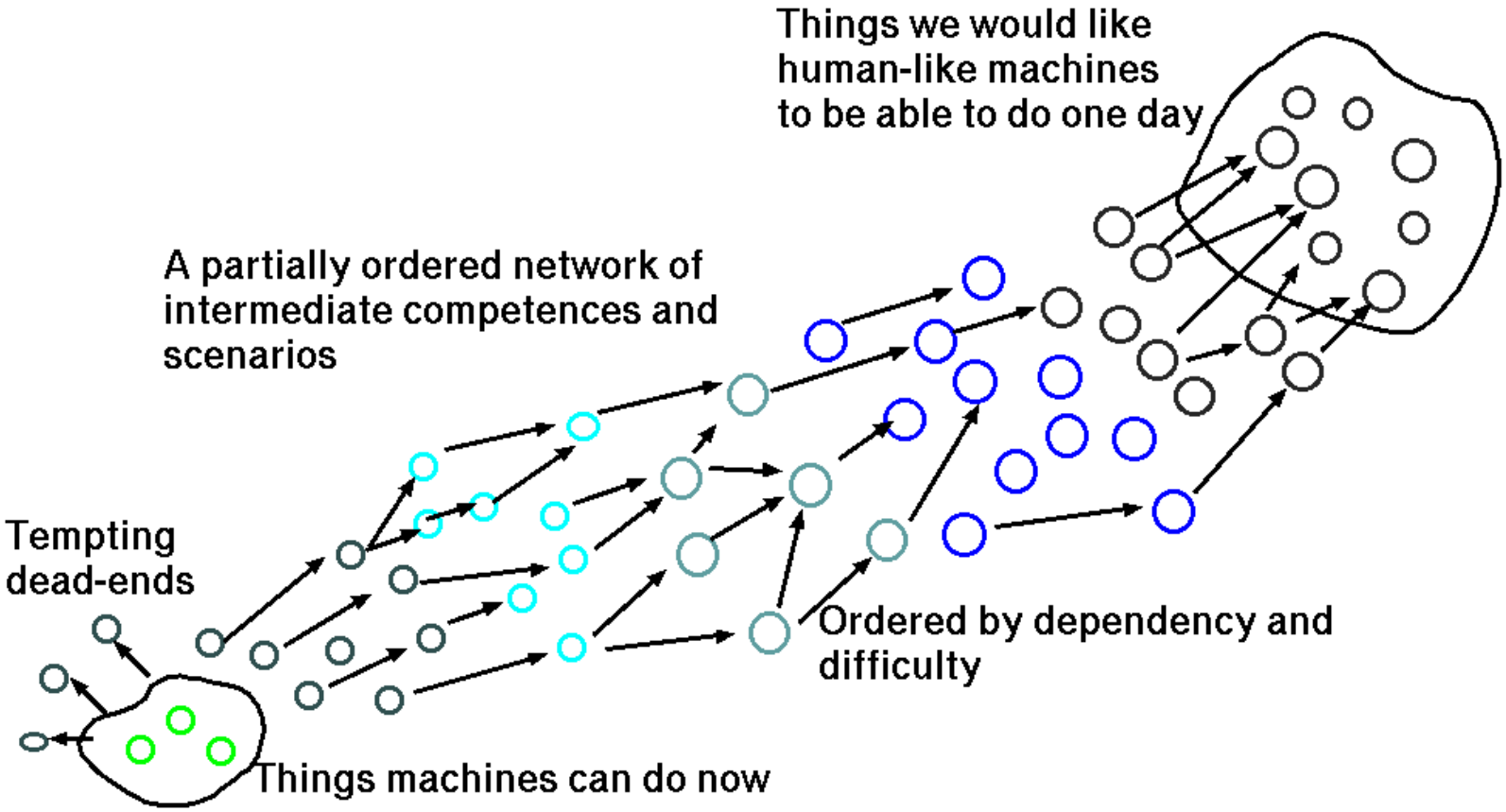
1 Some background

We start with some background to put the research in vision into the larger context of an attempt to specify requirements for future robots with human-like capabilities.

This is of interest both because it may help with the practical goal of designing more intelligent and more useful, flexible, companionable robots and also because it may help us understand human beings better by understanding the requirements for designs that explain what we can do.

The latter is my personal aim: I am more a biologist (and philosopher) than an engineer, but the engineering methodology is required for doing biology, psychology and philosophy well.

Steps towards a research roadmap



Forward chaining research asks: how can we improve what we have already done?
Backward chaining research asks: what is needed to achieve our long term goals?

See the introduction to GC5 in the booklet and on the web: researchers don't put nearly enough effort into analysing requirements.

Many of the hardest tasks are concerned with seeing 3D motion and affordances

Doing science

We need to move beyond ‘Here’s my architecture’.

For real scientific knowledge we need to have a theory about the space of possible designs and how design options relate to task requirements.

This leads to the idea of studying relations between

– design space (space of possible designs), and

– niche space (space of possible sets of requirements).

REQUIREMENTS FOR ANIMAL & ROBOT VISION

Vision is a process involving multiple concurrent simulations at different levels of abstraction in (partial) registration with one another and sometimes (when appropriate) in registration with visual sensory data and/or motor signals.

Max Clowes: Vision is controlled hallucination.
We add: multi-level controlled hallucination.

The theory has different facets, which link up with many different phenomena of everyday life as well as experimental data, and with a host of problems in philosophy, psychology (including developmental and clinical psychology), neuroscience, biology and AI (including robotics).

It raises new questions for AI, psychology, neuroscience and others.

Example: watch this video of child playing with a toy train set.

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

Perceiving structures vs perceiving affordances

Structures

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

Affordances (positive and negative)

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

modal, as opposed to categorical, types of perception.

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

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

Beware the tabula rasa trap: millions of years of evolution were not wasted!

We should not consider only human competence

Humans are a result of billions of years of evolution producing many different solutions to the problems of coping with a complex environment.

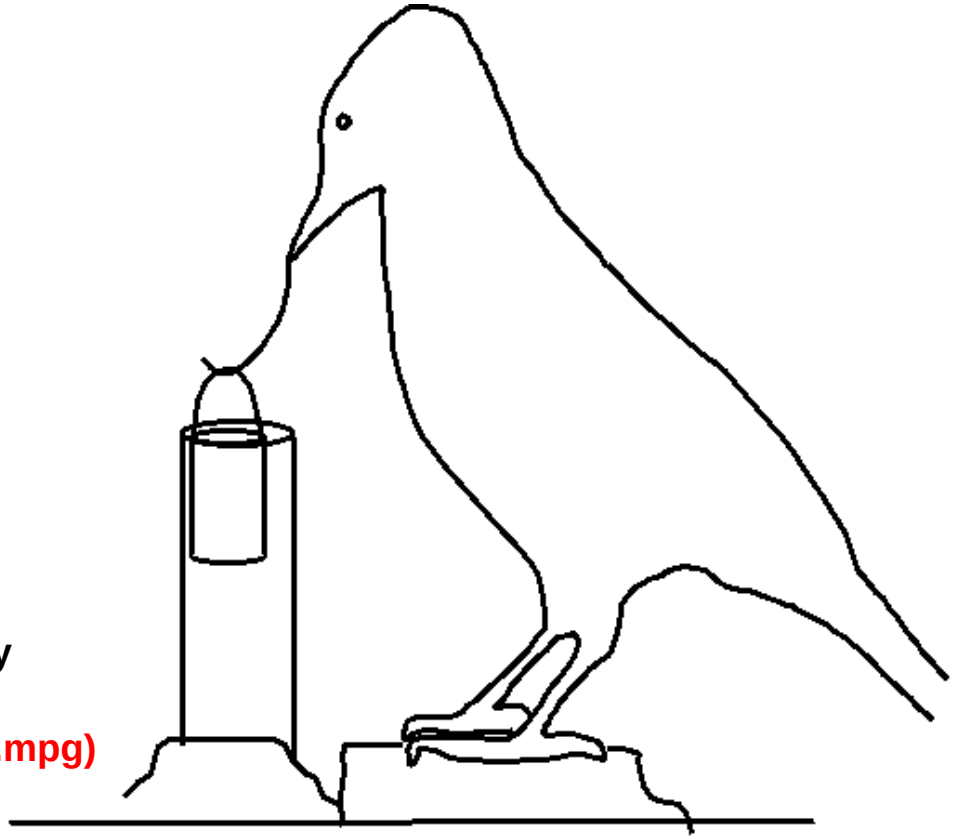
Betty the hook-making New Caledonian crow.

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 have a deep understanding of structure, process and causation.

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?

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 something to copy) seems to be far beyond the capabilities of current AI vision systems, let alone thinking about possible actions to transform one to the other – e.g. the angle of approach required to grip cup or saucer or spoon in a particular location, e.g. the left-most point of the saucer's rim, or the tip of the teaspoon's bowl.

Some tasks for a crow-challenging robot? (2)

Consider how, prior to the action, the agent has to

- **identify parts of objects, or parts of parts, e.g. the edge of the handle, or the far edge of the handle or a certain portion of the edge of the saucer**
- **see and understand their shapes and relationships**
- **identify possible actions: grasping this thing here from this direction**

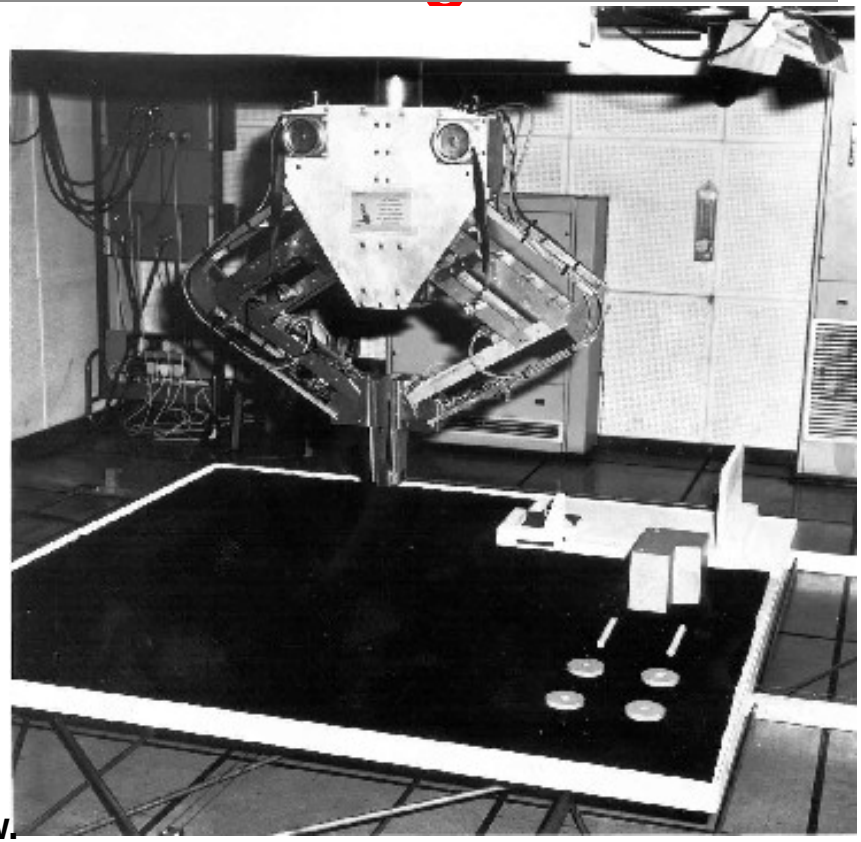
  Could such deliberative premeditation use the action schema (operator) with approximate, qualitative parameters instead of the more definite actual parameters that would be used if the action were performed?
- **think about various effects of actions, including changing effects of continuous processes**

NOTE: there are problems here partly analogous to problems of reference and identification in language, except that the mode of reference is not linguistic and what is referred to typically cannot be expressed in language because it is anchored in non-shared structures and processes.

(Internal 'attention' processes are partly like external pointing processes: virtual fingers.)

Compare Freddy the 1973 Edinburgh Robot

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



Freddy II could assemble a toy car from the components (body, two axles, two wheels) shown. They did not need to be laid out neatly as in the picture.

However, Freddy had many limitations arising out of the technology of the time.

E.g. Freddy could not simultaneously see and act: partly because visual processing was extremely slow.

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

There is more information on Freddy here

http://www.ipab.informatics.ed.ac.uk/IAS.html

http://www-robotics.cs.umass.edu/ pop/VAP.html

In order to understand the limitations of robots built so far, we need to understand much better exactly what animals do: we have to look at animals (including humans) with the eyes of (software) engineers.

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 (sometimes using laser range-finders).

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: the 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).**

2 A vision system has to be part of a larger architecture

In my ‘Talks’ directory

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

there are several presentations on architectures, and on a conceptual framework called CogAff for thinking about types of architectures supporting multiple kinds of functionality operating in parallel.

As an example of the application of the CogAff framework, it is conjectured that the human architecture has a kind of complexity illustrated very sketchily in the next slide.

There’s no time to explain this now, but many of the features of vision that are mentioned in the rest of this presentation depend on the fact that vision has multiple functions because visual mechanisms are connected to many different subsystems in the architecture.

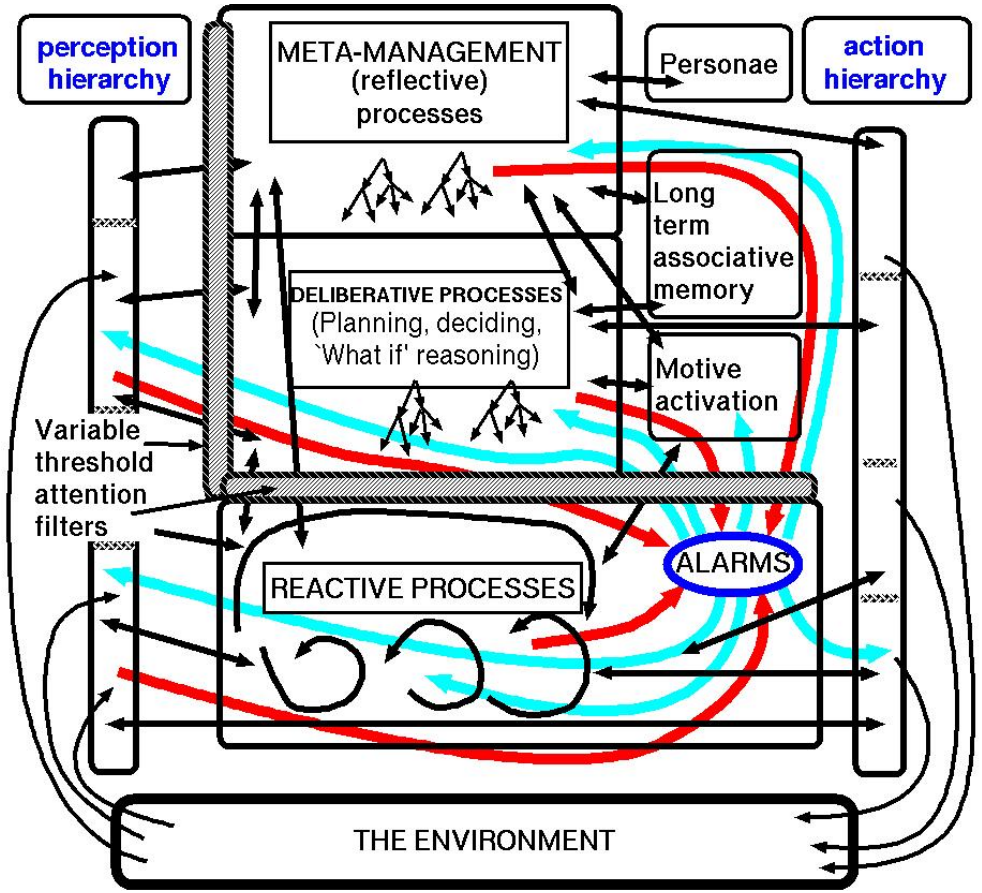
Because of this, evolution produced multilayered vision systems (and multilayered action systems) that, at least in humans, do not have a fixed structure but can be extended during learning and development, including learning to read, learning to understand abstract diagrams, and learning to see new kinds of affordances, e.g. the properties of hooks.

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

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

Many required sub-systems are not shown.

Different kinds of process simulation may go on in different parts of the architecture – some very old and widely shared, some relatively new and found in very few species.



A Shift of View

- For many years I assumed (like many other people) that if we could understand perception of static scenes we could later deal with motion.
- I also thought (as explained below) that perception of a static scene involved forming (static) descriptions of its contents (at different levels of abstraction), and that a theory of perception of motion might later be derived from that.

A Shift of View

- For many years I assumed (like many other people) that if we could understand perception of static scenes we could later deal with motion.
- I also thought (as explained below) that perception of a static scene involved forming (static) descriptions of its contents (at different levels of abstraction), and that a theory of perception of motion might later be derived from that.
- Then I learnt about Gibson's theory of affordances, which made it necessary to relate perception of static scenes to the *possibility of* (and constraints on) actions and their consequences that are not occurring, but might occur.
- For a while I assumed that a theory of perception of affordances could be tacked onto a theory of perception of structure by representing the perceived affordances as collections of something like condition-action rules associated with various parts of a scene.

A Shift of View

- For many years I assumed (like many other people) that if we could understand perception of static scenes we could later deal with motion.
- I also thought (as explained below) that perception of a static scene involved forming (static) descriptions of its contents (at different levels of abstraction), and that a theory of perception of motion might later be derived from that.
- Then I learnt about Gibson's theory of affordances, which made it necessary to relate perception of static scenes to the *possibility of* (and constraints on) actions and their consequences that are not occurring, but might occur.
- For a while I assumed that a theory of perception of affordances could be tacked onto a theory of perception of structure by representing the perceived affordances as collections of something like condition-action rules associated with various parts of a scene.
- In retrospect it seems silly to have forgotten that vision evolved in organisms embedded in a dynamically changing environment – so its primary function must be not to discover what exists in the environment, but what is happening in the environment, including the perceiver's movements and actions.
- Add the observation that what is happening, and what is potentially important to an organism, is not a unique process, but a collection of processes at different levels of abstraction, e.g. a wave moving horizontally towards the shore and millions of molecules mostly moving roughly up and down in the same place.

3 Example: A child playing with a train-set on the floor

The video mentioned above shows a child about three and a half years old doing things with a train set that surrounds him as he sits in the middle, turning this way and that, pointing at things behind him to answer questions, pushing the train through a tunnel, changing his position to replace a tree knocked down by the back of his head when he puts his head down to look through the tunnel.

http://www.cs.bham.ac.uk/~axs/fig/josh_tunnel.mpg (5MB)
http://www.cs.bham.ac.uk/~axs/fig/josh_tunnel_big.mpg (15MB)
(High resolution version.)

My claim that this child is running various simulations of things going on in the environment begs the question: 'What kind of thing is a simulation?'

My provisional answer is that anything that is capable of usefully representing a process can be called a simulation for present purposes, even if it is a static structure accessed sequentially.

Later I'll say more about what I do and do not mean.

NOTE: I am not making any use of Grush's distinction between 'emulation' and 'simulation', though it is possible that it will turn out that what I mean by 'simulation' is what he means by 'emulation'.

Snapshots from tunnel video

A child playing with his train illustrates the theory.



- **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?**

Vision is concerned with what is and is not happening in the environment – that's potentially of relevance to the perceiver: ongoing situations and processes.

Tunnel vision

Think about the child playing with and talking about his toy train, with track, tunnel and other things on the floor around him.

How many different levels of abstraction occur in

- **the processes he needs to perceive,**
- **the processes he needs to use in controlling his actions,**
- **the processes he needs to think about, explain, modify, predict, ...**

Is there a sharp division between

- **seeing geometric structures, relationships, changes and**
- **seeing causal and functional relations?**

  Is there a sharp distinction between what the child sees as caused by his action, and what he sees as merely happening in the environment?

  Could the same mechanisms represent both?

  Compare

  - **Movement of the truck he is holding and pushing**
  - **Movement of the truck adjacent to the one he is pushing**
  - **Movement of the trucks in the tunnel that cannot be seen**
  - **Reappearance of the front of the train from the far end of the tunnel**

We return to perception of causal relations later.

Background

- **There are many views of the nature and function(s) of vision, including the following:**
  - **Vision 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?)**
  - **Behaviour controls perception, including vision. (W.T.Powers)**
  - **Vision is unconscious inference (Helmholtz)**
  - **Vision is controlled hallucination (Max Clowes) Pretty close**
  - **Grush on Emulation theory of representation (BBS 2004)**
- **I'll try to present phenomena that require a richer deeper theory.**

  It will be evident that the new theory uses many of the above ideas, and assembles them with some new details. Some of the ideas are criticised.

Relationship with CoSy project

A change of view came while I was working on the CoSy project

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

I have been thinking about many of the problems for many years, but what made things click into place recently was examining very closely the perceptual and representational requirements for a robot manipulating 3-D objects on a table-top, e.g. watching a hand picking up a cup, or assembling a meccano model.

Try thinking about it yourself!

Using one or two hands, perform simple, everyday actions on cups, spoons, scissors, paper, string, a handkerchief, nuts and bolts, tin-openers, your food, a sweater you put on or remove ....

and watch very, very closely.

How can your brain represent the information you use, including

- **all the things and processes you see, as complex 3-D objects move while changing their shapes and mutual relationships,**
- **what you anticipate,**
- **your recollection of what just happened,**
- **your thoughts about what would have happened if you, or someone else, had done something different?**

PERHAPS YOU WILL INVENT THE SAME THEORY.

The theory is not totally new

There are many precursors of different kinds:

Some old philosophical theories of minds as idea-manipulators.
Kant's *Critique of Pure Reason* (1780) (Including his theory of mathematical knowledge)
Helmholtz: perception is unconscious inference
Kenneth Craik in 1943 (animals use predictive models)
Ulric Neisser and others (1960s): theories of vision as analysis by synthesis, and hierarchical synthesis.
Karl Popper (our hypotheses can die in our stead)
William T Powers: Behaviour controls perception.
Lots of control engineering using 'predictive' models.
Max Clowes: Vision is controlled hallucination
David Hogg's work on perceiving a walking person (1983)
My own work in the 1970s on multi-level perception and visual reasoning
Work by Tsotsos on motion perception.
Roger Shepard and others on mental rotation tasks.
Steve Kosslyn on imagery
Phil Johnson-Laird on reasoning with mental models
JJ Gibson on perceiving affordances (and his earlier ideas about 'perceptual systems')
Minsky's *Society of Mind* and other work.
Arnold Trehub: (1991) The Cognitive Brain
Alain Berthoz (2000) *The Brain's sense of movement*,
Murray Shanahan AISB 2005
Philippe Rochat, 2001, *The Infant's World*,
R. Grush, 2004, The emulation theory of representation: ... BBS, 27,

And probably more: but does any combine all the elements proposed here?

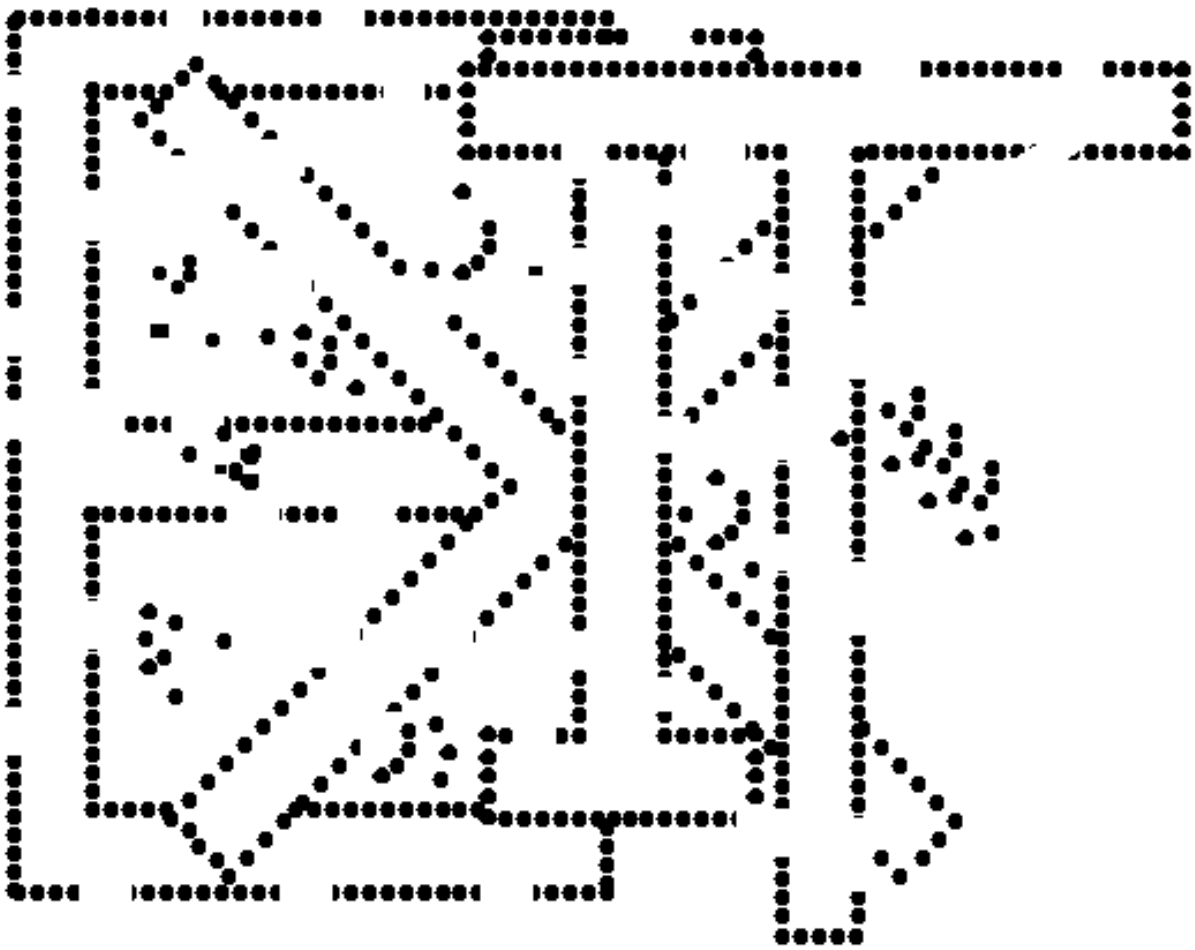
4 From structures (in the Popeye system) to processes

About 30 years ago a project at Sussex University explored some aspects of the theory that perception of complex and noisy structures could be facilitated by a visual architecture in which processes at different levels of abstraction, concerned with different ontologies, ran concurrently with a mixture of bottom up and top down control, including top-down control of attention.

But there was nothing in this about perceiving processes at different levels of abstraction, as is proposed here. Yet some of the ideas remain relevant.

How do we process noisy pictures? (1)

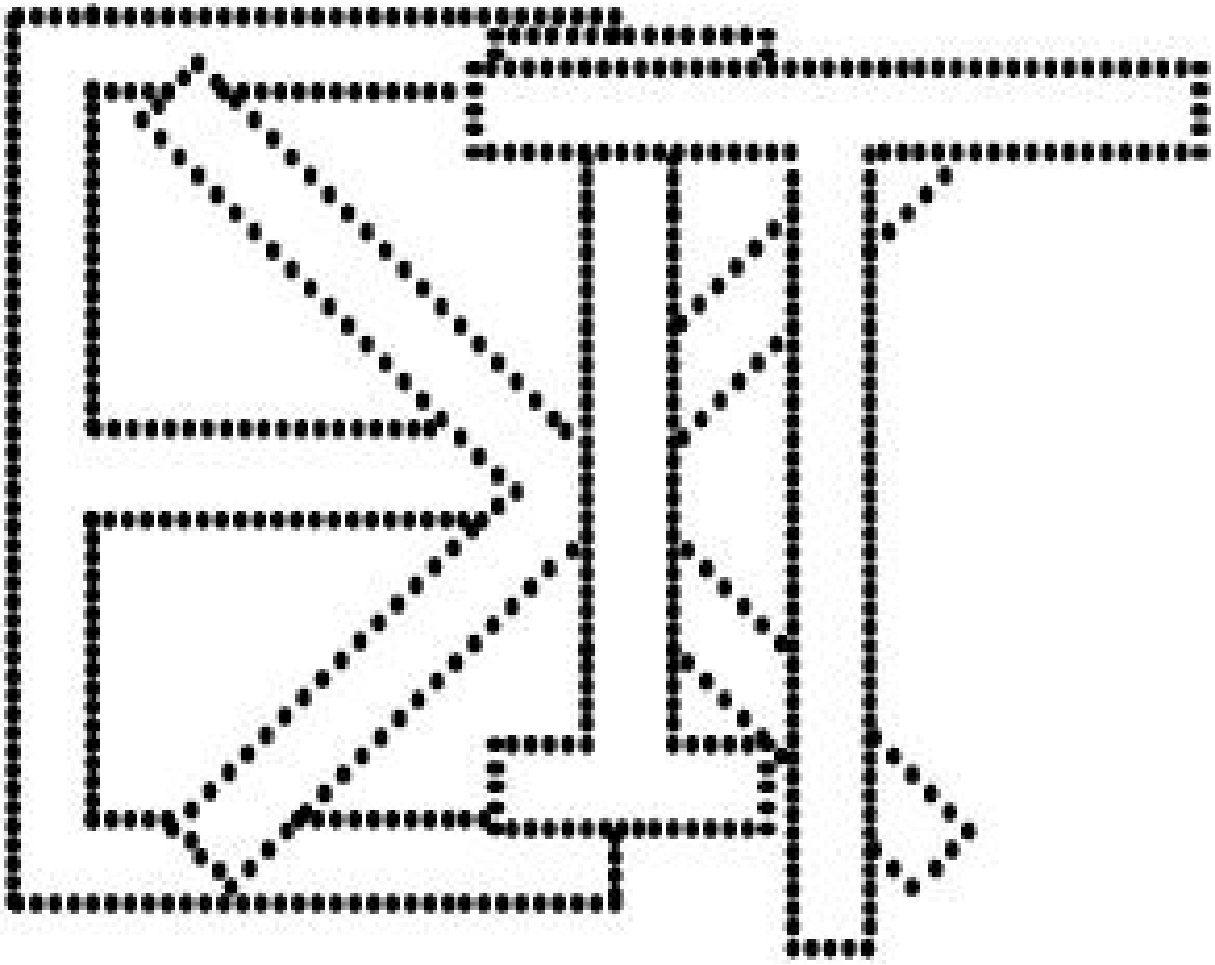
DO YOU SEE A WORD?



How do we process noisy pictures? (2)

How do we process noisy pictures? (3)

DO YOU SEE A WORD?



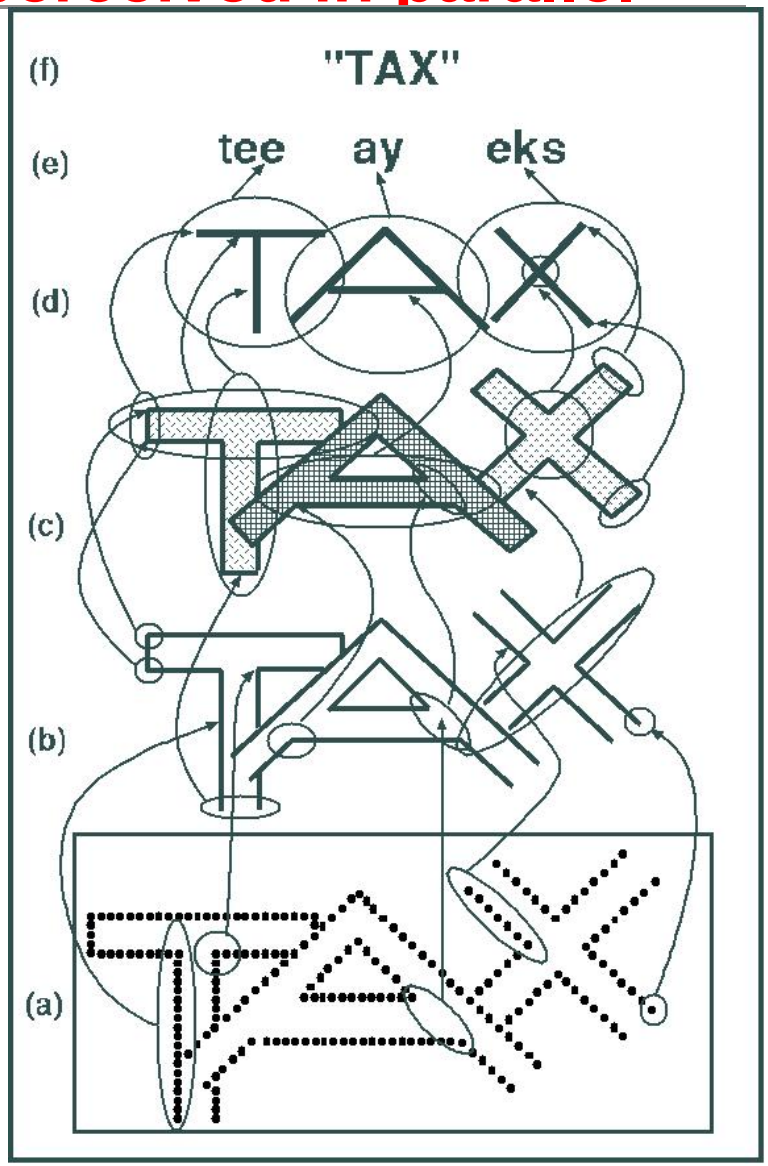
Multiple levels of structure perceived in parallel

Old conjecture: 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



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

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

Seeing a cup at multiple levels of abstraction

Many levels of structure and of affordances.

The identification of ‘objects’ is not fixed by the environment: e.g. thinking about different places to grasp.

But even that is not all that goes on in vision

From Structures to Processes

In the light of earlier observations we can replace the idea that

1. **seeing involves multi-level structures in partial registration 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.**

NOTE:

After developing these ideas I found that Philippe Rochat's book *The Infant's World* claims on pages 103-7 that there is evidence that even at 4 months infants are capable of 'dynamic imagery', used to predict trajectories of objects when the pass out of view.

The walking man

- **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 an image, rather than claiming to show how to interpret a sequence of images as representing a process.

- **Making Popeye see dotty images of moving overlapping laminas forming different words at different times would have been a very different task ....**

How to see a static scene as a process

If all this is right, our ability to see processes is used even when we look at a static scene:

it's just that then the process is one in which nothing changes.

- **But if something started changing we would see it, using the same mechanisms as were previously perceiving the static configuration.**
- **A static scene is just a special kind of process, in which nothing changes.**
- **Whether the things change or not the system has to be prepared for many possibilities.**
- **Thus perception of a static structure already involves perception of possibilities for motion (mostly latent: the simulation capabilities may be turned on if motion occurs, and left dormant otherwise).**

This could be seen as a minimal notion of affordance.

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

- **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 to keep track of (partly by running simulations?) the actions of others at the same time as planning and performing one's own actions.**
- **What are the evolutionary precursors of this, e.g. in hunting animals and prey of hunting animals, including parents defending young from predators?**
- **Concurrency is also important in social learning, since many social interactions are concurrent rather than simply based on turn-taking: e.g. dancing, old friends embracing, lifting or pushing a heavy article, and mating.**
- **Conjecture: our architecture evolved to support at least three sorts of concurrency:**
  - **Perceiving multiple concurrent external processes**
  - **Representing the same process at different levels of abstraction**
  - **Different concurrent actions in an individual, such as walking (including posture control), working out where to walk, discussing philosophy with a companion, using different parts of the information-processing architecture.**

Liberation from the here and now

CONJECTURE

The same mechanisms (or similar mechanisms produced using evolution's 'duplicate then differentiate' strategy) can be used

(a) Without using sensor-specific or motor-specific representations

(b) In relation to things that are not currently perceived

- **Past**
- **Remote**
- **Future**

Contrast:

multi-modal integration vs a-modal abstraction

Contrast:

Learning about (intra-somatic) sensorimotor contingences

vs

Learning about objective (extra-somatic) condition-consequence contingencies

We are not insects

The vast majority of animals (microbes, crustaceans, fish, reptiles....) may be able to get by with much less powerful and flexible perceptual systems.

They may always be involved in control of current actions (including quite sophisticated dynamical systems with predictive control mechanisms – using feedforward loops – e.g. flying insects).

But what humans and a few other species goes far beyond that: and much research on vision and robotics, including some research in neuroscience(?), does not take account of the requirements – e.g. to be able to remember what you did, to understand what went wrong, to think about what someone else may do, to plan several steps ahead....

Theories of insect intelligence may not be adequate for chimps, cheetahs and crows, let alone humans.

Note: all this may be unfair to some insects.

(And what about the octopus?)

Simulation capability exceeds behavioural capability

If human brains (and perhaps others) can construct and run simulations of processes of many kinds, there is no need for each one to be closely related either to the specific motor system that would be used to produce such processes or to the sensory systems that would be used to perceive such a process.

After all, we can perceive many processes we cannot produce, e.g. waterfalls – and we shall later give examples of perceiving and thinking about 'vicarious affordances', i.e. affordances for others.

So we have an ability to experience and appreciate processes that are richer and more complex than anything we can produce using our own bodies.

- **Evolution apparently 'discovered' the benefits of structural and causal disconnection between representation and thing represented, long ago (in a subset of animals only?): can we replicate this in our designs?**
- **Compare**
  - **the ability of a prey animal to think about what a predator might do**
  - **the ability of a composer to think up a multi-performer composition, and specify it in a musical score.**
  - **the ability of a general to prepare orders for various concurrently active platoons.**
  - **the ability of some programmers to design, implement, and debug programs involving concurrent processes (e.g. operating systems).**

So....

Many current theories of embodied cognition ignore the extent to which evolution discovered the power of disembodied cognition for a small subset of species

Infant humans seem to have minds with learning and developmental capabilities that can use a variety of different bodily forms available from infancy to achieve a common adult humanity.

For example, consider the thalidomide babies born limbless in the 1960s, and the artist Alison Lapper, celebrated here

http://www.ldaf.org/pages/dail/dailarticles.htm#lapper
http://www.mymultiplesclerosis.co.uk/misc/alisonlapper.html

She can clearly see many structures and processes that can be seen by people with normal arms and legs.

Reminding the audience: relevant things you probably know

There are many aspects of our everyday experience that people may or may not notice that seem to involve this ability to run some sort of simulation of environmental processes.

So this is not really a theory that's new to you, even if you previously never thought about it.

- **E.g. when you see something moving behind an opaque object you don't see the moving object as being truncated – you see it as having a hidden portion that continues to move (like the child in the video pushing his train into a tunnel), and typically you know roughly where the hidden parts are as the motion continues (though of course stage conjurers can fool us because we are not infallible).**
- **Many cartoons and jokes depend on our ability to run simulations derived from the information presented, e.g. pictorially or verbally.**
- **Droodles depend on this ability too. In fact many/most(?) forms of visual art do.**

Some cartoons showing 'snapshots' of extended processes follow. Some project into both future and past, some only one or the other.

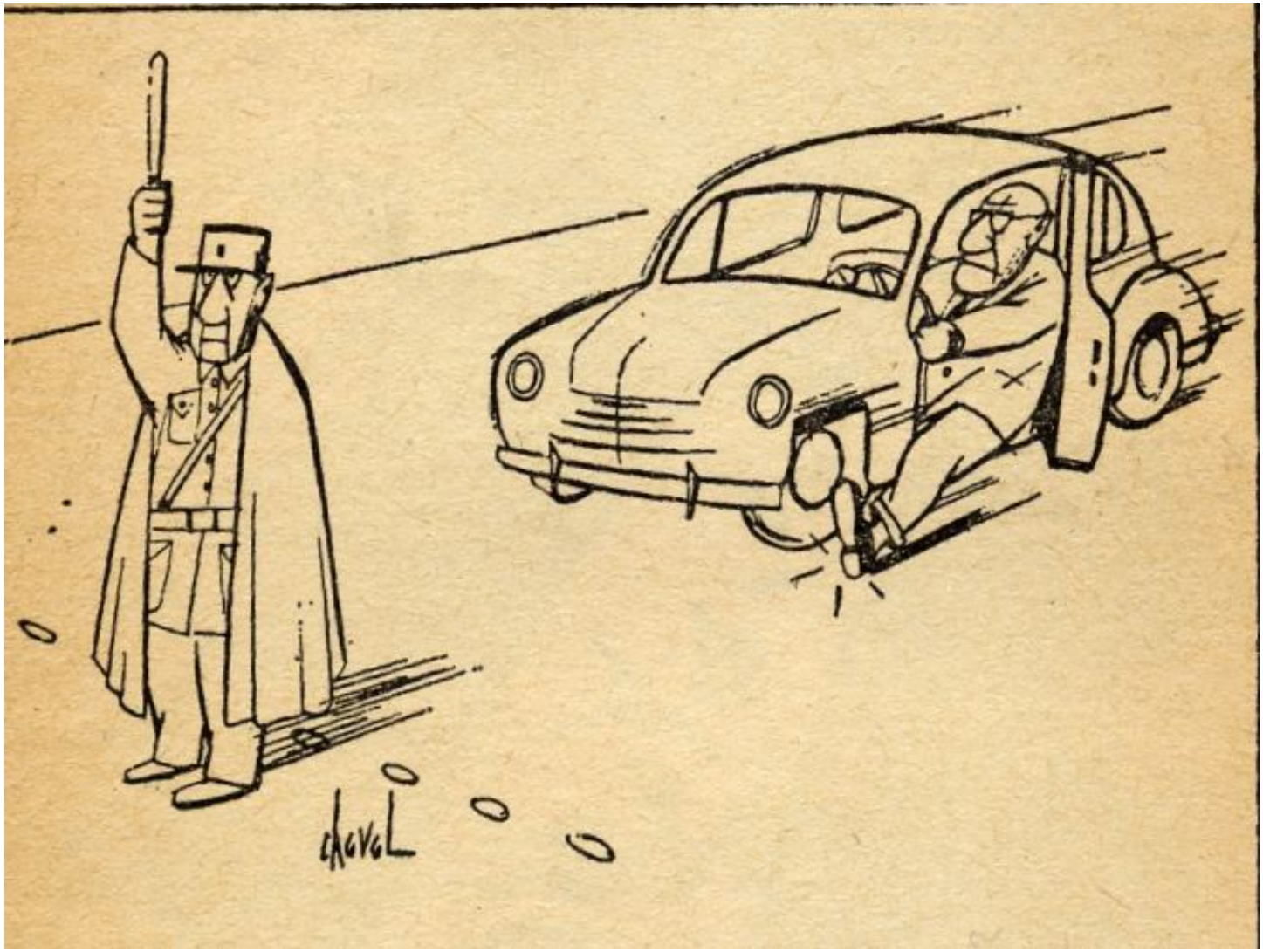
5 Cartoons and miming

‘French Cartoons’ Published 1955

Ed William Cole and Douglas McKee, : Panther Books

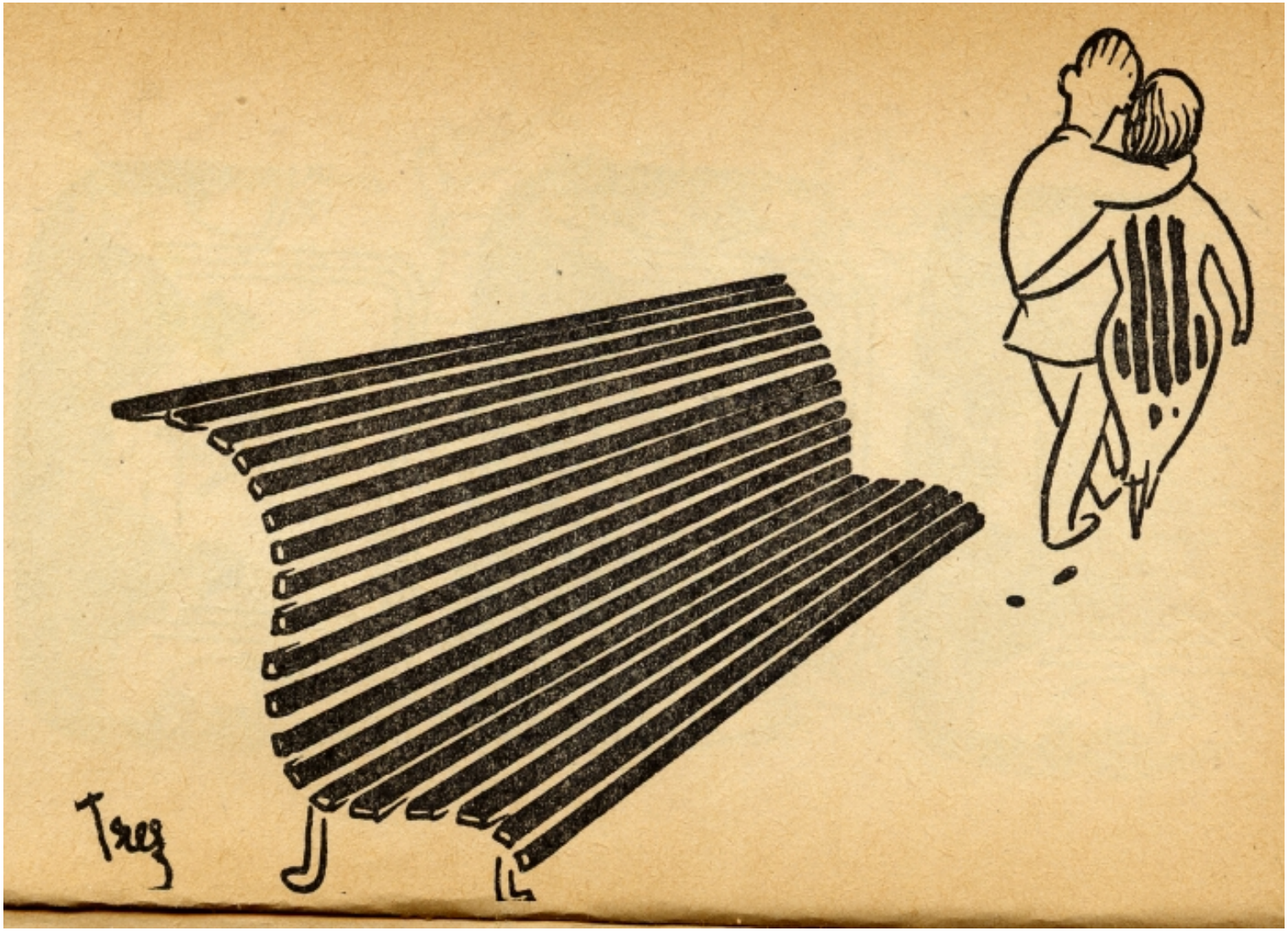
When you look at the cartoons that follow, what past and future processes come to mind and how do they relate to details of the scene?

Mostly future



Another kind of footbrake ???

Mostly past



Understanding the picture involves ‘running a simulation’ but at a high level of abstraction with many details of the previous history left out.

We produce joke actions also

Using a tennis ball and badminton shuttlecock to simulate eating an ice-cream – he never actually licked the ball.

We often use external simulations, including gestures, diagrams, working models. However most of our examples below will be cases of purely internal simulation.

Perhaps a major function of play in young mammals is developing simulation capabilities through learning about different things to simulate (as opposed to developing motor skills, muscles, etc.)

Evolution (and processes in individual development) somehow gave us the ability to make use of either internal or external objects, when running simulations. My 1971 IJCAI paper claimed that reasoning with diagrams is essentially the same thing whether done on paper or in the mind.

Brain mechanisms for this are still waiting to be discovered.
(See the interesting discussions in BBS paper and commentary by R.Grush, 2004 – found after much of this had been written).

Sensorimotor vs Condition consequence contingencies

Insects may be restricted to learning conditional probabilities relating total sensory and motor signal-sets.

In some cases that would be combinatorially explosive – e.g. all the ways of perceiving grasping, whether done with mouth, or left hand or right hand, or two hands holding an object.

An organism that can abstract from all the intra-somatic sensorimotor details and represent extra-somatic relationships between surfaces and their consequences (e.g. if something moves) independently of how movements are produced or sensed has a great advantage in generality and economy.

That includes being able to perceive and think about actions done by others: perceiving vicarious affordances.

So mirror neurons should have been called 'abstraction neurons'.

CONJECTURE: this ability to represent objective structures and processes, a kind of disembodiment, was a major evolutionary development.

It's not clear whether that is present at birth – though much else is.

6 Perceiving causation

Two kinds of causation: Humean (probabilistic, evidence based) and Kantian (deterministic: based on hypothesised structures)

Perceiving causation

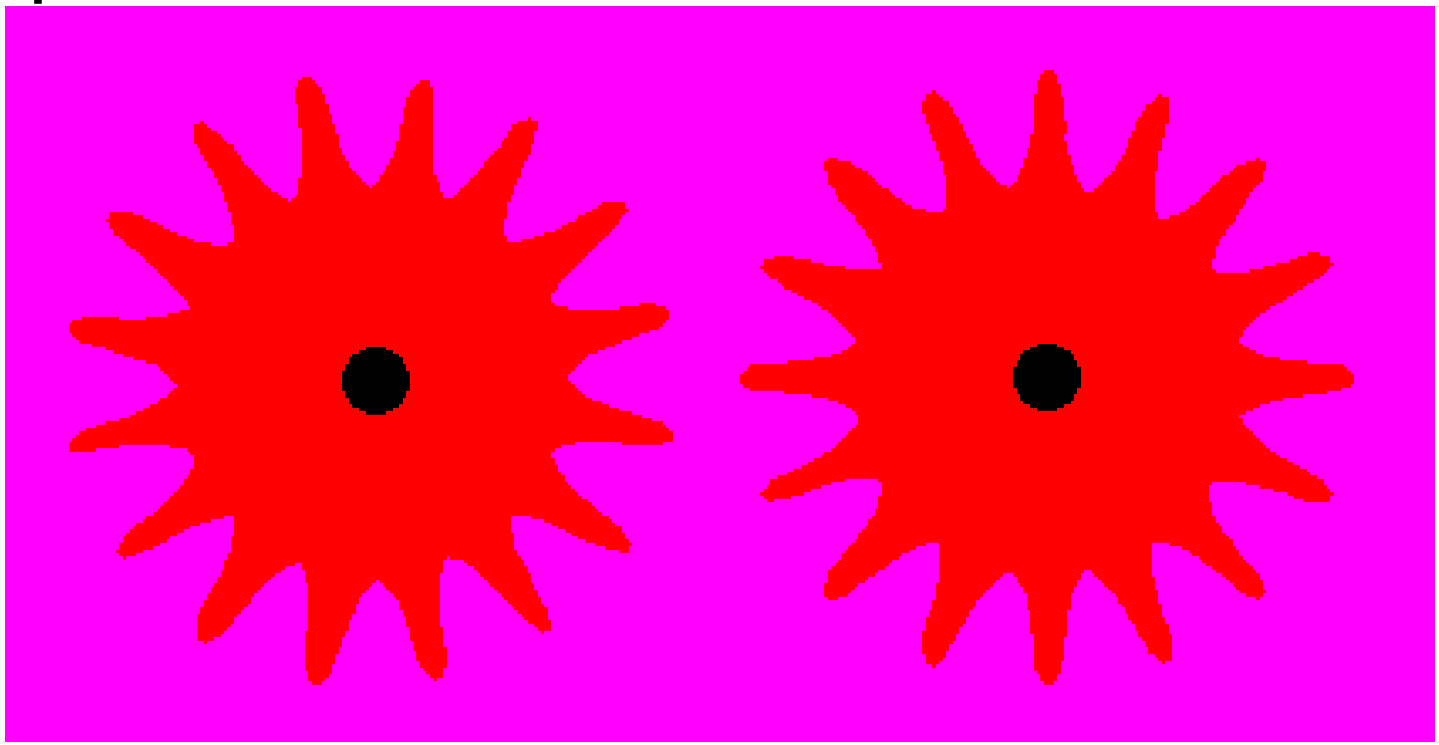
Our ability to perceive moving structures, and our meta-level ability to think about what we perceive, is intimately bound up with perception of causation and affordances.

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: we infer it from statistics.



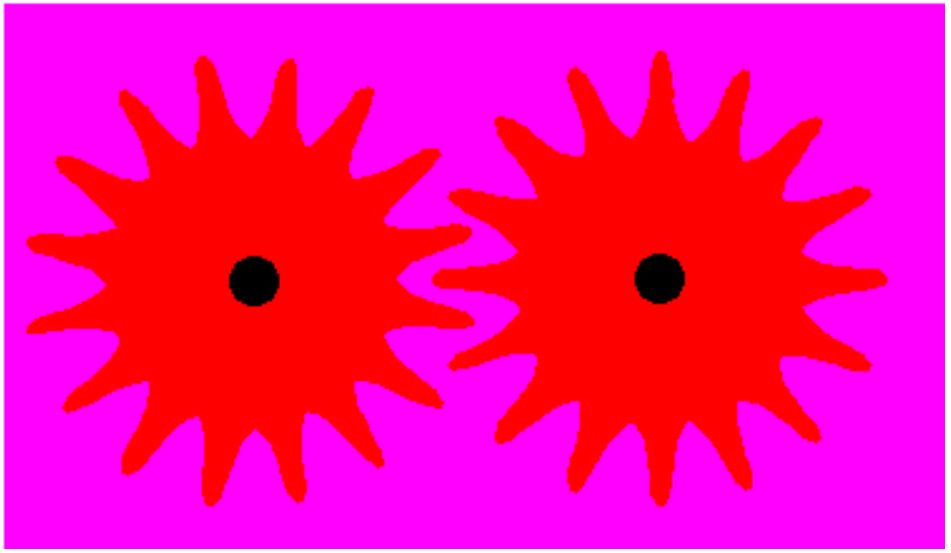
Can you tell by looking what will happen to one wheel if you rotate the other about its central axis?

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).

Humean and Kantian Causation

- When the only way you can find out what the consequence of an action will be is by trying it out to see what happens, you may acquire knowledge of causation based only on observed correlations. This is 'Humean causation' – David Hume said there was nothing more to causation than constant conjunction, and this is now a popular view of causation: causation as statistical (often represented in Bayesian nets).
- However if you don't need to find out by trying because you can see the structural relations (e.g. by running a simulation that has appropriate constraints built into it) then you are using a different notion of causation: Kantian causation, which is deterministic and structure-based.
- I claim that as children learn to understand more and more of the world well enough to run deterministic simulations they learn more and more of the Kantian causal structure of the environment.
- Typically in science causation starts off being Humean until we acquire a deep (often mathematical) theory of what is going on: then we use a Kantian concept of causation.
- This requires learning to build simulations with appropriate constraints.

For more on this see this talk

http://www.cs.bham.ac.uk/research/projects/cosy/papers/#pr0505
COSY-PR-0506: Two views of child as scientist: Humean and Kantian

7 Geometry-based causation

Perceiving causation in changing geometric structures.

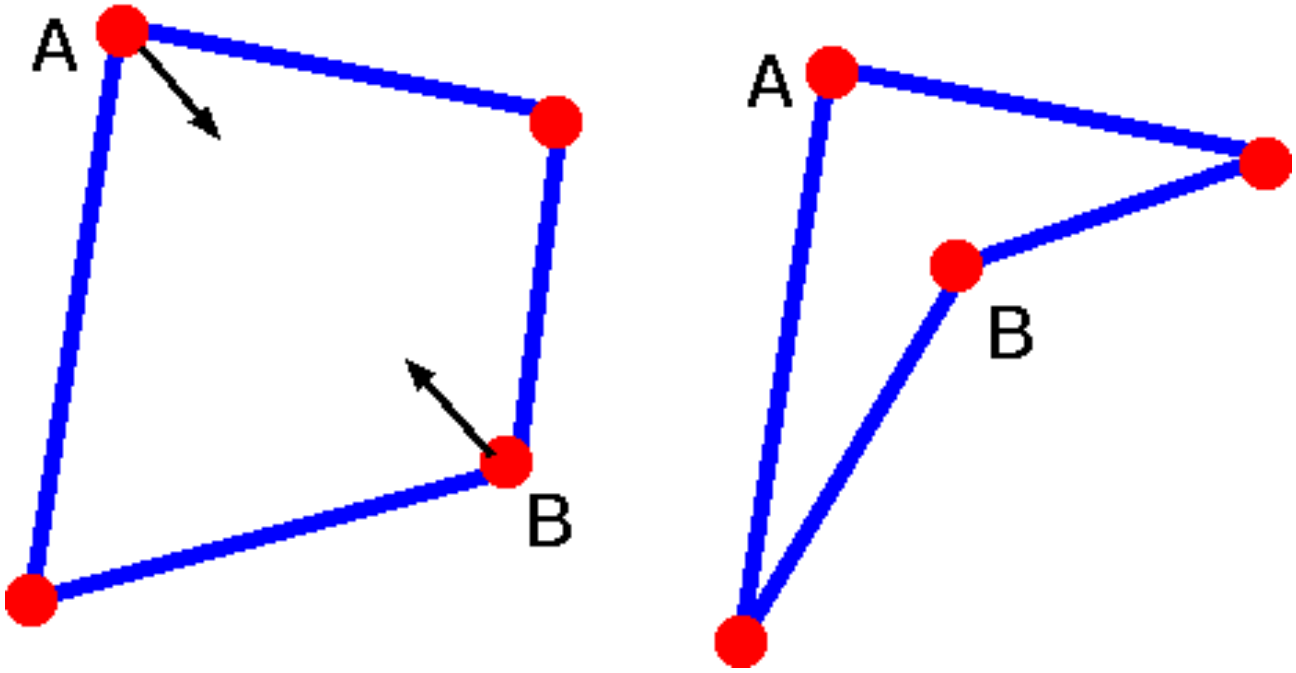
We can often see and understand consequences of motion of one part of a structure, including being able to predict effects on other parts.

But not when the structures are too complex, or have too many degrees of freedom.

Every kind of human competence has fairly low complexity limits, even though humans are enormously flexible in deploying and combining their competences.

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. ???



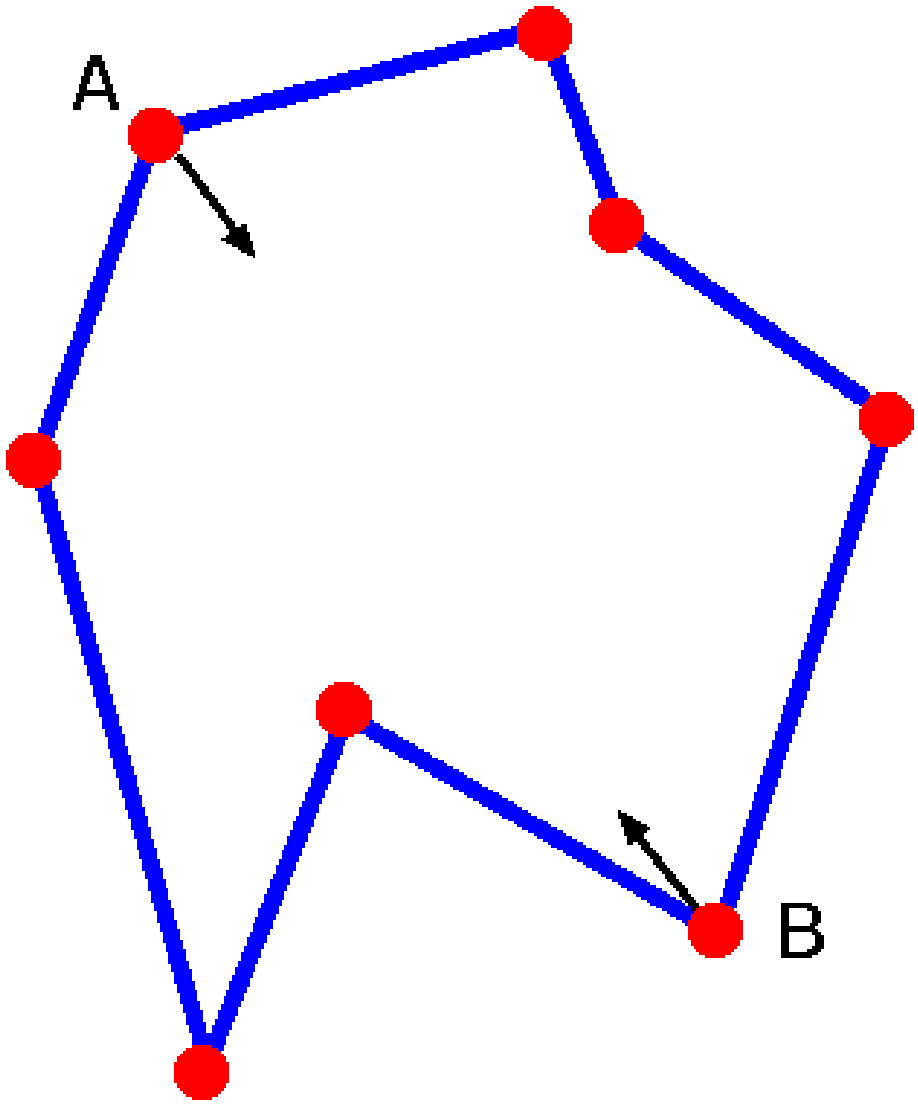
- **What happens if one of the moved joints crosses the line joining the other two joints?**
- **We can change the constraints in our simulations: what can happen if the joints and rods are not constrained to remain in the original plane?**

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.

The moral?

- **Processes we can imagine, see, think about are not necessarily related to what our own bodies can do: the importance of embodiment is currently being grossly oversold.**

- **Humans do not scale up, though we do 'scale out' – many different competences are available that can be combined in different ways.**

How they are acquired, represented, stored, accessed and combined, is largely unknown.

(That's one difference between what I am saying and 'global workspace theory', which doesn't address those questions.)

8 Multi-modal perception of causation

We can combine information from different senses to produce a running simulation of what is going on.

(As Grush (2004) points out.)

In some case what is represented in the simulation is not sensed at all, until some time after the simulation starts.

Mixed mode input to an integrated simulation

- What you hear, like what you see, can be a process occurring in the environment, for instance hearing someone moving round you when your eyes are shut.
- If you are sitting in a room with a door opening into a corridor, subtle aspects of the changing sound of footsteps (which you process unconsciously) may produce a percept of an unseen person moving to the door, so that you know when he will become visible – a device used often in movies.
- Likewise when you see the unseen person's shadow changing.
- So the process you hear occurring and the things you see occurring may exist in the same integrated simulation — which is just as well since they exist in the same spatial environment.
- Likewise what a dentist sees and feels with the probe as she looks into the patient's mouth need to be in the same perceived part of the world, and when you use a hand to feel the underside of the table you are looking at you see and feel the same table.
- If you push a pencil up through a hole in the table you see and feel the same moving pencil.

Sensory modality and mode of representation

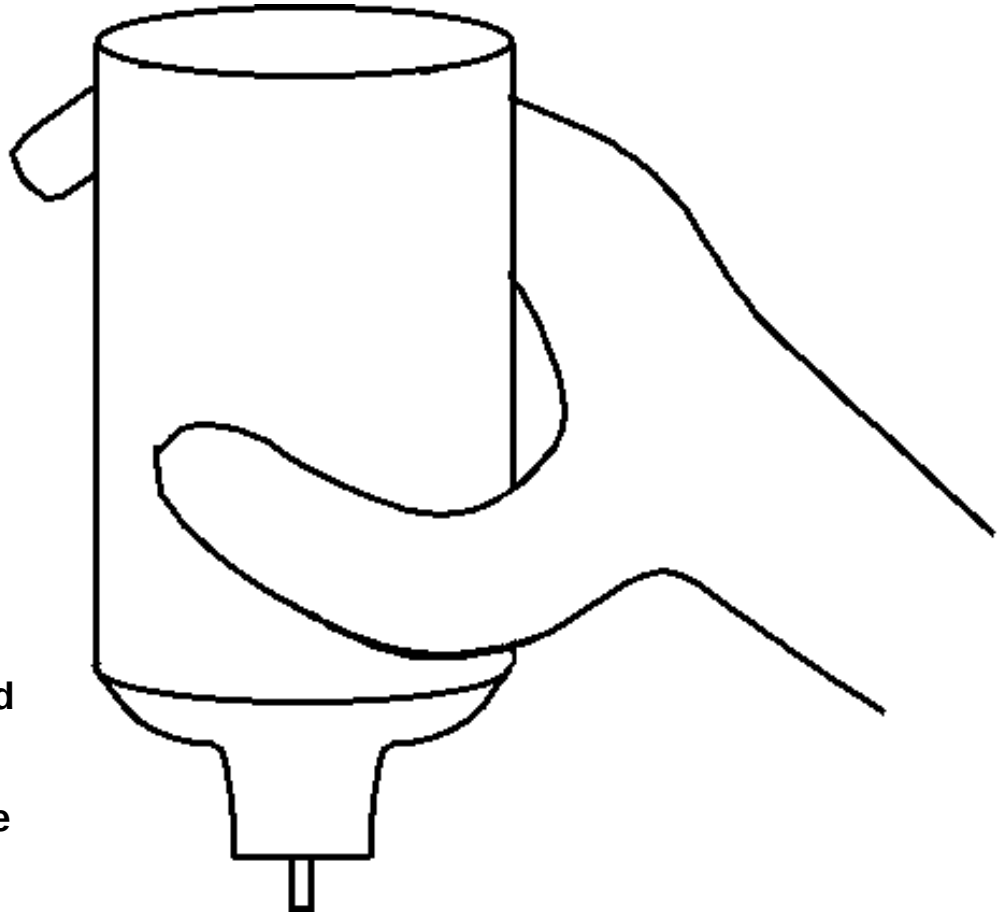
- Sensory modality driving a simulation need not determine the nature of the percept.
- A unitary, amodal, percept of a process can be driven by input from diverse sensory modalities – e.g. seeing, hearing, feeling the same thing happening.
- What is simulated does not determine the nature of the medium used to implement the simulation, as long as it has a rich enough structure and appropriate mechanisms to create, modify, access and use the contents.
- 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 support the required functionality.

Visual reasoning about something unseen

An example of disconnection between simulation and 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 the 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.

9 Many distinct competences have to be learnt

The competences described above are not all present at birth, though some of the mechanisms required to acquire them are (while other learning mechanisms have to be produced by learning).

They are not pre-configured by genetic mechanisms, like innate abilities or innate latent genetically-determined competences that emerge long after birth (e.g. sexual competences, or migration in some birds).

The learnt, meta-configured competences need powerful bootstrapping mechanisms.

See

A. Sloman and J. Chappell (2005), The Altricial-Precocial Spectrum for Robots, *Proceedings IJCAI'05* pp. 1187–1192.
http://www.cs.bham.ac.uk/research/cogaff/05.html#200502

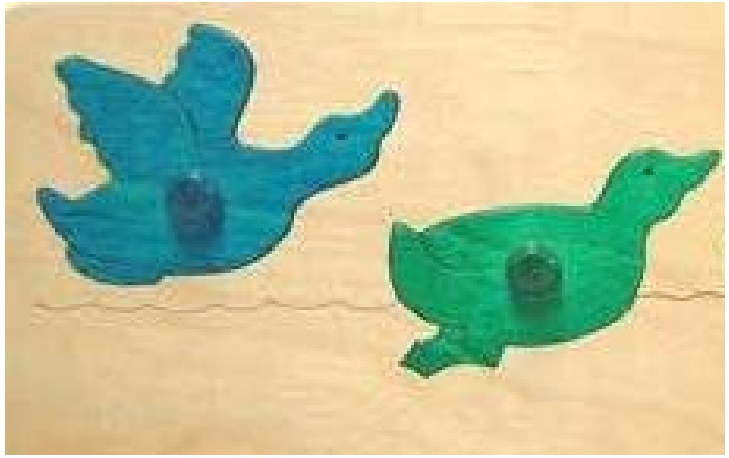
A. Sloman and J. Chappell (2005), Altricial self-organising information-processing systems, *AISB Quarterly*, 121, Summer 2005, pp. 5–7,
http://www.cs.bham.ac.uk/research/cogaff/05.html#200503

What the bootstrapping mechanisms achieve is extremely dependent on what is in the environment (including the culture), which is why altricial species with many meta-configured competences can differ enormously in what they know and can do, unlike precocial species, in which most competences are pre-configured, like deer which run with the herd soon after birth.

The examples that follow indicate some of what a child has to learn to see, before it can control its actions so as to achieve its goals, like inserting a puzzle piece where it belongs.

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.

CONJECTURE

In the first five years

- a child learns to run at least hundreds, possibly thousands,
- of different sorts of simulations,
- using different ontologies
- and different kinds of constraints on possible motions

  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

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).

10 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.

Contributors to simulation features

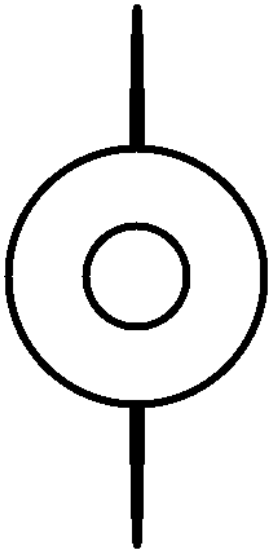
- We have so far seen that both shape and material can contribute to features of a simulation, including the constraints on what can and cannot change and what the consequences of change are.

- Another thing that can be important is viewpoint.

  E.g. viewpoint can interact with opacity of materials, as well as with the mathematics of projection from 3-D to 2-D.

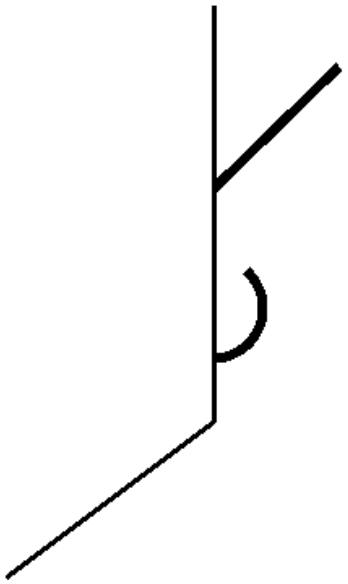
Sometimes a simulation includes a viewpoint

Droodles illustrate our ability to generate a simulation (possibly of a static scene) from limited sensory information (sometimes requiring an additional cue, such as a phrase ('Mexican riding a bicycle', or 'Soldier with rifle taking his dog for a walk').



In both of these two cases the perceiver is implicitly involved: one involves a perceiver looking down from above the cycling person, whereas the other involves the perceiver looking approximately horizontally at a corner of a wall or building.

In both cases the interpretation includes not only what is seen but also occluded objects: the simulation depends on knowing about opacity.



This does not imply that we have opaque objects in our brains: merely that opacity is one of the things that can play a role in the simulations, just as rigidity and impenetrability can.

The general idea may or may not be innate, but creative exploration is required to learn about the details.

We can see things from more than one viewpoint

- **Vicarious affordances: a parent watching a child needs to be able to see what is and is not possible in relation to the child's needs, actions, possible intentions, etc. (It is also useful to be able to perceive a potential predator's affordances.)**
- **This may include such things as visualising the scene from the child's viewpoint, including working out what the child can and cannot see – and the possible consequences of the child seeing some things and not seeing others.**
- **Some people can draw pictures of how things look from some other place than their current location.**
- **This ability to contemplate the world from multiple viewpoints, not just one's own current viewpoint, is essential for planning, since at some future state in the plan one's location and orientation could be very different from what it is now, yet it still needs to be reasoned about in extending the plan.**
- **The ability to perceive and use information about 'vicarious' affordances (affordances for others) and the ability to perceive affordances for oneself in the past (e.g. thinking about a missed opportunity) or future (planning to use opportunities that have yet to be created) may use the same mechanisms because both are disconnected from current viewpoint.**

  Could that be the main point of substance behind all the fuss about "mirror neurons"? They should have been called abstraction neurons.

Seeing things from the viewpoint of your hand

The importance of hand-eye uncoordination!

- The evolution of body-parts for manipulation that can move independently of a major sensor perceiving what's happening (hands *vs* beak or mouth) had profound implications for processing requirements.
- Most animals are restricted to doing most of their manipulation with a mouth or beak, which cannot move much without the eyes moving too.
- If your eyes move as your gripper moves, because they are closely physically connected, then the sensorimotor contingencies linking actions and their sensory consequences will have strong, useful regularities that can be learnt and used.
- If a gripper can move independently of the eyes then the variety of relationships between actions and sensed consequences explodes.

  The explosion can be reduced by modeling action at a level of abstraction removed from sensory changes: e.g. by representing actions as altering 3-D structures and processes (including subsequent actions), independently of how they are sensed.
- The mapping between sensory data and what is perceived becomes very indirect, and there may need to be several intermediate layers of interpretation: perception becomes akin to constructing a structured theory to explain complex data. (Compare the 'dotty picture' example, above.)
- This is one of many reasons for NOT regarding perception as simply concerned with detecting sensorimotor contingencies.

Seeing from no particular viewpoint

Dealing with a changing scene perceived by a moving observer may, for some purposees, require a representation of what is happening that is viewpoint independent as well as being modality independent.

Sensorimotor 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 sensorimotor 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.**

11 No good theories about shape perception exist

A huge amount of work on machine vision totally ignores shape and is concerned only with recognition, classification, prediction, or tracking, more or less treating the world as two-dimensional.

However there are some attempts to get machines to perceive shape.

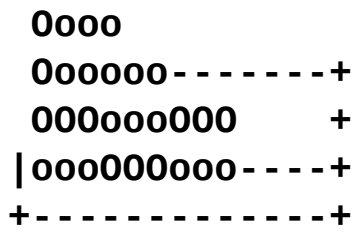
Unfortunately these mostly seem to use inadequate requirements for shape perception. E.g. using vision and laser-scanning or whatever, to produce a detailed 3-D model of space occupancy which can be given to computer graphics programs to project images from any viewpoint in different lighting conditions may be very useful for many applications (e.g. medical imaging, and computer games) this does not give the computer a kind of understanding of shape that is required for manipulating objects.

Structures vs combinations of features

It is important to understand the difference between

- **Categorising**
- **Perceiving and understanding structure.**

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



What is seeing without recognising?

There's a huge amount of work on visual recognition and labelling e.g. statistical pattern recognition. (Using totally arbitrary collections of benchmark images.)

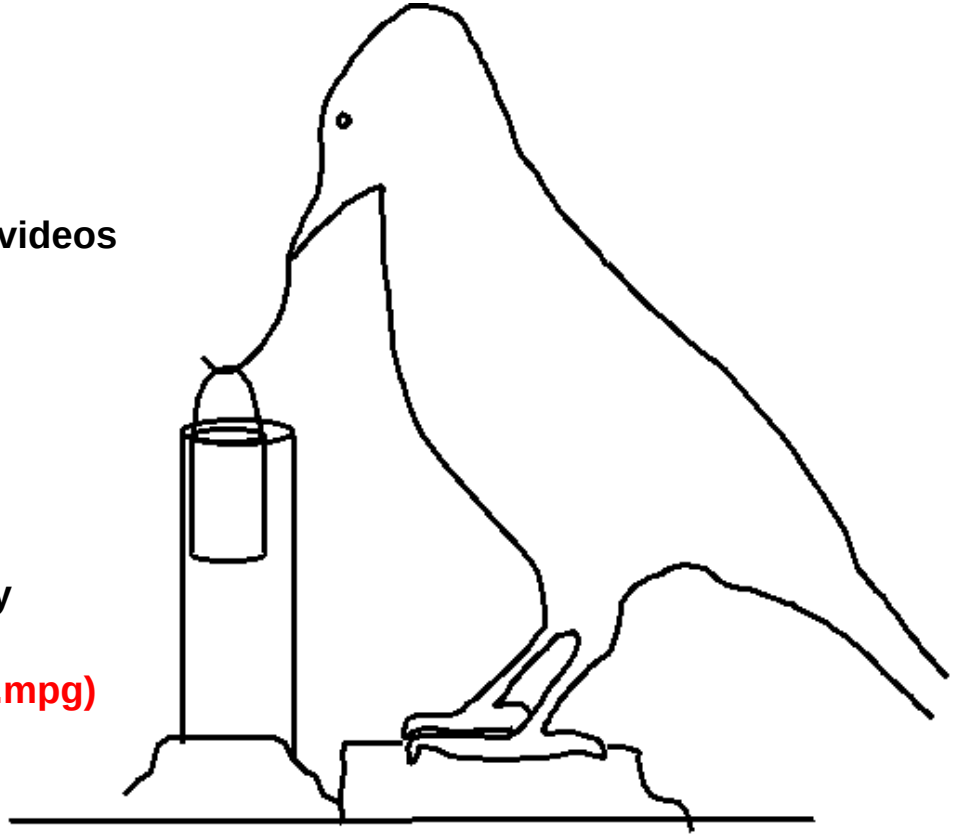
But does that tell us anything about perception of structure?
Much work on vision in AI does not get beyond categorisation.

There is some work that attempts to identify structure from visual images, but the form in which structure is represented is merely a volumetric model, which may be very suitable for generating graphical displays from different viewpoints, but does not include any understanding of the structure by the computer – it leaves the main representational problems unsolved.

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

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?

Understanding how hooks work

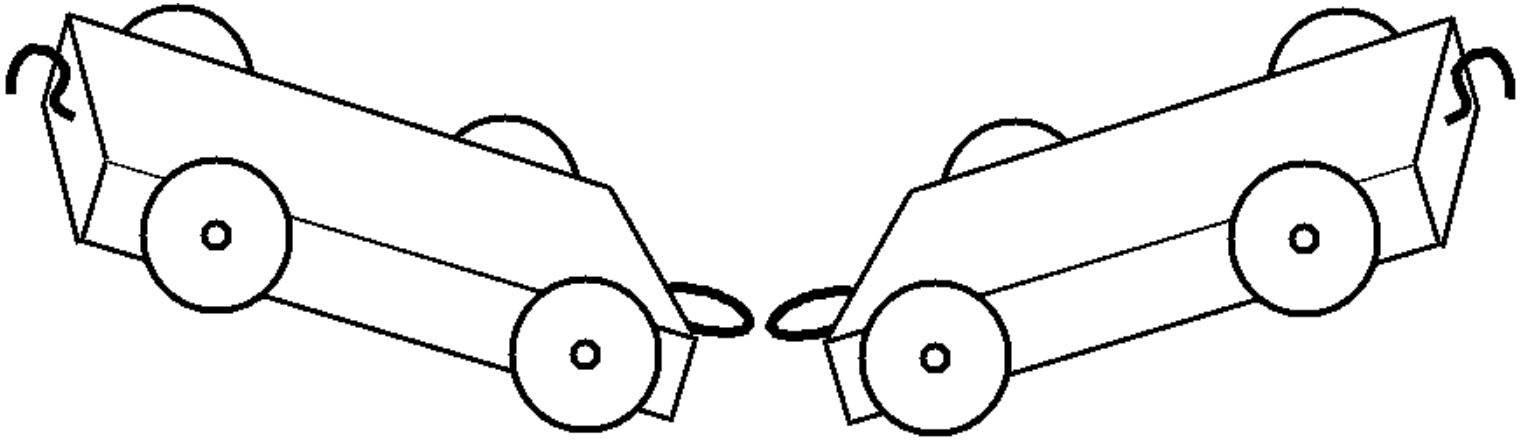
- Betty seems to understand how hooks work when she uses hooks to lift a basket of food out of the glass tube.
- The depth of understanding seems even greater when she demonstrates her ability to make hooks from straight pieces of wire in several different ways. I have also seen her make a hook from a long thin flat strip of metal.
- The behaviour is clearly not random trial and error learning behaviour: she seems to know exactly what to do, even though she does things in slightly different ways, e.g. making hooks using different techniques.
- Note that in Betty's environment far more distinct motions are possible than in the multi-rod linkage a few slides back: how does she confidently select a course through the continuum of continua?

  The answer cannot simply be: by running a simulation, because the simulation might have the same problem of under-determination.
- A young child does not start off understanding how a hook and a ring can interact in such a way as to allow the hook to pull the ring and what it is attached to.
- At some stage that (Kantian) understanding develops.
  But I don't think anyone knows how – even if some psychologists know when.
- The next slide points to a video showing a child who has not yet got there.

12 A child can appear less competent than a crow

We next show a video of a 19 month old child who is competent in many ways but seems to fail to understand how a hook and ring are used to join up a toy train.

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
  The date is June 2003, when he was 19 months old. (Born 22 Nov 2001)

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.

13 Running 2-D or 3-D simulations to answer questions

Perhaps the child who fails to join up the train does not understand because he has not yet learnt to simulate processes in which a hook and a ring form a connection that is useful for pulling.

Why not? Why are some competences innate, and some learnt. Why are some learnt very early and some only later.

Maybe we still have to understand the dependency relations between hundreds, or thousands, of sub-competences.

There are many problems we can solve, by running 2-D or 3-D simulations.

Some examples follow.

Simulating potentially colliding cars

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

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

Whereabouts will they meet – more to the left or to the right, or in the middle?

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 – more to the left or to the right, or in the middle?

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

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

Me: Why?

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

What is missing?

- **Knowledge?**
- **Appropriate representations?**
- **Procedures?**
- **Appropriate control mechanisms in the architecture?**
- **A buggy mechanism for simulating objects moving at different speeds?**

Mr Bean's underpants

This paper (from a conference on thinking with diagrams in 1998)

http://www.cs.bham.ac.uk/research/cogaff/00-02.html#58

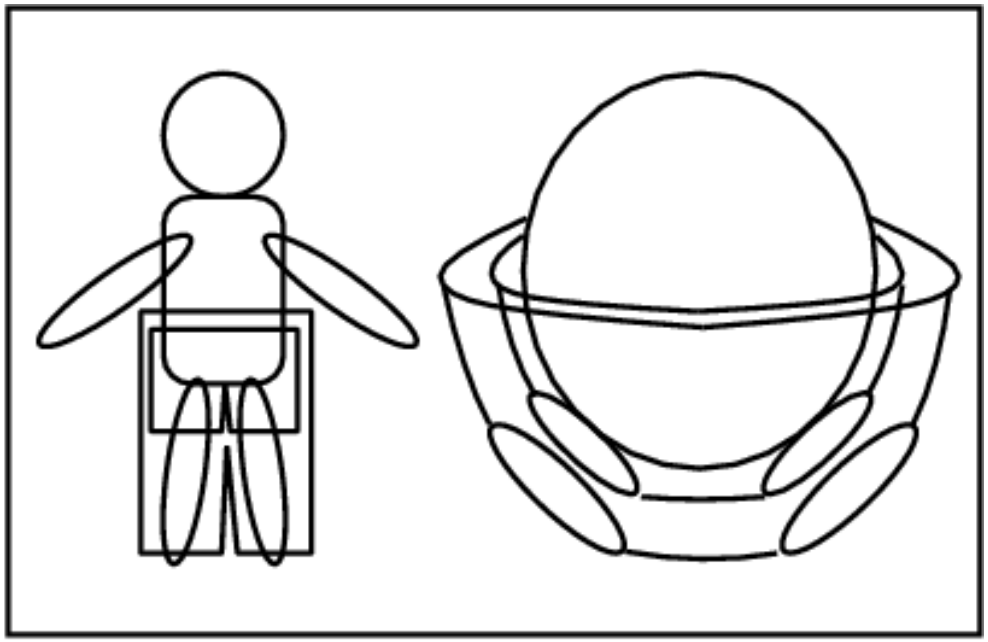
discusses how we can reason about whether Mr Bean (the movie star) can remove his underpants without removing his trousers.
People often don't see all the possibilities at first.

The paper discusses how changing the simulation to a topologically 'equivalent' one can help us count the possible ways to perform the task.

Children can learn to perform such actions (as party tricks) physically long before they can reason with the mental simulations.

What changes as the simulation ability develops?

In part it seems to require an introspective ability to understand the nature of the simulations we use.



See

Jean Sauvy & Simonne Sauvy *The Child's Discovery of Space, From Hopscotch to Mazes: an Introduction to Intuitive Topology* (Translated P.Wells 1974).

KANT'S EXAMPLE: 7 + 5 = 12

Kant claimed that learning that 7 + 5 = 12 involved acquiring *synthetic* (i.e. not just definitionally true) information that was also not *empirical*. I think his idea was related to the simulation theory of perception – but I am guessing.

You may find it obvious that the equivalence below is preserved if you spatially rearrange the twelve 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?

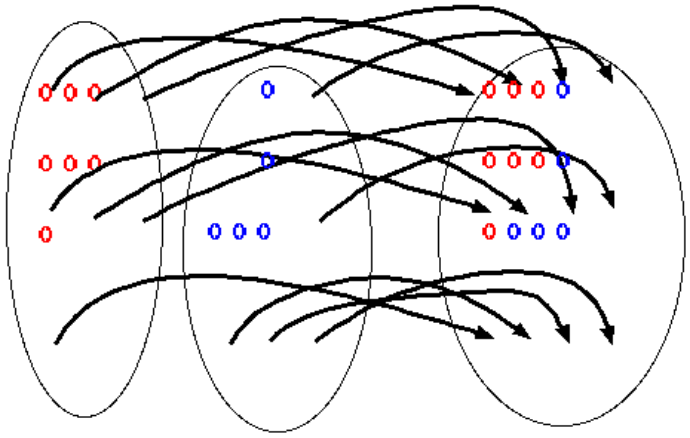
What sort of equivalence are we talking about?

I.e. what does "=" mean here?

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

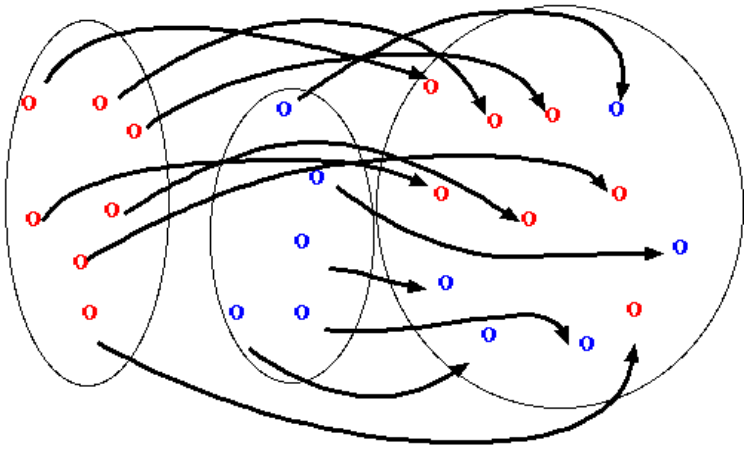
That can be defined logically, but the idea can also be understood by people who do not yet grasp the logical apparatus required to define the notion of a bijection — if they have a way of thinking about the consequences of motion of the blobs.

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 reasoning will also work for other additions, e.g. 777 + 555 = 1332

Humans seem to have a 'meta-level' capability that enables us to understand why the answer is 'yes'. This depends on having a model of how our model works – e.g. what changes and does not change if you add another pair of objects joined by a string.

But that's a topic for another occasion.

14 What the simulation theory does and does not say

So far I have given many examples, and talked very vaguely about perception and reasoning as involving various kinds of simulations, using different ontologies with different sorts of constraints, different viewpoints, etc.

But the theory is easily misunderstood – and also still has many gaps.

I'll now try to make it a little more precise, including saying what I am NOT claiming.

The concurrent simulation theory in more detail

- Different simulations of the same scene may be used in different sub-mechanisms running simulations at different levels of abstraction and serving different functions.
- Some parts of simulations may go beyond sensory data, e.g. including unobserved sub-mechanisms (Kant)
- Some of the processes are continuous some discrete.
- The continuous and discrete processes may both have different levels of resolution.
- There may be gaps in the simulation at all levels (for different reasons)
- Mode of processing can change dynamically: parts of the simulation may be selected for more detailed processing, or type of processing can be changed.
- Seeing static scenes involves running simulations in which nothing happens – though many things could happen (cf. seeing affordances).
- The mechanisms originally evolved to support perceptual and motor control processes but became detachable from that role in humans and can be used to think about things that could never be observed,

  e.g. search spaces, high-dimensional spaces, infinite sets, including operations on transfinite ordinals (move all the odd numbers after the even numbers and reverse their order).

  See my paper 'Diagrams in the mind' 1998
  http://www.cs.bham.ac.uk/research/cogaff/96-99.html#38

Development of perceptual sub-systems

The ability to run these simulations is not static, and may not even exist at birth:

- **Visual capabilities described here develop in part on the basis of developing architectures for concurrent simulations and in part on the basis of learning new types of simulation, with appropriate new ontologies and new forms of representation.**
- **The initial mechanisms that make all of this possible must be genetically determined (and there may be limitations caused by genetic defects).**
- **But the *contents* of the abilities acquired through various kinds of learning are heavily dependent on the environment – physical and social, and on the individual's history. Some innate content is needed for bootstrapping.**
- **For instance someone expert at chess or Go will see (slow-moving!) processes in those games that novices do not see.**
- **Expert judges of gymnastic or ice-skating performance will see details that others do not see.**
- **An expert bird-watcher will recognize a type of bird flying in the distance from the pattern of its motion without being able to see colouring and shape details normally used for identification.**

A deeper theory would explain the variety of types of changes involved in such developments: including changes in ontologies used, in forms of representation, and perhaps also in processing architectures.

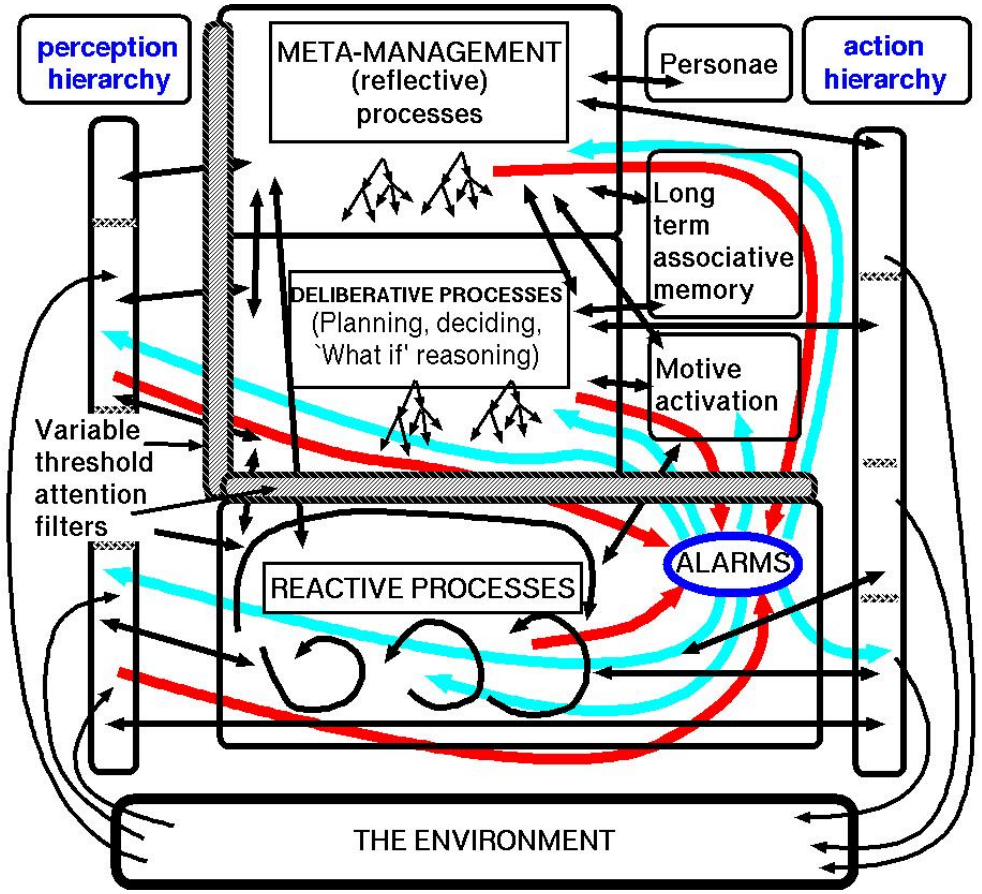
These will be changes in virtual machines implemented in physical brains.

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

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

Many required sub-systems are not shown.

Different kinds of process simulation may go on in different parts of the architecture – some very old and widely shared, some relatively new and found in very few species.



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

For more details, see the presentations on architectures here
http://www.cs.bham.ac.uk/research/cogaff/talks/

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'.)**

15 What I am NOT saying

The theory being proposed is easily misinterpreted.

The following slides attempt to explain what is not being said, by pointing out that some tempting interpretations of the theory are wrong.

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.

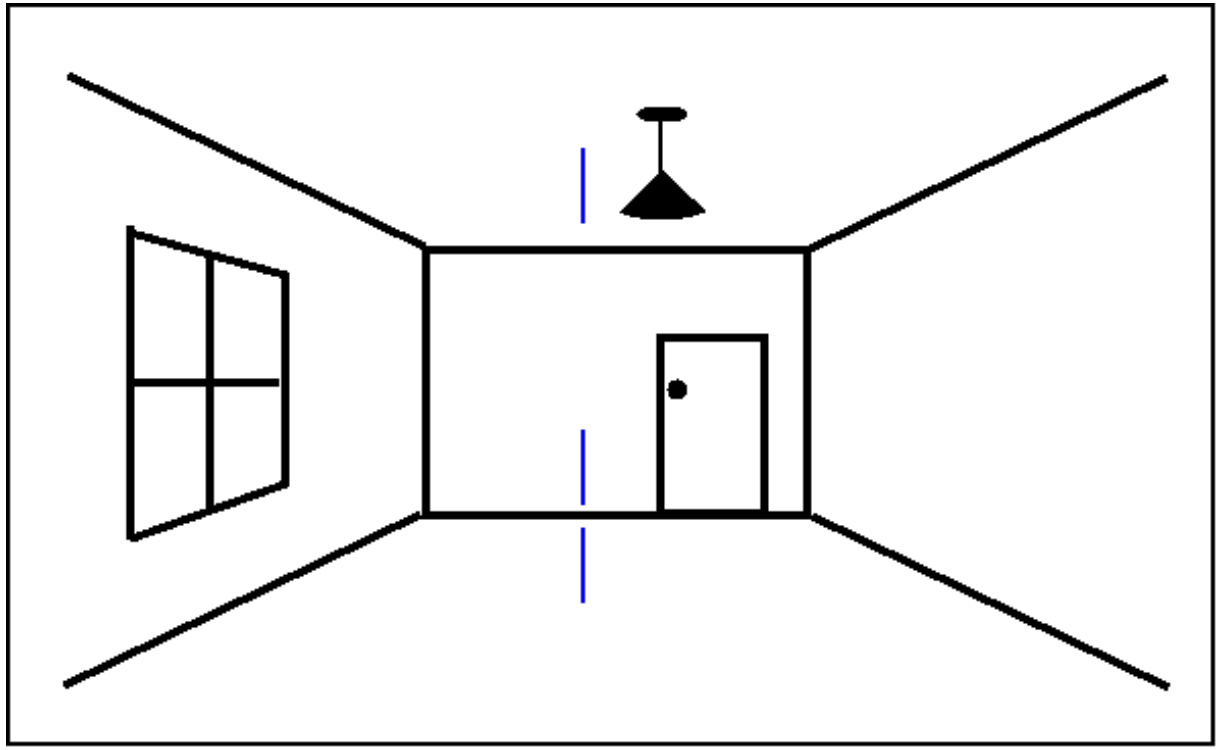
  There is considerable overlap with the BBS paper by R.Grush (2004): The Emulation Theory of Representation.

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.

MAJOR DISCLAIMER

I am not claiming that simulations have to be isomorphic with what they simulate

- As pointed out in my 1971 paper, analogical representations use relations to represent relations but they need not be the same relations:

  Think of a 2-D picture of a 3-D scene (the same 2-D relation 'above' can correspond to different 3-D relations in different parts of the picture – floor, far wall, ceiling).

  See http://www.cs.bham.ac.uk/research/cogaff/crp/chap7.html

- Not all simulations of spatial processes have to be spatial: it may often be simpler to use equations, for example, and psychological behavioural experiments may be wholly unable to determine which kind of implementation is used without having access to design information.

- Somehow we have developed enormously flexible ways of using mappings between one changing structure and another changing or static structure – it is a matter of learning what kinds of formalism with what kinds of constraints do and do not work for particular tasks.

  E.g. programming language constructs can map onto dynamic graphical displays.

- The ability I am talking about goes on being developed throughout life as we acquire more and more kinds of expertise.

- That means a complete theory will have to explain that acquisition process – and no finite theory will explain all past, present and future human competence.

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 what physicists understand.**

Terminology

- Some people distinguish simulation, emulation, imagery, etc.
- What I call a simulation is a representation of a process that can be used for a variety of purposes, e.g. recording, predicting, tracking, explaining, controlling.
- A simulation may itself be a process, or it may in some cases be a re-usable static trace of a process, e.g. an executable plan, even a plan with loops and conditionals – with a 'now' pointer.
- The same process may be simulated at different levels of abstraction:

  simulations run at a high level may be very much faster than what they represent.
- Different sorts of simulations are useful for different purposes.
- A child continually learns new sorts of simulations and new uses for old sorts.
- Some running simulations can change direction, can explore options.
- Some simulations are continuous, and some discrete, and some simulated processes are continuous and some discrete.

  A continuous simulation may represent a discrete process and *vice versa.*
  It is difficult for a continuous simulation do searching, e.g. in a space of possible explanations or possible plans: discretisation makes multi-step planning feasible.
- A simulation may change in complexity and structure as it runs (e.g. simulation of development of an embryo — unlike simulations that involve a fixed dimensional state vector).
- The things that change in a simulation need not be numerical variables.
- We probably don't yet know all the powerful ways of representing processes that evolution may have discovered and implemented in brains.
- In principle a simulation can itself be simulated (e.g. at a higher level of abstraction) – as in John Barnden's ATT-META system. http://www.cs.bham.ac.uk/ jab/ATT-Meta/

16 Re-runnable check-points

One of the consequences of discretisation is support for multi-step deliberation, e.g. systematic searching for a plan, including use of back-tracking.

Re-runnable check-points

- When searching for a solution to a problem we often have to explore a branching space of possibilities.
- Continuous simulations are not good tools for exploratory searching because there are always infinitely many possible branch points with infinitely many branches.
- This can be overcome by doing the searching with the aid of a discrete, more abstract, symbolic version of the simulation, and saving check-points, which can later be compared with one another.
- Ideally the check-points should be able to generate new lower-level runs of the simulation, when you back-track to a check-point.
- But for this, fully fledged deliberative mechanisms (for exploring answers to 'what if questions') could not really use simulations.
- So the development of discrete (symbolic) forms of representation was a major step for evolution. It had profound consequences including making mathematics and human language possible.

  Some animals probably use discrete symbols in internal languages.

  http://www.cs.bham.ac.uk/research/cogaff/81-95#43

Orthogonal environment-related competences 1

A typical child about five years old has much detailed knowledge of several distinct kinds, which can be combined in different ways in perceiving, understanding and planning actions in the environment:

- many kinds of physical stuff with different physical properties (e.g. water, sand, mud, wood, string, rope, paper, metal, stone, plastic, human skin, cotton wool, hair, butter, treacle, plastic film, aluminium foil, various kinds of food, wind, breath, fire and many more)
- different kinds of surface features – flat curved, smooth, rough, sticky, slimy, wet, sharp, textured in different ways, with ridges, furrows, dents, etc. etc.

  This decomposes further into yet more orthogonal sub-spaces.
- different shapes of whole objects, varying in topological and metrical aspects, with both continuous and discrete sub-spaces, at different levels of abstraction,

  E.g. there are discrete differences between numbers of holes, between being symmetric or not, having a long axis or not, etc. as well as a huge variety of types of continuous variation.
- different ways in which new, possibly more complex, wholes can be formed by combining or modifying things (in ways that depend on their shape, material, etc.)

  We could include 'negative' combinations, e.g. gouging out, carving, punching a hole, to make a new shape as in sculpture.

  Other shape-making transformations include bending, twisting, etc.

(continued...)

Orthogonal environment-related competences 2

.... Continued from previous page

- different sorts of spatial relations between different objects of similar or different material (e.g. containing, touching, being glued to, being hooked round, being a certain distance apart, resting on, being mixed, attracting, repelling, etc. etc.)

  There's a particularly important difference between 'rigid' containment (e.g. the streak of metal in a rock, the screw in a plank) and 'fluid' containment, e.g. water, sand or a small ball in a mug, a river flowing in its bed.

- different kinds of force that can be applied to things, e.g. prodding, poking, stroking, squeezing, twisting, pulling, pushing, screwing, patting, ...

- different sorts of process that can occur, including moving, rotating, changing shape, entering, coming out of, passing between, pushing, pulling, stretching, swaying, covering, uncovering, putting on (clothing), flocking, swarming, as well applying forces, and changing the application of forces .....

  Some of these may result from the individual's actions, some merely observed.

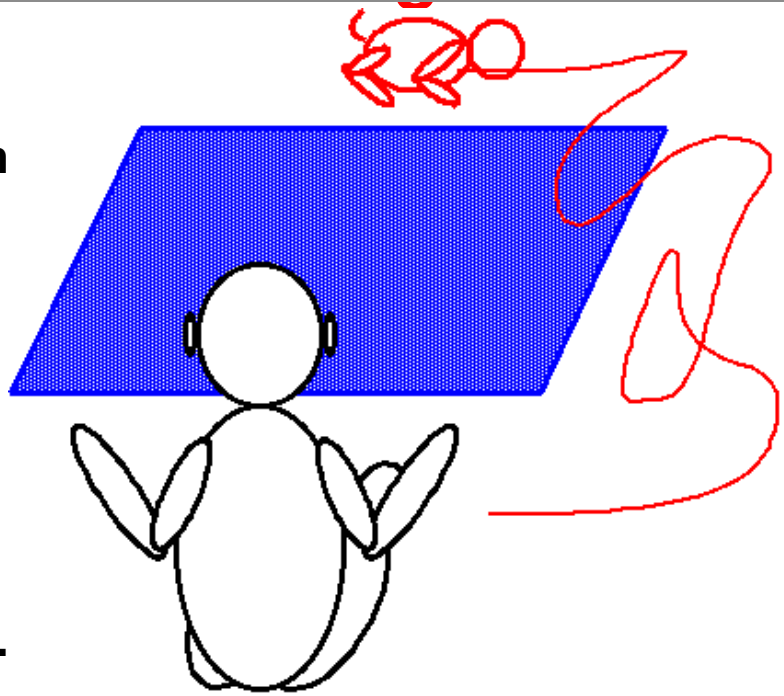
  As remarked previously, more complex things can be observed by an individual than produced by that individual, e.g. a busy street scene, a waterfall, a football match.

  (There may also be behaviours an animal (e.g. insect) can produce that it cannot perceive because its perceptual mechanisms lack the required sophistication.)

These lists are illustrative, not definitive or exhaustive, and do not include social abilities.

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.

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 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.**

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.

  Perhaps that is an example of the common pattern in evolution: duplication of structures or mechanisms followed by differentiation. (See the 'primacy' paper.)

Conjecture

Alongside the innate physical sucking reflex for obtaining milk to be digested, decomposed and used all over the body for growth, repair, and energy, there is, in some species, a genetically determined information-sucking reflex, which seeks out, sucks in, and decomposes information, which is later recombined in many ways, growing the information-processing architecture and many diverse recombinable competences.

Human-like robots will also need to be able to do that.

HOW ???

See also http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0601