Last Changed (January 26, 2008): liable to be updated.

Seeing Possibilities: A new view of Empty Space

Aaron Sloman

School of Computer Science, University of Birmingham

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

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

These PDF slides are in my 'talks' directory:

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

Video to warm up

Show video of 15 month old toddler manipulating a broom, while just about maintaining his balance – notices possibilities (e.g. corridor to go along, door to turn into at end) and manipulates the broom and his own location and posture in order to realise those possibilities.

He also detects impossibilities (motion of end of broom handle restricted by railing, forward motion of broom restricted by skirting board) and takes action to avoid the constraints.

That is included, along with some other videos assembled for my contribution to the CoSy Meeting of Minds workshop in Paris, sept 2006:

```
http:
//www.cs.bham.ac.uk/research/projects/cosy/conferences/mofm-paris-07/sloman/
```

I do not claim that he knows he is doing these things.

> At that stage of development he probably does not have a sufficiently well-developed reflective/meta-management layer in his information-processing architecture.

Similar questions arise about Betty, the New Caledonian crow, when she makes hooks

> (using several different techniques to achieve the same effect):
>
> does she know what she is doing and why it works, or does she merely do it?

Most current robots don't know what they are doing.

This talk is about seeing possibilities and impossibilities, and knowing what you are seeing, and making use of the information.

Look at some sliding possibilities

You can probably imagine the two objects in the picture sliding around on the surface of the page, rotating, and undergoing various transformations, e.g. stretching, compressing, shearing, and having their shape changed.
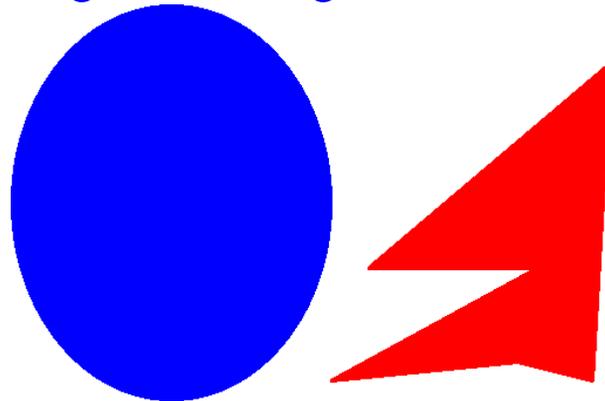
Suppose you imagine the red object moving left without changing its shape, orientation, or size: it just moves horizontally:

- which portion of the red object will first make contact with the circle?
- which point of the circle will be first touched by the red object?

You don't need to have very precise information about the objects in order to be able to give a rough answer.

Notice that simple geometric methods for finding the answer, like enclosing each in a rectangular box and finding where the boxes will first make contact cannot be used.

Neither can enclosing one of the objects in a box.

The ability to see and imagine changing structures

Thinking about where the shapes will make contact involves being able to represent a spatial process in which various spatial relationships change concurrently (e.g. relationships between various parts of the two objects in the figure): the key idea is that visualising that process when it is not occurring has something in common with what goes on when such a process actually occurs – about which we currently know very little.

If we can develop a theory of what goes on when we see things move, that will include identifying the forms of representation that are used

   possibly several forms used concurrently, for different purposes.

For various reasons, the internal representations used when a spatial process is perceived and understood cannot simply be another process inside the viewer of exactly the same kind – not least because we would then have the problem of explaining how that process is understood.

I suggest that explaining what is going on in such cases will require advances in computational theories of how perception works developed in the last 50 years or so.

High level plan

A short history of the aims of vision

- Pre-Marr

- Marr

- Gibson

- Generalised Gibsonianism (GG)

Theories about how to interpret modal operators

- Purely deductive uninterpretend axiom systems (several of them)

- Possible worlds semantics

And lots of examples

Pre-Marr – 1960s onwards

Lots of image analysis routines

e.g. Azriel Rosenfeld: but they mostly just transformed images

Ideas about pictures having structure, often inspired by Chomsky's work on language

E.g. S. Kaneff (ed) *Picture language machines*

Analysis by synthesis/Hierarchical synthesis

Ulric Neisser *Cognitive Psychology*, 1967 (parallel top-down and bottom up processing using models)
Oliver Selfridge, PANDEMONIUM (partly neurally inspired?)

Model-based vision research on polyhedra

Roberts, Grape, (particular models); Clowes, Huffman, (model fragments);
Also recent work by Ralph Martin's group (Cardiff). `http://ralph.cs.cf.ac.uk/Data/Sketch.html`

Use of expert systems ideas to analyse pictures

Hanson and Riseman (UMASS)

Structure from

Stereo (many people)
Motion (Ullman, Clocksin....)
Intensity (Horn, ...)

Towards understanding relations between image fragments and scene fragments

Barrow and Tennenbaum "Recovering intrinsic scene characteristics from images" (1978)

Parallel work on pattern recognition, from earliest times (often disparaged by AI people)

Marr – brilliant, but had some very bad effects

Some of his main points:

- Reject artificial images, use natural images – rich in data, making tasks easier (??)

  He ignored the possibility of informed selection of artificial images to study well-defined problems.

- Layers of processing: primal sketch, 2.5D sketch, 3-D interpretations

- Use of generalised cylinders (Compare Biederman's geons)

  Generalised cylinders proved unsuitable as models for many objects.

- Use of different frames of reference for scene descriptions
  - Viewer centred
  - Scene centred
  - Object centred

- Function of vision is to produce descriptions/representations of what's out there:

  3-D geometry, distance, surface orientations, colours, textures, relationships.

Stressed three levels of theory (causing much confusion):

computational, algorithmic, and implementational

They were badly named and far too widely accepted as important.

For a critique by McClamrock (1991) see `http://www.albany.edu/~ron/papers/marrlevl.html`

Marr's work killed off some research of a different sort

Popeye project at Sussex University (Chapter 9 of Sloman 1978).
Barrow and Tenenbaum's analysis?

J. J. Gibson's Revolution

The Ecological Approach to Visual Perception, 1979

For organisms the function of vision (more generally perception) is not to describe some objective external reality but to serve biological needs

- Providing information about positive and negative affordances (what the animal can and cannot do in a situation, given its body, motor capabilities, and possible goals)

- Use invariants in static and changing optic arrays

    texture gradients, optical flow patterns, contrast edges, "common fate".

- Use actions to probe the environment so as to change the contents of the optic array

    The sensors and effectors work together to form "perceptual systems" (compare "active vision")

Vision is highly viewer centred and action-centred.

There are no internal representations, no reasoning.

Perception is immediate and direct ("pickup", not "intepretation")

The idea that the input to vision processing is not retinal images but an optic array, whose changes are systematically coupled to various kinds of actions, was a brilliant move: the retina is just a device for sampling the optic array.

The problem with James Gibson

Although he continually emphasised **information**, e.g. the information available in static and changing optic arrays, he denied the need for representations or information processing (computation) using a mysterious concept of "direct pickup" instead.

He provided many important insights regarding interactions between vision and action, and the episodic information in vision, but ignored other roles for vision e.g.

- multi-step planning,

   If I move the pile of bricks to the right, and push that chair against the bookcase, and stand on it, I'll be able to reach the top shelf.

- seeking explanations

   Can marks on the road and features of the impact suggest why the car crashed into the lamp post?

- understanding causation

   (apart from *immediate* perception of causation, as in Michotte's experiments)
   If this string is pulled down, that pulley will rotate clockwise, causing that gear to turn, and ....

- geometric reasoning

   The line from any vertex of any triangle to the middle of the opposite side produces two smaller triangles of the same area, even when the shapes are different.

- Design of new machines, tools, functional artifacts (e.g. door-handles).

- Perceiving intentional actions. Fred is looking for something.

Contrast Eleanor J. Gibson and Anne D. Pick, *An Ecological Approach to Perceptual Learning and Development,* OUP, 2000.
They allow for the development of a wider range of cognitive competences using vision.

Beyond JJG: Generalised Gibsonianism (GG)

Less constrained analysis on the many functions of vision,

including its roles in mathematical reasoning (geometrical, topological, logical, algebraic), and its various roles in a robot capable of seeing and manipulating 3-d structures (as in the CoSy project)

leads to an extension of Gibson's theories,

while accepting his rejection of the naive view (e.g. Marr) that the function of vision is provide information about what objectively exists in the environment.

In particular we should not expect **one** set of functions to be common to all animals that use vision.

Many species use vision only for the **online** control of behaviour, using many of the features of the changing optic array, and correlations of those changes with actions, to provide information about what can be or should be done immediately (e.g. the need to decelerate to avoid hard impact, the need to swerve to avoid an obstacle, the possibility of reaching forward to grasp something).

In contrast humans (though not necessarily newborn infants) and possibly some other species use vision for other functions that go beyond Gibson's functions.

Moreover, in order to cope with novel structures, processes, goals and actions, some animals need vision to provide lower level information than affordances, information that is potentially sharable between different affordances: "proto-affordances".

Beyond affordances and invariants

- Vision does not just have one function, but many, and the functions are extendable through learning and development – building extensions to the architecture.

  E.g. reading text, music, logic, computer programs, seeing functional relations, understanding other minds, ...

- Vision deals with multiple ontologies

- Vision is not just about what's there but (as Gibson says) about what can happen

- But what can happen need not be caused by or relevant to the viewer's goals or actions

  Trees waving in the breeze, clouds moving in the sky, shadows moving on the ground, leaves blown by the wind.

- Besides action affordances there are also epistemic affordances concerned with availability of information.

- Besides affordances for the viewer some animals can see vicarious affordances, i.e. affordances for others

  including predators, prey, potential mates, infants who need protection, etc.

- Seeing structures, relationships, processes, and causal interactions (or fragments thereof) not relevant the goals, needs, actions, etc. of the viewer can make it possible to do novel things in future, by combining old competences.

  Great economies and power introduced by using an ontology that is exosomatic, amodal, viewer-neutral. (Still missing from current robots?)

GG: The ability to see possible changes

Seeing simple proto-affordances involves seeing what processes are and are not possible in a situation.

Seeing compound proto-affordances involves seeing what (serial or parallel) combinations of processes are possible and what constraints exist at different stages in those combinations.

In each of these four configurations, consider the question:

> Is it possible to slide the rod with a blue end from the "start" position to the "finish" position within the square, given that the grey portion is rigid and impenetrable?

Other questions you could ask include

> When it is possible how many significantly different ways of doing it are there?

(Based on a similar idea by Jeremy Wyatt)

This task and the earlier task (ellipse and polygon) use the ability to detect the possibility of movements that are not happening, and the constraints limiting such movements, and and to visualise combinations of such movements, while inspecting them for consequences: Using what brain mechanisms?
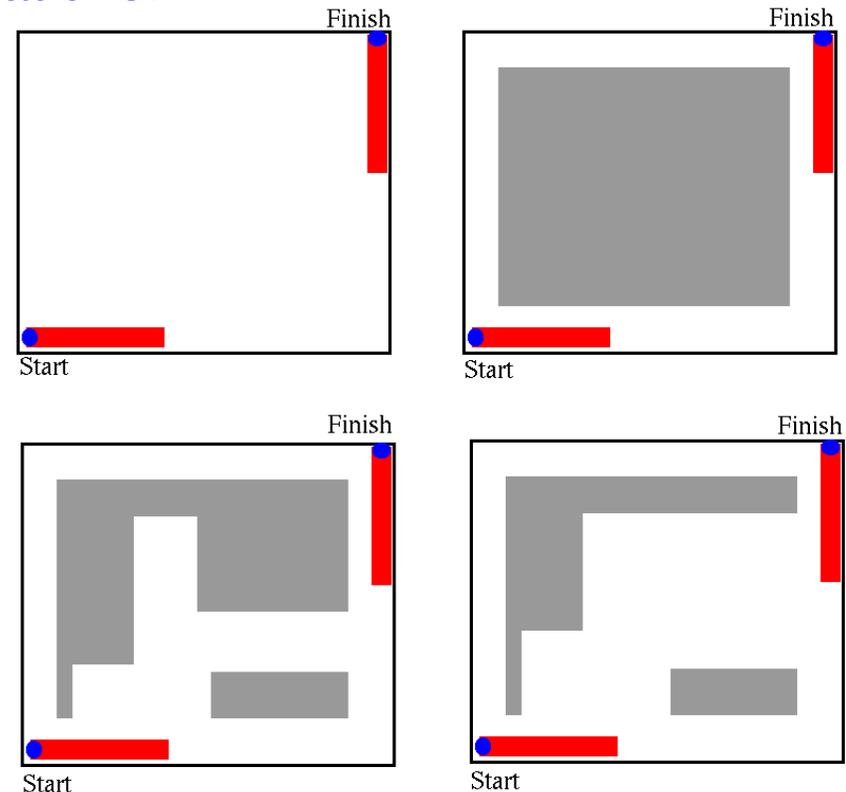
# Getting information about the world from the world

An action affordance concerns what can and cannot be done by the perceiver, whereas an epistemic affordance concerns what information is and is not available in the environment.

Actions can change both action affordances and epistemic affordances.

Things you probably know:

- You can get more information about the contents of a room from outside an open doorway

  (a) if you move closer to the doorway,

  (b) if you keep your distance but move sideways.

  Why do those procedures work? How do they differ?

- Why do perceived aspect-ratios of visible objects change as you change your viewpoint?

- In order to shut a door, why do you sometimes need to push it, sometimes to pull it?

- Why do you need a handle to pull the door shut, but not to push it shut?

- Why do you see different parts of an object as you move round it?

- When can you can avoid bumping into the left doorpost while going through a doorway by aiming further to the right – and what problem does that raise?
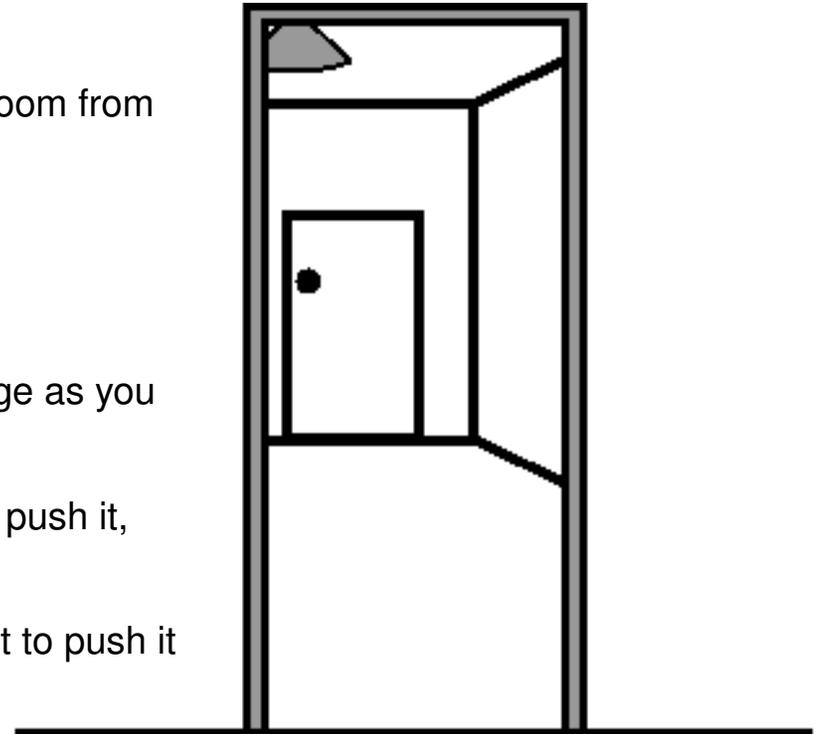
Carrying a chair through a door

Process fragments (proto-affordances) can be combined, in sequence or in parallel, in action or in hypothetical reasoning, to form new complex processes (actual or possible).

Affordances can interact in complex ways when combined, because of changing spatial relationships of the objects involved during the processes of performing the actions.

A large chair may afford lifting and carrying from one place to another, and a doorway may afford passage from one room to another, but attempts to combine the two affordances by lifting and carrying the chair to the next room may fail when the plan is tried.

A very young child may not be able to do anything about that, but an older child who has learnt to perceive the possibility of rotation of a 3-D object, may realise that a combination of small rotations about different axes combined with small translations some done in parallel, some in sequence. can form a compound process that results in the chair getting through the doorway.

Is any other type of animal is capable of understanding that?

Even the very familiar process of grasping an object is a complex combination made of various successive sub-processes, and some concurrent processes, with concurrently changing relationships between different parts of the surface of the object and different parts of the grasping hand.

Problem: What needs to be added to traditional AI planners to enable them to construct plans involving such continuous, concurrent, interacting processes?
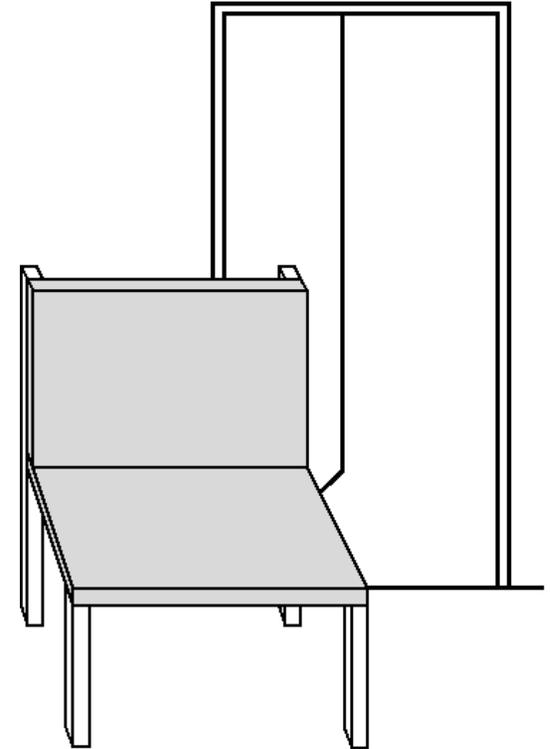
Interacting processes

Processes that occur close in space and time can interact causally in a wide variety of ways, depending on the precise spatial and temporal relationships and constraints.

It is possible to learn about the consequences of such interactions by observing them happen, but humans and some other animals sometimes need to be able to consider and work out consequences of possible combinations that they have never previously observed.

The ability to think about and reason about novel combinations of familiar types of processes is often required for solving new problems.

One source of fallibility of mathematical generalisations about interacting spatial structures is the fact that whatever space encloses those processes could, in principle also contain something else that interferes with their normal consequences.

Thus the necessity in such causation, and the validity of spatial mathematical reasoning is always conditional, but often we don't understand the conditions well enough to formulate them apart from the nearly vacuous *ceteris paribus* (other things being equal).

Perhaps slightly better: *provided nothing else intervenes*.

Don't forget Lakatos.

Why are those functions of vision?

Richard asked why the alleged list of functions of vision isn't just a list of parts of an architecture that should be investigated in parallel with vision, though it may use the results of visual processing.

Two answers

- The processing that is unique to those functions often needs to be done done in registration with visual representation (or partial registration where the processing is at a fairly high level of abstraction).
  E.g. when proof-reading text you want to know where the bits that need changing came from on the page.

- Secondly the notion that different modules can all be developed in isolation is a myth that
  - leads to harmful fragmentation in AI
  - produces subsystems that cannot be integrated (they may scale up well on their specialised benchmark tests, but they don't "scale out")
  - is one of the main themes of the CoSy project and the EU Cognitive systems initiative

Proto-affordances – objective representations

Proto-affordances: Things that can happen or change in the scene, independently of whether they are relevant to the perceiver's needs, goals, capabilities, etc. and the can be amodal and exosomatic.

These can involve fragments of objects, of surfaces

relations between fragments

can be 2-D or 3-D.

larger ones can be composed of smaller ones

learning to see them and to manipulate the possibilities involved is the basis of human mathematical competences – as well as planning, predicting, explaining competences.

requires architectures that support concurrent processing including both object level and meta level: noticing patterns.

requires ability to abstract from details, to represent something powerful and reusable.

What can be learnt by interrogating nature

Topics for further investigation (I have time only to illustrate):

- some of the ways nature can be interrogated, e.g.
  - perceiving
  - acting and perceiving
  - getting information from others who have already acquired information

- some of the kinds of information that can be acquired by such interrogation, e.g.
  - about what particular things and what types of things exist in the environment
  - about possibilities for change and limitations of possibilities
  - generalisations about what happens when
  - limitations and benefits of particular forms of representation
  - the need to modify or extend current ontologies (CRP, 1978 Chap 2)

- some of the things that can be done with the information, e.g.
  - achieving practical goals (changing the environment, including online control)
  - understanding causation and making correct predictions
  - explaining WHY things are as they are, in two ways:
    - o Deriving consequences from theories (about hypothesised mechanisms)
    - o Investigating limits of what is possible in a world for which a certain form of representation is appropriate (e.g. a certain sort of geometry, a certain kind of logic).

- Information-processing architectures, mechanisms, and forms of representation required for all this to work (Including architectures that grow themselves.)

The child as explorer

A child who plays with toys and various parts of its body, later also learns to play with information structures.

Besides learning to manipulate objects,

a child also learns many hundreds of ways of acquiring, manipulating and using information in the first years of life;

but understanding why they work, and what their limitations are, comes later.

Acquiring object-level and meta-level knowledge

Gilbert Ryle distinguished "knowing how" and "knowing that".

We can distinguish **object-level** practical knowledge and **meta-level** practical knowledge.

- Most animals, very young children, and current robots have only object-level practical knowledge, i.e. know-how (including knowing how, knowing that, knowing who, knowing when, ...)

- This may come from evolution (the only source for most animal species), or from training, e.g. learning associations between goals+circumstances and actions that will achieve the goal in a situation, or by building up records of what's where when.

  This is misleadingly labelled "episodic" memory and misleadingly contrasted with "semantic" memory.

- Such know-how is implicit in all feedback control mechanisms that achieve or maintain some state.

- But it is possible to have the object-level knowledge and lack meta-level knowledge: most animals don't know what they can and cannot do, under what conditions they can do it, why the right actions succeed, why the wrong actions fail, etc.

  Even human self-knowledge of this sort is always limited.

- Some mechanisms that provide the meta-level knowledge can also contribute to mathematical knowledge.

- Most current AI robotic research aims only at giving object-level know-how.

I'll return to meta-level knowledge later.

The need for patterns of motion, or change

Several of the examples have involved things changing in some experimental situation:

- A person moving nearer to a door and seeing more of a room
- A person moving past objects and seeing them in some order
- A person moving sideways and seeing different parts of a room
- A person moving past objects and seeing them in some order
- Cutting processes that increase the number of objects
- Counting processes
- Coin-turning processes
- Processes of re-ordering items

Perceiving a process clearly produces processes in the perceiver: things change in the perceiver.

If what is seen is remembered and re-usable, that implies that some information structure is stored which can be accessed later: a representation of the process (not necessarily representing every detail of the process).

Much is unknown about what forms of representations are good ones, what forms brains use, what forms should/could be used by robots – though there has been a lot of work on auditory memories suggesting that what is stored is itself a process, in some of those cases: rehearsal.

Re-usable information about processes

The ability to use a remembered process to produce or recognise a new process of the same type implies that there is some sort of pattern structure in the process representation: it can be re-instantiated to new instances of the pattern – perceived or created.

So our claim that patterns could be discovered in processes is not a very surprising claim – if the discovery of patterns is already a requirement for repetition or recognition.

However that leaves entirely unspecified what the form of that pattern is.

E.g. it could be an algorithm for generating instances.

If the pattern allows different forms, e.g. counting sequences of different lengths, the pattern may be stored in the form of a grammar of some kind.

Modal logics

Theories about how to interpret modal operators

Namely

- Necessary

- Possible

- Contigent

- Impossible

- Purely deductive uninterpretend axiom systems (several of them)

- Possible worlds semantics
  Accessibility relations

- Perhaps what we need is "possible fragements of this world" semantics.

Inspectable structures and processes

Some of what we have said about the difference between object-level knowledge or know-how, and meta-level knowledge, e.g. about discovering limitations of what is possible, or what must necessarily occur in certain conditions, depends on those process-representation patterns being inspectable.

We already have AI systems that can inspect some of their own data-structures and some of their own operations. (Cf. Sussman's HACKER 1975)

That's not all that different from what we need for logical information to be stored, and to be re-usable, and to be testable for validity or inconsistency.

Learning about occlusion and epistemic affordances

A child can learn various things about the effects of moving in such a way as to change what it sees: at first empirically, and later understanding why

Moving from side to side can provide evidence, in the form of optical flow, that one object partially occludes another.

(a)   (b)
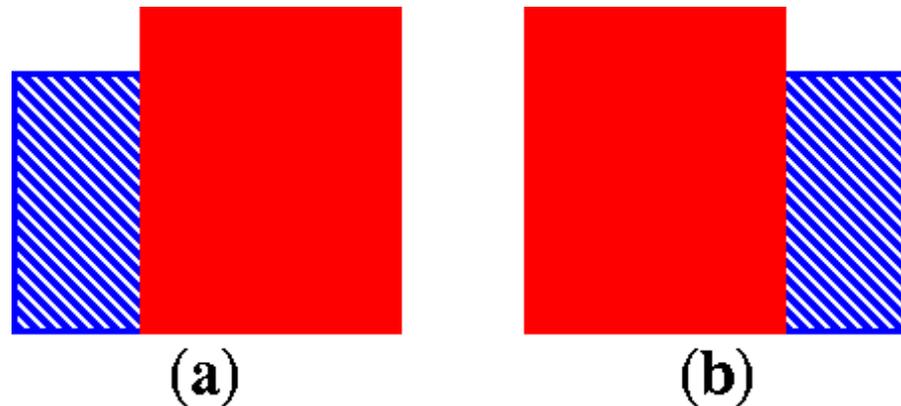
If you wish to see more of a partially occluded object you can do so by moving sideways in the direction in which the occluded object protrudes.

E.g. in situation (a) move left to see more of the blue object, in situation (b) move right.

A child could learn that the occluded object is further away than the occluding object, and "further" is transitive.

BUT....

How what is seen changes with motion

As you move from side to side, or rotate your head (or eyeballs) to look in different directions (including downwards or upwards) or move backwards or forwards, what you see changes systematically in many detailed ways, providing information both about what is in the environment and how you are moving, but also about what is and is not possible in the current situation, and about what information is and is not available.

The importance of this was emphasised by J.J.Gibson in *The Ecological Approach to Visual Perception*, and many examples relevant to child development are presented by E.J.Gibson and A.D.Pick in *An Ecological Approach to Perceptual Learning and Development*.

However the variety of types of information available in the environment is even richer than they suggested, and the ways in which the information can be represented, manipulated and used more diverse than they thought: there is a lot more than sensorimotor invariants.

E.g. J.J.Gibson focuses much on the use of perception for controlling action, but ignores the use of perception for finding explanations, or for designing new things.

However, E. Gibson and Pick do draw attention to a child's need for representation of future possibilities, e.g. alternative routes round an obstacle.

J.J. Gibson also did not address the difference between
   (a) being able to acquire and use information, and
   (b) understanding why things are as they are, including predicting and explaining novel effects.

Information can be used with, and without, understanding

Many animals, e.g. insects, use many of the sorts of mathematical facts discussed here, but they do not know that they use them, or why they are usable with confidence.

Robots can also be built that learn and use associations without understanding what they are doing or why it works.

That description fits all current robotics, as far as I know.

- The systematicity in the relationships between perceptual contents and changes and what is happening in the environment can be used through purely reactive highly trained associative mechanisms (e.g. neural nets).

- However it can also be used in a different way for reasoning, predicting, explaining, and solving novel problems creatively.

    This occurs in robots that can use planning mechanisms to find new routes.

- Robots can do such things without knowing what they are doing or why it works – would would require something closer to mathematical competence.

- The relationship between perceptual competence and mathematical competence is, I believe, closely related to Kant's philosophy of mathematics.

Seeing uses exquisite, and changing, structural correspondences between what is in the environment, where the viewer is, and how the viewer moves.

We have already given some examples, including the multiple changing projections of fixed 3-D shapes into the 2-D optic array.

As the Gibsons noted: passively observed changes are rich in information, and actively produced changes can provide even richer information.

They also generate philosophical problems, e.g. about qualia, the changing contents of perceptual mechanisms!

TO BE EXPANDED

We need to talk more about similarities and differences between vision and other forms of perception: are there auditory, or haptic, or olfactory inference patterns?

You can see something cause something else to happen: can you hear causation happening or smell it happening, with the same kind of necessity, or inevitability in the events concerned?

Seeing motion possibilities

For any configuration that happens to occupy some part of space, there are always variants that are possible: learning to see what changes are and are not possible is a crucially important aspect of learning to see – for an animal or robot that can act in the world.
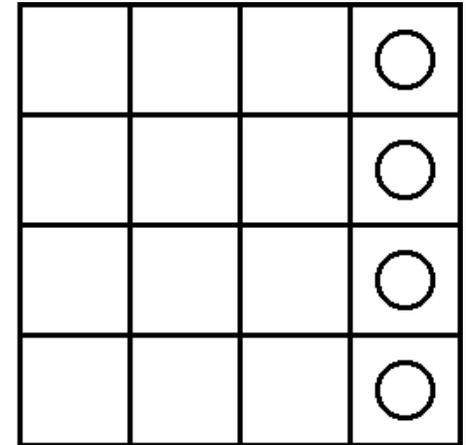
Some possibilities are obvious, others not so obvious.

For example if you have coins placed on a board divided into squares, it is obvious that you can slide them around into different places.

Suppose you consider only moves that are diagonal: no coin can go straight up or down or horizontally.
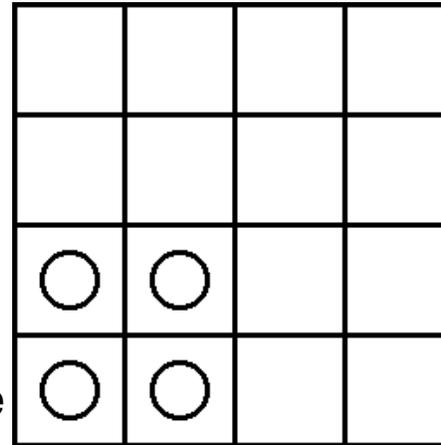
(a)   (b)

Question:

Using only diagonal moves, can you transform configuration (a) to configuration (b)?

What is the minimum number of diagonal moves?

Some people will find the answer obvious, whereas others will have to experiment.

You may find a some general pattern in the combinations of diagonal moves that convinces you that from any 2x2 starting configuration, the coins can always be converted to a row of four, or a column of four, using only diagonal moves.

What about converting them to a diagonal of four?

Seeing impossibilities

## What happens if we try a different pair of configurations?

Question:

Using diagonal moves, can you transform configuration (a) to configuration (b)?

This task may look easier than the previous one, because the starting and ending configurations look very similar: in both cases it is just a vertical column of coins.
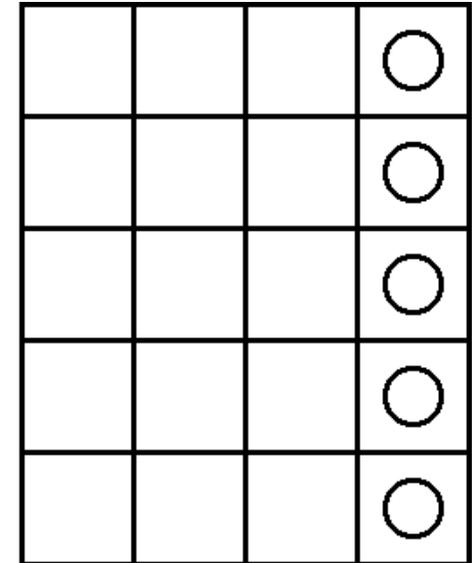
But if you try it you will eventually find that it is impossible.

Question:

How can you understand why it is impossible?
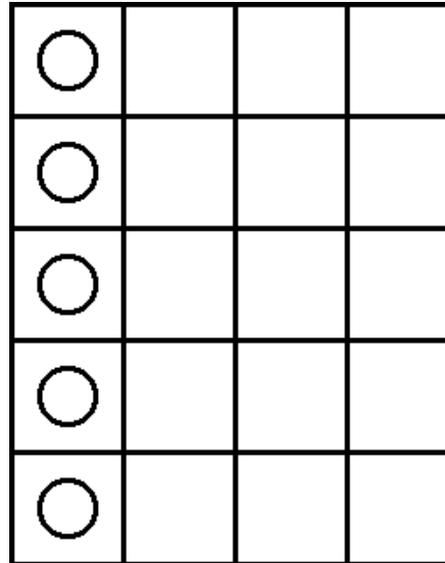
(a)   (b)

One thing you could do is try all possible collections of diagonal moves: which is not too difficult because the board is finite and there is only a finite number of configurations using that number of coins.

But an exhaustive analysis is very tedious: can you do better?

Mathematical intelligence essentially involves laziness, i.e. productive laziness.

In this case we want a way to see why the transformation is impossible, in a much simpler and cleaner way than by trying all possible moves.

One way to do this is to see some possibilities that are quite different from the possibility of moving coins around.

Discovering parity

I watched someone working on the five coin problem: at first she thought it was going to be easy. Then she tried, and found a problem. After trying a few different ways, she began to suspect it was impossible. After tracing routes she saw a pattern relating locations in the left column to relations in the right column. That pattern made it easy for her co conclude that the task was impossible. She had seen the link with chess boards.

Someone once discovered that a grid of squares has an interesting property: they can be divided into two colours in such a way that squares of the same colour are never adjacent: they meet only at corners.

That fact is used in chess boards.

On the basis of that clue you may be able to work out why it is impossible to perform the transformation from (a) on the previous slide to (b) on the previous slide, by moving coins only diagonally.

That requires you to notice a pattern that is never changed by such moves.

I don't know at what age young children are capable of discovering and using the information about diagonal colouring, but more importantly I don't know what has to change in their information-processing architectures to enable them to make the discoveries discussed here.

Perhaps we can come up with good hypotheses by trying to design robots that are capable of such discoveries.

Playing with the arithmetisation of geometry

Descarte's arithmetisation of geometry was one of the greatest and most important intellectual achievements in human history

Without it, Newton's mechanics would have been impossible

(Stephen Muggleton drew my attention to this.)

A child who had learnt about numbers might discover for the purposes of a game (e.g. the game "battleships") that it is convenient to label rows and columns of a rectangular grid with numbers: then each square in the grid can be identified using two numbers.

That invites various kinds of playing: e.g. what happens if you write into each box the sum of the two numbers that identify it?

Suddenly a deep link between colouring possibilities, diagonal moves and the difference between even and odd numbers becomes evident.

What enables a learner to realise that it does not matter how many squares there are, and that it works even if the grid has holes, like the odd slab of chocolate illustrated earlier.

You might suspect that if the colouring process went round a hole in the grid it might come round and be inconsistent with the starting layout.

Why is this impossible?

What happens if you write the **difference** of the two numbers, into each box?

Try the **product** of the two numbers: is there anything interesting to be found in the resulting pattern?
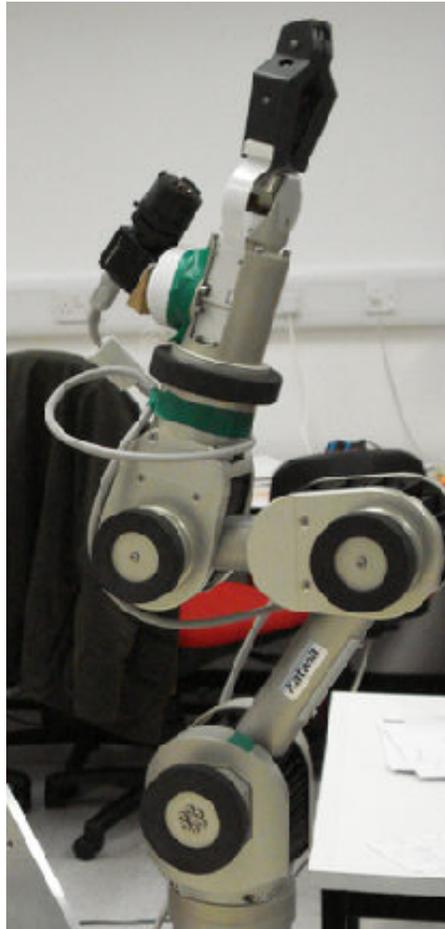
A pattern in the preceding examples

We can see some common patterns in the preceding examples, which may help us design more human-like machines, help us understand better how humans and some other animals work, and perhaps even help us design far better educational strategies.

The pattern seems to be something like this:

- Competences are acquired that allow actions to be performed on objects in the environment.

- Mechanisms involving those competences require use of representations of the structures and processes.

- Those representations of possible occurrences can to some extent be created and manipulated independently of what is actually actually going on in the environment.

- Consequences of the manipulations can be used for predicting or explaining actual occurrences, or for planning new ones to achieve goals.

- The forms of representation can themselves become objects of play and exploration, sometimes with the aid of externalisations (e.g. diagrams).

- This can allow the representations acquired for different competences to be combined playfully and the consequences explored (with or without external aids).

- A meta-management architectural layer observing things that happen in play and in use can notice and store patterns that have some interesting feature.

- Often those patterns allow new problems to be solved.

- Sometimes trying to solve a specific problem also leads to discovery of a new and powerful pattern.

The CoSy PlayMate robot

Our robot has a camera on its wrist, looking past the gripper

[SHOW TYPICAL PICTURES]

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

Epistemic affordances in grasping

The rigid relationship between eyes and mouth can be used to control motion towards an object to be grasped by biting.

The images represent two views as the eyes move down towards the object to be grasped by biting.
One of the images is taken when the gripper (i.e. mouth or beak) is still some way from the block to be grasped and the other is taken when the gripper is lower down, closer to the block.
Now, if the eye (or camera) is directly above the gripper is the gripper moving in the right direction?
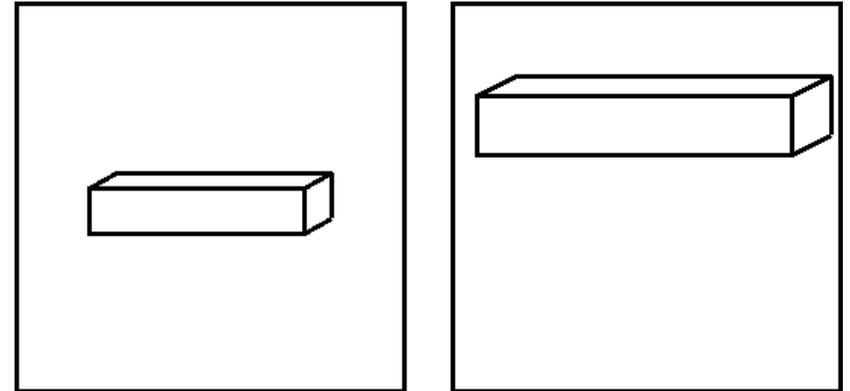
An agent can use the epistemic affordance here by reasoning about the effects of its movements on what it sees and how the effects depend on whether it is moving as intended or not (e.g. in this situation aim higher).

Instead of explicit reasoning using general knowledge about space and motion, an individual could simply be trained to predict how views should change if the target is being approached, and to constantly adjust its movements on the basis of failed predictions.

I.e. it can either use explicit knowledge, and the ability to reason about changing 3-D relationships applicable in varied situations, or implicit pattern-based reactive knowledge, produced by training, applicable only to situations that are closely related to the training situations.

Reactive pattern-based competence may work fast, but not generalise well.

Uncertainty-reducing affordances

A very common problem in robotics is the uncertainty that comes from low resolution or noisy sensory input, or inadequate algorithms for interpreting sensor input (as in current machine vision).

The diagram shows various possible configurations involving a pencil and a mug on its side, along with possible translations or rotations of the pencil indicated by arrows.

Assume all the pencils lie in the vertical plane through the axis of the mug.

For each starting point and possible translation or rotation of the pencil, consider questions like:

  o Will it enter the mug?
  o Will it hit the side of the mug?
 o Will it touch the rim of the mug?

In some cases the answer is clear.

In cases where the answer is uncertain, because the configuration is in the "phase boundary" between two classes of configurations that would have clear answers we can ask how the pencil could be moved or rotated into a new initial configuration, to make the answer clear.

If pencil A moves horizontally to the right, will it enter the mug? If the answer is not clear, what vertical change of location of the pencil will make the answer clear?

If pencil G is rotated in the vertical plane about its top end will it hit the mug? If the answer is not clear what translations will make the answer clear?

Perceiving a scene can include perceiving possible ways of changing the epistemic affordances related to actions under consideration.
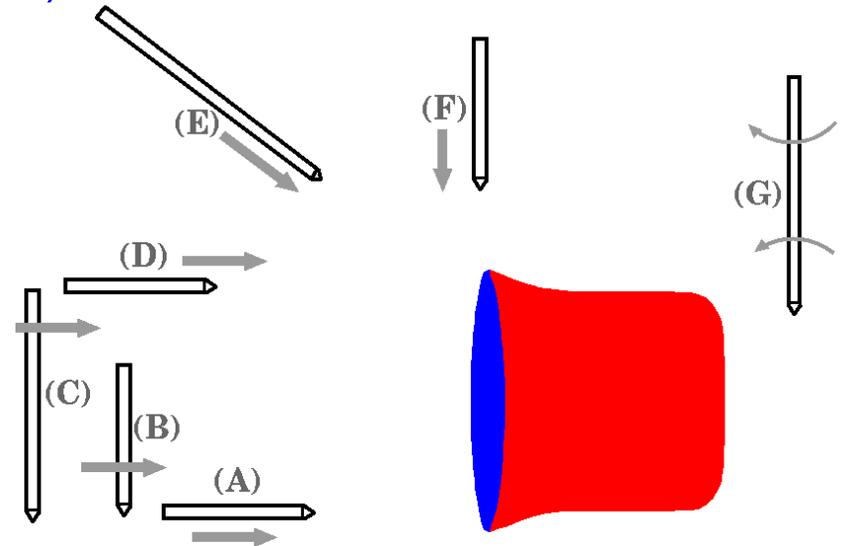
Two ways of dealing with uncertainty

Because uncertainty is so common in robotics, a vast amount of effort has gone in to ways of coping with uncertainty, including:

- Improving sensor quality

- Adding multiple sensors (e.g. multiple video cameras)

- Using different types of sensors (e.g. combining video cameras with laser range finders).

- Using sophisticated mathematics to compute probability distributions, and combining that with sophisticated decision-making algorithms to control actions.

A child or animal who is confronted with something uncertain, because of poor lighting, bad eyesight, dirty windows, occluding objects, distance of objects may not be able to adopt any of those engineering solutions.

(Except when it is possible to open the curtains or turn on a light.)

However they can learn other ways of coping with uncertainty, by using the epistemic affordances in the environment to remove or reduce uncertainty.

That typically involves changing what you are doing rather than changing the way you process information.

So alter your heading to remove uncertainty about a collision, look from a different viewpoint, or move an object, or rotate an object to remove uncertainty about things that are occluded or self-occluded.

Selecting an appropriate strategy can often use geometric or topological reasoning, rather than manipulating probability distributions and expected utilities.

Multimodal sensorimotor ontologies are not general enough

Full human competence in a 3-D environment requires more than a somatic ontology based on patterns in input and output signals.

For some purposes an exosomatic ontology (of 3-D surfaces, objects, substances, motions, causal interactions, etc.) is required.

For more on this see

http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0603
Sensorimotor vs objective contingencies

http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0601
Orthogonal Recombinable Competences Acquired by Altricial Species

http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0606
Requirements for going beyond sensorimotor contingencies to representing what's out there (Learning to see a set of moving lines as a rotating cube.)

Abstraction to an amodal exosomatic ontology

Five different cases of grasping occur here, with very different realisations (projections) in the image plane: what is common to the different cases can be abstracted to a 3-D relationship between two facing surfaces and an object between them.

Instead of a somatic sensori-motor ontology referring to contents of sensory and motor signals, this uses an exosomatic amodal ontology referring only to things in the environment, i.e. outside the body.

Using the exosomatic ontology makes it possible to predict, control, and understand motions in far more varied situations, e.g. using the fact that if two surfaces come together firmly with an object between them, then when the two surfaces move the other object will move with them.

Using exosomatic ontologies requires an architecture that supports hypothetical reasoning about 3-D geometric relationships, and causal consequences of changes.

Chains of causation

You can probably imagine various chains of causation by doing "What if reasoning about" this 3-D structure, with the initial causation being located in different places, including rotating, sliding, in various directions, etc.
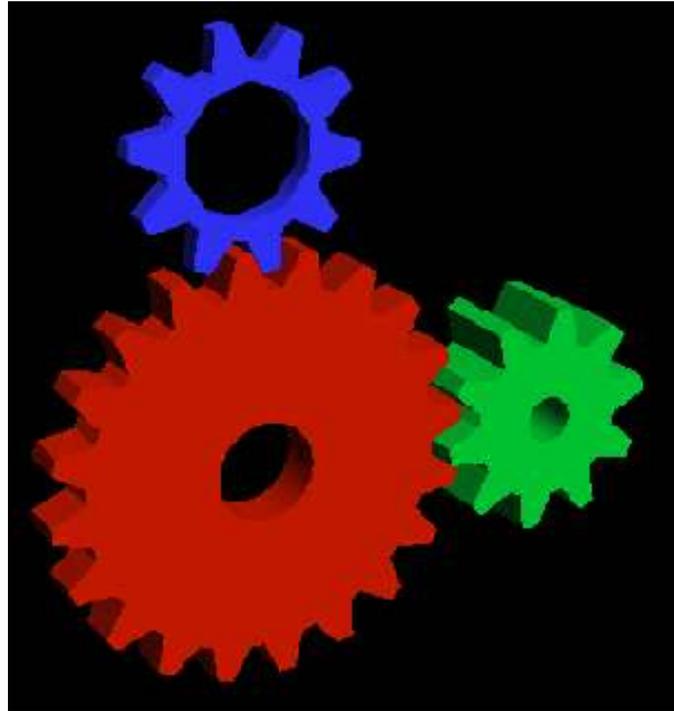
Snapshot of the 'glxgears' program running on linux

E.g. if the small blue wheel moves towards you then starts rotating it may leave the other two unaffected, but not if it rotates where it is.

Our ability to represent different combinations of processes that are possible in any situation is rich and varied – but limited, and sometimes partly dependent on previous practice.

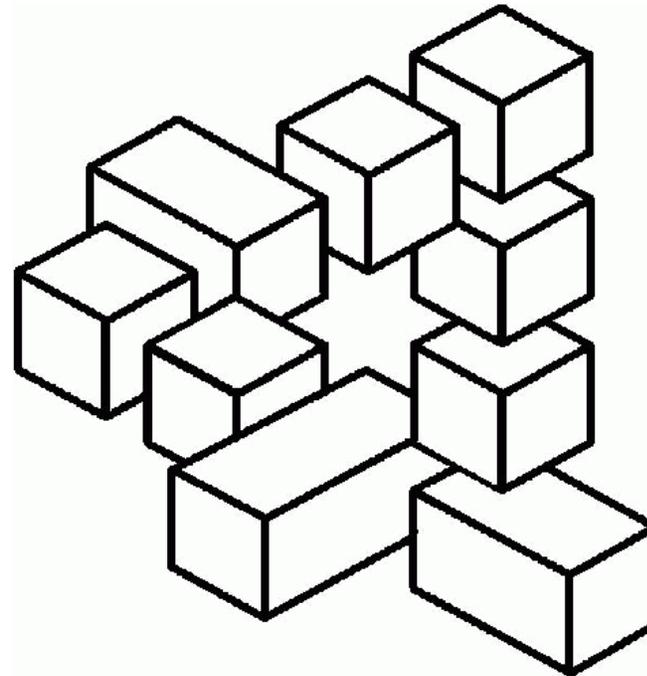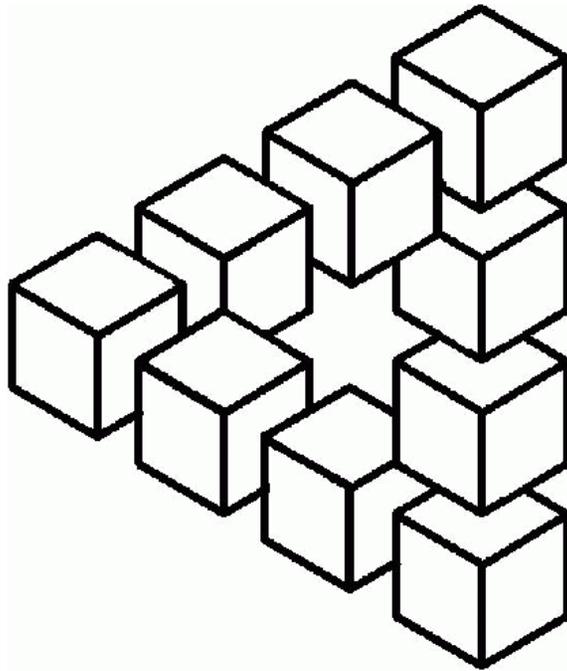# How we represent 3-D structures and processes is problematic

We can see a 3-D configuration of cu
LIKE THIS:

You could build a model of
what you see.
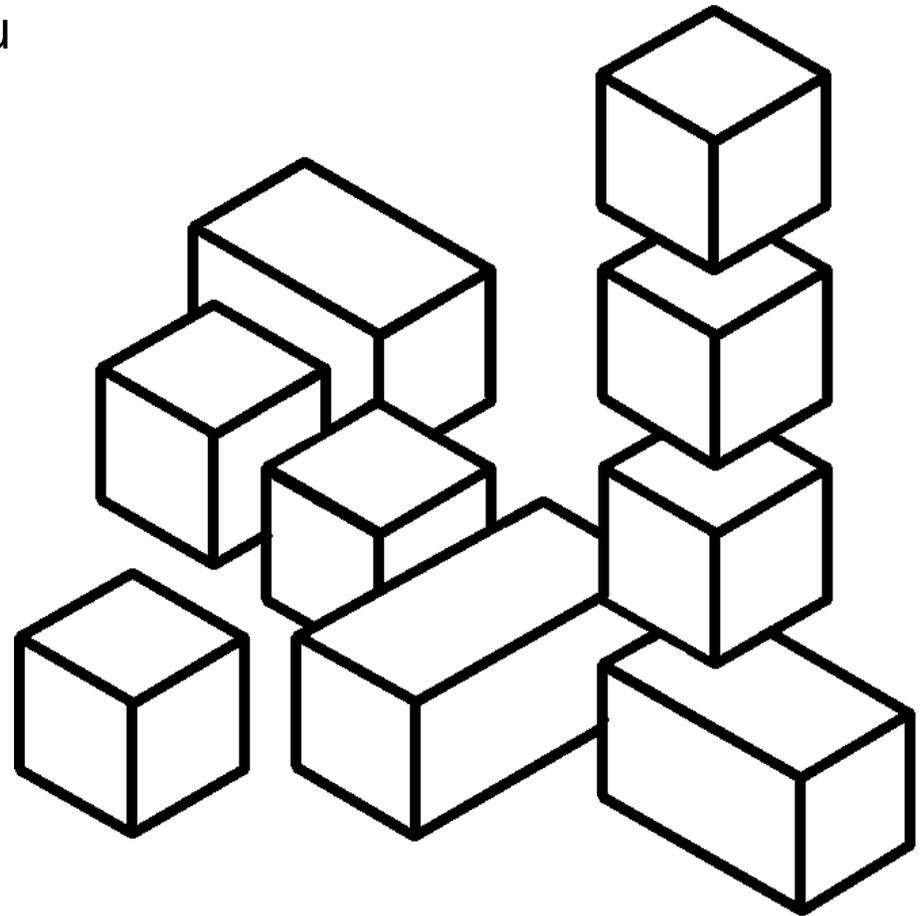
But what happens if we rearrange the cubes, ... ?

Like this?   or   Like this?

Given a pile of cubes, could you build configurations like these?

These examples – inspired by Oscar Reutersvärd (1934)[*] – show that 3-D perception does not involve building internal objects that are isomorphic with the things seen.

That's not at all surprising from the standpoint of Sloman 1971, which argued that analogical representations need not be isomorphic with what they represent. (See next slide)
http://www.cs.bham.ac.uk/research/projects/cogaff/04.html#200407

[*] http://www.sandlotscience.com/EyeonIllusions/Reutersvard.htm

Potentially inconsistent fragments

The crucial point is that the result of making sense of perceptual input is neither some sort of sensory copy of the stimulation pattern (e.g. a bitmap), nor an isomorphic model of what is taken to exist in the environment, but a collection of re-usable separate items of information about things, surfaces, processes, relationships, and possibilities in the environment derived from the sensory input (often using prior knowledge).

- In the 1960s it was thought by some that a major result of perception would be some sort of "parse tree" or "parse graph" for images and scenes, based on a grammar for spatial structures.   E.g. S. Kaneff (ed) *Picture language machines* 1970.

- But the proposed grammars were often very arbitrary, and the approach did not lead to good ways of representing information about 3-D structures and processes for a robot to use.

- An aspect graph, linking hypothesised views to actions that will lead to those views being seen has more 3-D information, implicitly, and allows partial ignorance to be represented.

- These ideas need to be generalised to allow more kinds of actions and more kinds of consequences to be represented, in addition to structural information.

- The form of representation needs to be capable of being driven by sensory input, and also by hypothesised future actions, e.g. in planning or predicting.

- Results of such manipulations need to be accessible for reasoning.

Converting mappings

A child may discover empirically a strategy for converting one mapping to another and implicitly understand that it will always work, without necessarily being able to articulate the strategy nor explain why it works.

This depends on the architecture allowing one process to observe that another process has some consequences that do not depend on the particularities of the example.

A one-to-one mapping from one set of objects to another (e.g. the grey arrows) can be converted to any other such one-to-one mapping (e.g. the black arrows) by swapping ends on one side, two at a time.

E.g. the right hand ends of the grey arrow from A to G and the grey arrow from D to E can be swapped, then the right hand ends of arrows from B to H and from C to F, etc. gradually eliminating differences between grey and black mappings.


Discovering that any one-to-one mapping between elements of two finite sets can be converted into any other by successive changes can make use of simultaneous perception of spatial and temporal relationships.

Formulating the general algorithm is left as an exercise. Could a robot do this?

I don't think this is how children normally come to understand the invariance:
What alternatives are there?
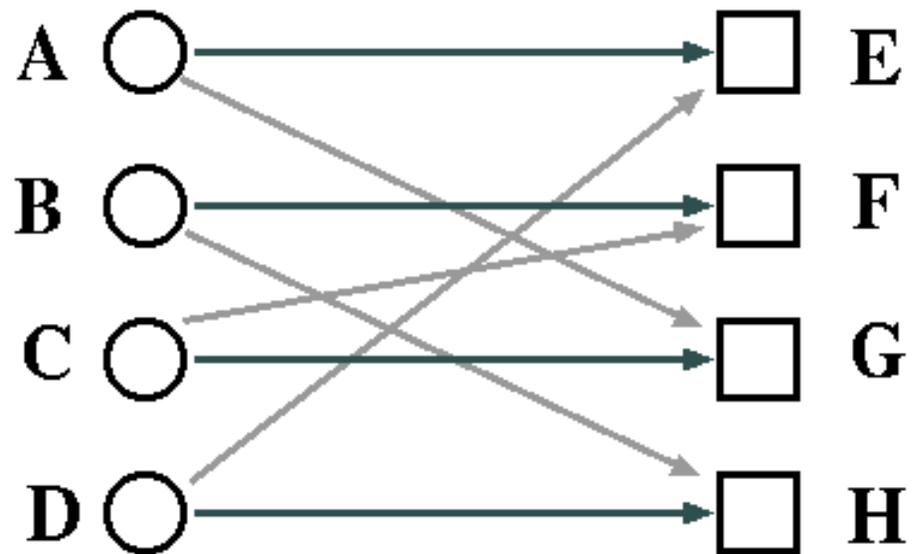
Pre-mathematical discoveries
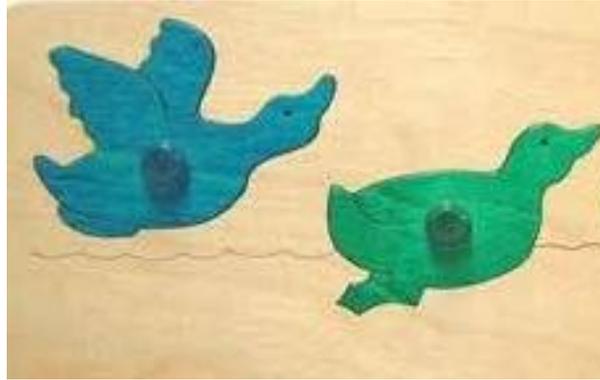
There are developments that would not normally be described as mathematical, yet are closely related to mathematical competences.

For example a very young child who can easily insert one plastic cup into another (of the sort shown in the figure may be able to lift a cut-out picture from its recess, and know which recess it belongs to, but be *unable* to get it back into the recess: the picture is placed in roughly the right location and pressed hard, but that is not enough.

The child apparently has not yet extended his or her ontology to include boundaries of objects and alignment of boundaries.

Some time later the child copes easily.

How such extension of competences happens is not at all clear, but what has to be learnt, namely facts about boundaries and how they constrain possible movements, is something that can be studied mathematically, and might be so studied later.

Specialised mathematical education builds on general abilities to see structures and processes and see how some structures can constrain or facilitate certain processes, including processes of information acquisition.

High level perceptual processes can ignore low-level details

- I am suggesting that when we watch or imagine things moving we simulate the motion (i.e. we create and run representations) at different levels of abstraction.

- Some of them we probably never become conscious of as they are used only in relatively automatic control of common processes, for instance as optical flow patterns are used in posture control.

- What we say we are conscious of is often closely related to what we can report, to ourselves or to others, and that will typically be things happening at a high level of abstraction, that are relevant to our current goals and needs, though we can direct our attention to details just for the sake of examining details, and we can become aware of details that are too rich and complex to be reported, even to ourselves, e.g. watching swirling rapids in a fast flowing river or hundreds of leaves stirring in the wind.

- What we are conscious of seeing may depend on what the current task is, and sometimes we do not notice details even if a low level system processes them – e.g. because what we attend to when answering a question includes only the contents of the more abstract simulations.

- But that does not mean that the details have not been processed, as I have shown elsewhere: one of your subsystems concerned with posture-control may be conscious of optical flow even when you are not.

Development of perceptual sub-systems

The ability to run simulations while seeing 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.

Things to do

There is still much to do, and many topics to discuss, including:

- The variety of extrapolations to limiting cases, e.g. infinite discrete sequences, infinitely long lines, infinitely large areas, infinitely thin lines, infinitely small points, infinitely dense textures,...

- Many issues to do with continuity

- Extending the notion of number from discrete, countable, sets to amount of something that can vary continuously, eg. length.

- How can a child come to understand the notion of half the area or volume of an asymmetric spatial region or volume.

- How to extend the idea of number to a measure of an arbitrarily shaped area: the importance of rectangular grids and the limiting case as grid size shrinks.

- Using finite spatial structures to represent infinite sets and infinite ordinals.

- Using what you can imagine to help you imagine what you can't imagine.

- What to think about Euclid's parallel axiom: is there some way of constructing a pair of straight lines that forces them to go on indefinitely exactly the same distance apart?

- Does the construction come unstuck before grids of lines with different orientations are considered?

- Need to go back to the elastic sheet proof of Euler's theorem: what would a robot need to be able to imagine the process of stretching a polyhedron's surface flat?

Unanswered questions

The form of representation, the mechanisms for manipulation, and the architecture for combining the various information-processing components of an intelligent individual are still barely understood.

A brave attempt at theory construction can be found in

Arnold Trehub, 1991, *The Cognitive Brain,*
`http://www.people.umass.edu/trehub/.`

The retinoid theory seems to be only a partial model, though richer than many others.

The work of Eric Baum may also be relevant, and his approach (looking closely at how humans solve particular problems) overlaps with what I have been doing.

Eric Baum's web site `http://www.whatisthought.com/eric.html`

"A Working Hypothesis for General Intelligence" 16 pages Draft October, 2006.
`http://www.whatisthought.com/working.pdf`

There is probably a lot of other relevant work that I don't know about or have forgotten (and may be unwittingly reproducing!).

We may be able to move towards a design specification if we study and analyse more and more examples in order to work out detailed information-processing requirements, which may lead us to features that may suffice to explain the desired behaviours.

In conclusion

- I have tried to identify an array of features of normal perception, action, learning, reasoning, control, planning, and explaining which seem to be products of complex developmental processes influenced by both evolution and the current environment (including other humans), and which also are able to play a role in generating mathematical explorations and supporting mathematical reasoning.

- Making this more precise and detailed will require considerably extending the state of the art in robotics and AI, and giving robots new ways of representing and reasoning about spatial structures and processes, as well as giving them architectures that support self-observation of a kind that drive new learning and developments.

- These mechanisms required for intelligently coping with the environment, including other intelligent individuals, can, as some science fiction writers have pointed out, produce both philosophical activities, and when they become really buggy, even theological activities.

    (Isaac Asimov: "Reason" in *I Robot*)