Last Changed (June 10, 2015) – liable to be updated.

What's vision for, and how does it work?

From Marr (and earlier) to Gibson and Beyond

(With some potted, rearranged, history)

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

These PDF slides are in my 'talks' directory:
http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#gibson
And also in slideshare.net http://www.slideshare.net/asloman

A partial list of requirements for visual systems in intelligent animals and machines can be found here (work in progress): http://www.cs.bham.ac.uk/research/projects/cogaff/misc/vision

An online discussion of Gibson by philosophers may (or may not) be useful to some:
http://theboundsofcognition.blogspot.com/2011/01/noe-on-gibson-on-affordances.html

The problem

- Very many researchers assume that it is obvious what vision (e.g. in humans) is for, i.e. what functions it has, leaving only the problem of explaining how those functions are fulfilled.
- So they postulate mechanisms and try to show how those mechanisms can produce the required effects, and also, in some cases, try to show that those postulated mechanisms exist in humans and other animals and perform the postulated functions.
- The main point of this presentation is that it is far from obvious what vision is for – and J.J. Gibson's main achievement is drawing attention to some of the functions that other researchers had ignored.
- I'll present some of the other work, show how Gibson extends and improves it, and then point out much more there is to the functions of vision and other forms of perception than even Gibson had noticed.
- **In particular, much vision research, unlike Gibson, ignores vision's function in on-line control and perception of continuous processes; and nearly all, including Gibson's work, ignores meta-cognitive perception, and perception of possibilities and constraints on possibilities and the associated role of vision in reasoning.**

High level outline 1

A short, incomplete, illustrative, history of theories of functions and mechanisms of vision

- Previous millennia, E.g. Aristotle
- Previous centuries: lots of philosophers, including Hume, Berkeley, Kant, Russell, and non-philosophers, e.g. (non-philosopher? polymath?) von Helmholtz.

  Helmholtz proposes a "sign" theory, according to which sensations symbolize their stimuli, but are not direct copies of those stimuli. While Müller explains the correspondence between sensation and object by means of an innate configuration of sense nerves, Helmholtz argues that we construct that correspondence by means of a series of learned, "unconscious inferences."
  `http://plato.stanford.edu/entries/hermann-helmholtz`

- AI Vision work – 1960s on

  Lots of work on finding and representing structure in images (e.g. lines, junctions, and also higher order, possibly non-continuous structures, e.g. views of occluded objects). (E.g. A. Guzman)

  Also work on trying to find 3-D interpretations, using 3-D models (e.g. Roberts, Grape), or using constraint propagation from image details, e.g. Huffman, Clowes, Winston, Waltz, ...)

  Then Barrow and Tenenbaum, on "intrinsic images" (Barrow & Tenenbaum, 1978) (1978), etc. (an idea that seems to have become popular recently)

  Use of soft constraints/relaxation, G. Hinton, use of statistics and neural net inspired mechanisms.

  Many ideological battles between researchers whose systems were **all** very limited, compared with animal vision.

- Nearly all researchers assumed that the functions of vision were obvious and only mechanisms/explanations were in dispute. Gibson shattered that.

High level outline 2: pre-Marr – 1960s onwards

Lots of image analysis routines

e.g. Azriel Rosenfeld: Many of the algorithms just transformed images

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

E.g. S. Kaneff (ed) *Picture language machines* (1970)

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 polyhedral models);
Clowes, Huffman, (model fragments: edges, faces, vertices);
Later work - Marr/Nishihara generalised cylinders; Biederman geons
More recent, more systematic work on polyhedra by Ralph Martin's group (Cardiff).
`http://ralph.cs.cf.ac.uk/Data/Sketch.html`

Use of "expert systems" techniques to analyse pictures Hanson and Riseman (UMASS)

Structure from:

Stereo (many people – badly influenced by results from random dot stereograms (Julesz)
Motion (Longuet-Higgins, Clocksin, Ullman, Spacek ....)
Intensity/shading (Horn, ...)
Texture and optical flow (Gibson)

More general 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)

High level outline 3: Marr 1

David Marr: papers at MIT in late 1970s, died 1981, posthumous book 1982

Died tragically young.

Enormously influential, partly for bad reasons, though very clever.

Claimed previous AI vision research was all wrong because it used artificial images (e.g. line drawings full of ambiguities) and vision doesn't require sophisticated information processing architectures, just a one-way pipeline of processing stages ending with model construction (using generalised cylinders).

Convinced many people that vision is (merely?):
a process of producing 3-D descriptions of shape, size, distance, orientation, etc, from 2-D data
"reversing" the production of an image by projection.

Went against his own anti-top-down approach by trying to fit scene structures to generalised cylinders.

Stressed three levels of theory (causing much confusion):

(1) Computational, (2) Algorithmic, (3) Implementational.
In fact, this is mainly a confused way of introducing the old engineering distinctions:

(1) Requirements analysis (What's X for?)

(2) Design (What are the high level design features for systems that meet the requirements?)

(3) Implementation (How can the design be implemented in lower-level mechanisms, physical, electronic, physiological, or computational.)

All of those can have different levels – e.g. implementation in virtual machines, implemented in lower level virtual machines, ... implemented in physical machines.

Marr's Levels were badly named and far too widely accepted as important, by people without engineering experience.

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

Marr 2

Some of his main points:

- Reject artificial images – e.g. line drawings used as test images.
  – use natural images (actual photographs of real 3-D objects)
  – rich in data, making tasks easier (??) [Illustrated using teddy bear photo]
  He discounted the possibility of informed selection of artificial images to study well-defined problems.
- Processing pipeline: primal sketch $\Rightarrow$ 2.5D sketch $\Rightarrow$ 3-D interpretations
- Use of generalised cylinders (Compare Biederman's geons)
  Generalised cylinders proved unsuitable as models for many objects.
- The function of vision is to produce descriptions/representations of what's out there:
  3-D geometry, distance, surface orientations, curvature, textures, spatial relationships, colours(?).
- He shared the common assumption that metrical (Euclidean) coordinate frames are required.

  But he allowed that different frames of reference can be used for scene descriptions
  - Scene centred (Use a global coordinate system for everything visible in the scene)
  - Object centred (Attach coordinate systems to objects, or object parts)
  - Viewer centred
    o egocentric (Represent scene objects in terms of relationships with perceiver).
    o allocentric (Represent scene objects in terms of relationships with another viewer).
    (I am not sure Marr made this distinction. Others have.)

Marr 3: functions of vision

"... the quintessential fact of human vision – that it tells about shape and space and spatial arrangement".

Comments:

(a) This ignores the functions of vision in on-line control of behaviours:

e.g. visual servoing while grasping, moving, avoiding obstacles –
or assumes (wrongly) that these control functions can all be subsumed under the descriptive functions.
Compare: (Gibson, 1966) (Sloman, 1983)

(b) Marr's view fits the common idea of vision as "reversing" the projection process – using information in the optic array (or image) to construct a 3-D model of what's in the scene.

But that common idea is mistaken: visual systems do not represent information about 3-D structure in a 3-D model (information structure isomorphic with things represented) but in a collection of information fragments, all giving partial information.

A model cannot be inconsistent.

A collection of partial descriptions can.

See pictures of impossible 3-D objects by Reutersvard (in later slides), Penrose and Escher.

The way we see such pictures would be impossible if certain views of 3-D vision were correct (percepts as 3-D models isomorphic with the objects perceived: are impossible for percepts of impossible objects – the models, and therefore the percepts would be impossible too).

NOTE:
Marr's emphasis on spatial structure made it awkward for him to talk about colour vision.

Marr 4: A hint of a move towards Gibson's ideas.

On p.31 of **Vision**, he wrote "Vision is a process that produces from images of the external world a description that is useful to the viewer

(That's Gibsonian)

and not cluttered with irrelevant information."

Not cluttered? My visual contents often are!

Also:

p. 32 "Vision is used in such a bewildering variety of ways that the visual systems of animals must differ significantly from one another".

"For a fly the information obtained is mainly subjective..."

(Mainly concerned with image contents?? Or something like affordances for the fly??)

Moving beyond Marr and predecessors

Marr:

"... the quintessential fact of human vision – that it tells about shape and space and spatial arrangement".

What else could there be, besides shape and space and spatial arrangement?

Lots!

Gibson noted some of it.

(Starting in the 1960s (Gibson, 1966).)

There is a useful but brief discussion of Gibson's ideas and how they relate to AI in section 7.v of (Boden, 2006), pp. 465–472.

Philosophers who are ignorant of that work by Gibson think the ideas came up much later, when philosophers and others started discussing embodied cognition, around 20 years after Gibson's first book.

Some of the ideas were also in (Simon, 1969).

Unfortunately, some philosophers teach their students to think of a philosophy as a subject that can be pursued in ignorance of science because philosophy studies "purely conceptual" problems, or something like that.

Moving Beyond Marr and predecessors 2

Are visual contents spatial?

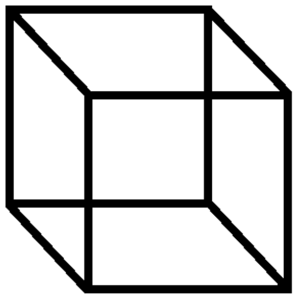
Yes – but only **some** of them.

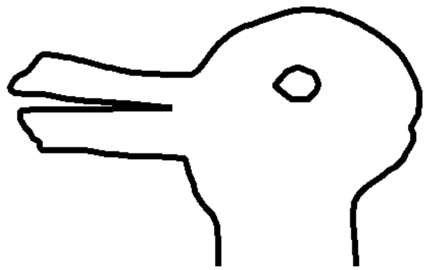
Contrasting two sorts of ambiguous image makes this clear:

Ask what changes in what you see:

In some cases it is shape, and space and spatial arrangement, as Marr claimed.

But not in all cases (Sloman, 2001):


Necker Cube


Duck-rabbit

Everything that changes in the "Necker flip" is spatial: orderings, directions, relative distances, orientations.

In the duck-rabbit, however, there is no geometric flip: the change is much more abstract and involves both changes in how parts are identified (e.g. ears *vs* bill) and more abstract notions like "facing this way", "facing that way" which presupposes that *other organisms can be seen as perceivers and potential movers*.

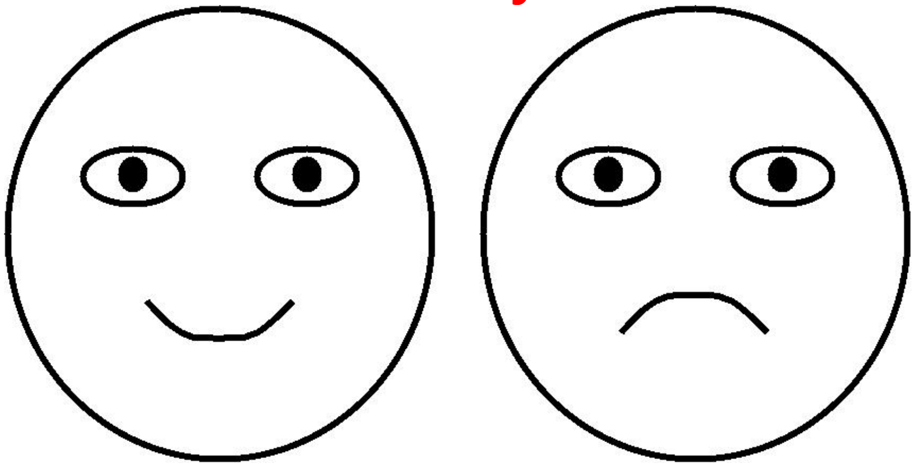
That's the basis of vicarious affordances and social affordances.

We can see some things as more than physical objects.

It's perception, not cognition because of how the perceptual details are in registration with the optic array.

Gibson noticed other abstract features of visual contents: features related to possible actions.

Visual contents can be very abstract, non-physical



Do the eyes look different? Why? How?
Compare illusory contours – Kanizsa

http://en.wikipedia.org/wiki/Illusory_contours

Problems

- What sorts of non-shape (including non-physical) information can a visual system get from the environment? (intention, mood, emotion, effort... compare Johansson movies)
- Why are these kinds of information useful – to what sorts of agents?
- How is the information represented in the perceiver?
- What mechanisms derive it from what information?
- What are the roles of evolution and development/learning in the creation/modification of such mechanisms?
- How do these mechanisms and forms of representation interact with others?
- How do the interactions help the perceiver, or the group.

The eyes are geometrically identical.

Beyond sensorymotor contents

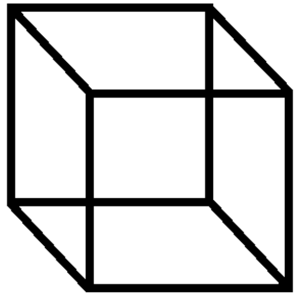
There are many philosophers and scientists who assume, like the old empiricist philosophers, that all semantic contents must be derived from experience, i.e. that all concepts must be defined in terms of processes of abstraction from sensory contents, or sensory-motor contents.

This "symbol–grounding theory" is a reinvention of "concept empiricism".

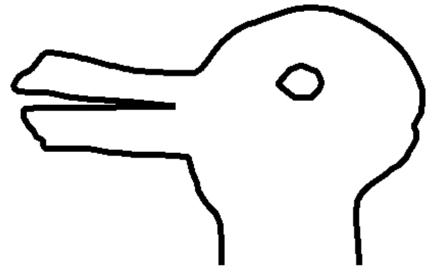
- However Kant argued in (Kant, 1781) that it is impossible to have experiences at all without having some concepts initially, e.g. of spatial structure and relations of adjacency, containment, etc.
- Then 20th Century Philosophers (Carnap, Nagel, Pap, and others) showed that many of the concepts required for scientific theories (e.g. atom, charge on an electron, gene, valence) cannot be derived from experience of instances.
- Alternative to symbol-grounding theory: some concepts are implicitly defined by their roles in deep explanatory theories, where the structure of the theory determines the possible models of the theory.
- In order to rule out unwanted models, "bridging rules" can be used, e.g. specifying how to measure values corresponding to certain concepts (e.g. "charge on an electron") and how to make predictions. (Carnap, 1947)

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

Beyond sensorymotor contents



Necker Cube



Duck-rabbit

The existence of ambiguous figures shows clearly that the sensory (e.g. retinal) contents do not uniquely determine what concepts are used in a percept.

Moreover in cases like the necker cube, what changes when the percept flips is 3-D structure.

The concepts of 3-D structure cannot be defined in terms of concepts of patterns in 2-D structures or 2-D structures plus motor signals.

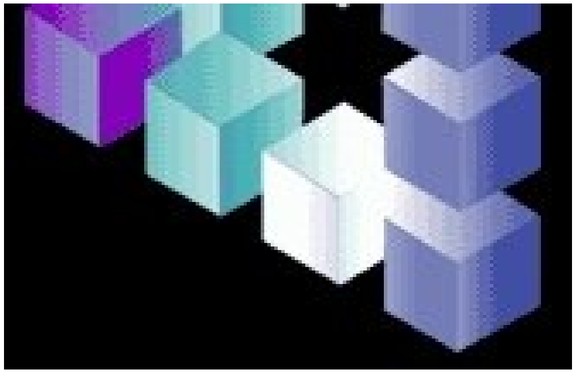
Going from retinal contents to 3-D percepts involves creative interpretation that goes beyond the data.

Likewise going from retinal contents to percepts of animal parts and the perception of an animal facing one way rather than another is not just a matter of grouping or abstracting sensory or sensory-motor contents, or storing 2-D or 3-D structural information.

See also (Sloman, 2009c) and
`http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#models`

The point is also illustrated by affordances: perceived possibilities and constraints, as shown below.

3-D structures – and possible actions



Here (on the right) is part of a picture by Swedish artist, Oscar Reutersvärd (1934) which you probably see as a configuration of coloured cubes.

As with the Necker cube you have experiences of both 2-D lines, regions, colours, relationships and also 3-D surfaces, edges, corners, and spatial relationships.

You probably also experience various affordances: places you could touch the surfaces, ways you could grasp and move the various cubes (perhaps some are held floating in place by magnetic fields).

E.g. you can probably imagine swapping two of them, thinking about how you would have to grasp them in the process – e.g. swapping the white one with the cube to the left of it, or the cube on the opposite side.

Experienced Possibilities, and representations

All of this suggests that much of what is experienced visually is collections of possibilities for change and constraints on change –

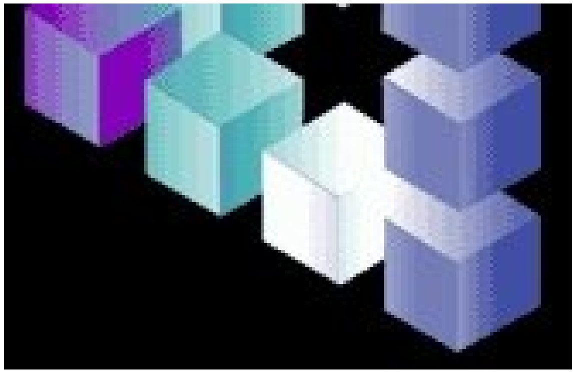
where these possibilities are attached to various fragments of the scene (via fragments of the image).

I.e. the possibilities and the constraints (e.g. obstacles) are located in 3-D space, in the environment, just as the objects and object fragments are.

Question: how are all these information fragments represented?

- Anyone who thinks it is possible for organisms or robots to do without representations must think it is possible for them to do without information.
- For information must somehow be encoded and whatever encodes the information is a representation – which could include a physical or virtual structure and much of its context.
- The representations used may be physical structures or structures in virtual machines – like the information a chess program has about the current state of the board, and the information about possible moves and their consequences.

2-D and 3-D Qualia: back to Reutersvard

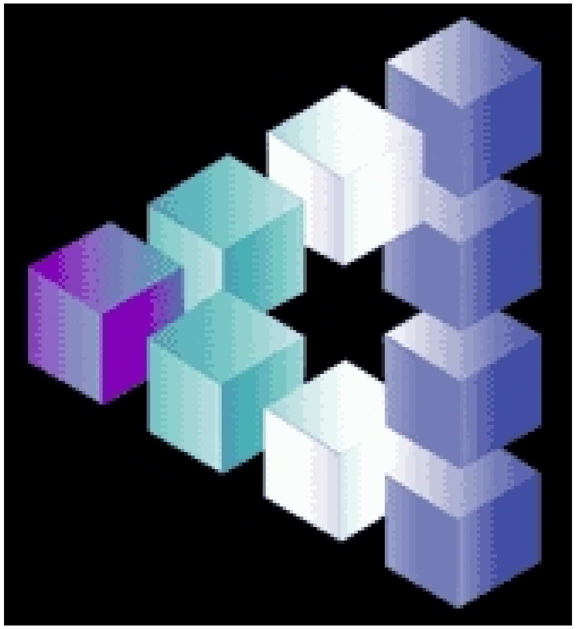


Here (on the right) is part of a picture by Swedish artist, Oscar Reutersvärd (1934) which you probably see as a configuration of coloured cubes.

As with the Necker cube you have experiences of both 2-D lines, regions, colours, relationships and also 3-D surfaces, edges, corners, and spatial relationships.

You probably also experience various affordances: places you could touch the surfaces, ways you could grasp and move the various cubes (perhaps some are held floating in place by magnetic fields).

E.g. you can probably imagine swapping two of them, thinking about how you would have to grasp them in the process – e.g. swapping the white one with the cube to the left of it, or the cube on the opposite side.



The second picture on the right (from which the first one was extracted) has a richer set of 2-D and 3-D contents.

Again there is a collection of 2-D contents (e.g. a star in the middle), plus experience of 3-D structures, relationships and affordances: with new possibilities for touching surfaces, grasping cubes, moving cubes.

The picture is outside you, as would the cubes be if it were not a picture. But the contents of your experience are in you: a multi-layered set of qualia: 2-D, 3-D and process possibilities.

But the scene depicted in the lower picture is geometrically impossible, even though the 2-D configuration is possible and exists, on the screen or on paper, if printed: the cubes, however, could not exist as shown.

So your qualia can be inconsistent!

That's impossible according to some theories of consciousness.
They can also go unnoticed: See `http://tinyurl.com/uncsee`

There are many sorts of things humans can see besides geometrical properties:

- that one object is supported by another,
- that one object constrains motion of another (e.g. a window catch),
- that something is flexible or fragile,
- which parts of an organism are ears, eyes, mouth, bill, etc.,
- which way something is facing,
- what action some person or animal is about to perform (throw, jump, run, etc.),
- whether an action is dangerous,
- whether someone is happy, sad, angry, etc.,
- whether a painting is in the style of Picasso...
- what information is available about X
- which changes in the scene or changes of viewpoint will alter available information about X.

  and other epistemic affordances.

Reading music or text fluently are important special cases, illustrating how learning/training can extend a visual architecture. (Sloman, 1978, Ch.9)

Attaching affordance information to scene fragments

I have argued that a lot of the content of visual perception (at least in adult humans) takes the form of information about what is and is not possible (possibilities and constraints) in various locations in the perceived scene – along with realised possibilities when processes are perceived

(e.g. objects or surfaces moving, rotating, deforming, interacting).

- The Reutersvard image shows how quite rich collections of possibilities can be "attached" to various scene fragments,
- Where some of the information is about possibilities and conditionals:
  - X is possible, Y is possible, Z is possible
  - if X occurs then, ...
  - if Y occurs then, ...
  - if Z occurs then, ...
- including "epistemic affordances"
  (if motion **M** occurs, then information **I** will become available/or become inaccessible).
- This can be seen as (yet another) generalisation of aspect graphs which encode only information about how changes of viewpoint (or rotations of a perceived object) will change what is and is not visible.
  `http://people.csail.mit.edu/bmcutler/6.838/project/aspect_graph.html#1`
- An important special case of this idea concerns the effects of different kinds of material on what would happen if various things were to happen:
  e.g. material being: rigid, plastic, elastic, liquid, viscous, sticky, etc.
  See `http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#babystuff`

The idea of generalising the concept of "aspect graph" as suggested here arose in a discussion with Jeremy Wyatt several years ago.

J. J. Gibson started a revolution

The Senses Considered as Perceptual Systems, 1966,
The Ecological Approach to Visual Perception, (Gibson, 1979).

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

in importantly different ways for different organisms.

The different functions arose at different stages in biological evolution, as both the physical environment became more complex and more structured, and the organisms and their predators, prey, mates, and offspring became more complex.

Some of the consequences were physical – development of independently movable manipulators: hands.

That had implications for information processing requirements.

E.g. use of an exosomatic rather than a somatic ontology.

Somatic = Marr's 'subjective' ??

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)
- Using invariants in static and changing optic arrays

  texture gradients, optical flow patterns, contrast edges, "common fate".
- Using actions to probe the environment, e.g. so as to change contents of the optic array

  The sensors and effectors work together to form "perceptual systems"(Gibson, 1966)
  (compare "active vision")

Vision is highly viewer-centred and action-centred.

The crazy bits in Gibson's theories:

- There are no internal representations, no reasoning.
- Perception is immediate and direct ("pickup", not "interpretation")

Also a brilliant idea:

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 mainly a device for sampling optic arrays: cones of information converging on viewpoints.

Perhaps in combination with brain area V1?

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)

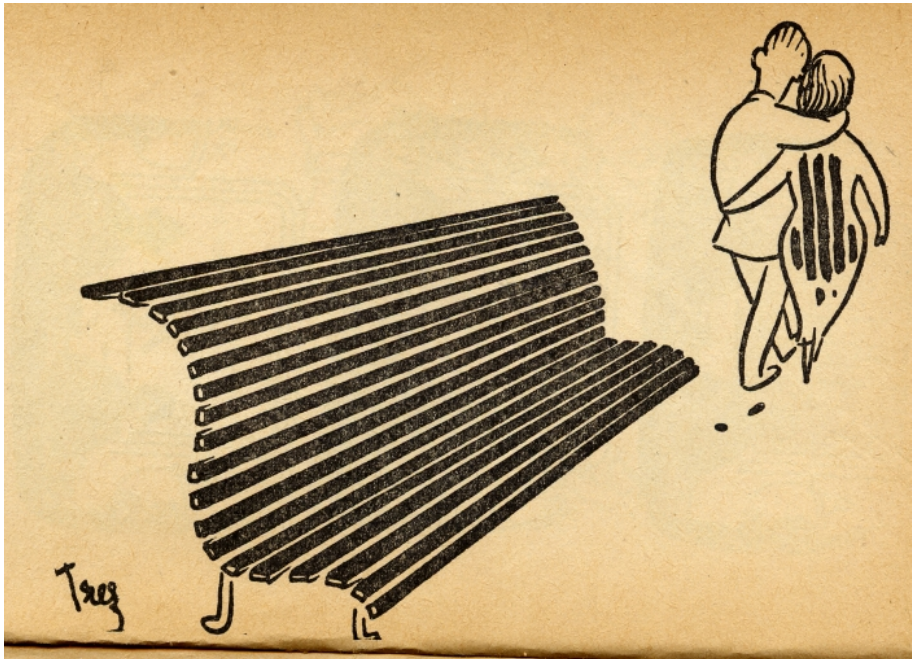
  If this string is pulled down, that pulley will rotate clockwise, causing that gear to turn, and ....
- geometric reasoning

  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.
But even they don’t allow for learning by working things out.

Example: Perceiving causation

Often our ability to perceive causal connections is used in humour:



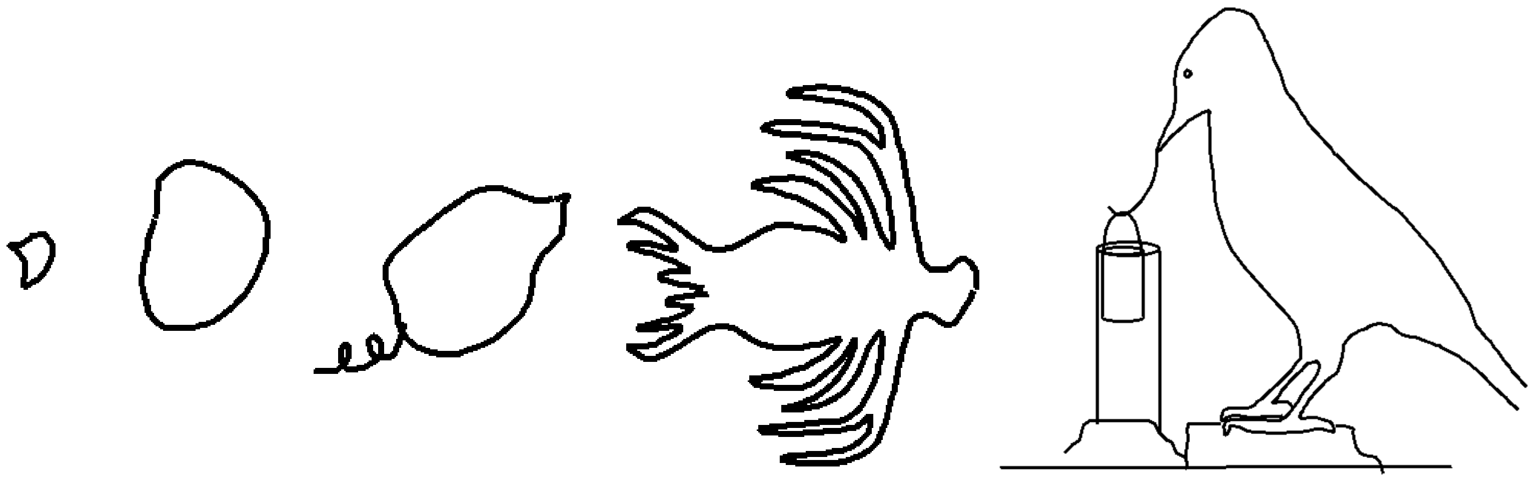
From a book of “French Cartoons”.

Some cartoons depend on our ability to see processes of various sorts.

In the above it’s a past process reaching into the present.

Some cartoons present a future process extending from the present –
e.g. a pompous pontificator heading unwittingly for a fall, collision, come-uppance, etc.

All organisms are information-processors but the information to be processed has changed and so have the means



From microbes to hook-making crows:
How many transitions in information-processing powers were required?

Contrast these transitions:

- transitions in physical shape, size and sensory motor subsystems
- transitions in information processing capabilities.

Fossil records don't necessarily provide clues.

`http://users.ox.ac.uk/~kgroup/tools/crow_photos.shtml`

`http://users.ox.ac.uk/~kgroup/tools/movies.shtml`

Environments have agent-relative structure

The environments in which animals evolve, develop, compete, and reproduce, vary widely in their information-processing requirements and opportunities.

If we ignore that environmental richness and diversity, our theories will be shallow and of limited use.

In simple environments everything can in principle be represented numerically, e.g. using numbers for location coordinates, orientations, velocity, size, distances, etc.

In practice the optic cone may lack the required clarity and detail; but for many purposes it may suffice to perceive partial orderings and topological relationships, as illustrated in `http://www.cs.bham.ac.uk/research/projects/cosy/papers/changing-affordances.html`

In more complex environment things to be represented include:

- Structures and structural relationships, e.g. what is inside, adjacent to, connected with, flush with, in line with, obstructing, supporting...
- Different sorts of processes, e.g. bending, twisting, flowing, pouring, scratching, rubbing, stretching, being compressed.
- Plans for future actions in which locations and arrangements and combinations of things are altered (e.g. while building a shelter).
- Intentions and actions of others.
- Past and future events and generalisations.

How can all those be represented? How can the information be used?

Varied environments produce varied demands

Types of environment with different information-processing requirements

- Chemical soup
- Soup with detectable gradients
- Soup plus some stable structures (places with good stuff, bad stuff, obstacles, supports, shelters – requiring enduring location maps.)
- Environments that record who moved where (permitting stigmergy).
- Things that have to be manipulated to be eaten (disassembled)
- Controllable manipulators
- Things that try to eat you
- Food that tries to escape
- Mates with preferences
- Competitors for food and mates
- Collaborators that need, or can supply, information.
- and so on .....

How do the information-processing requirements change across these cases?

For more on evolutionary transitions see
`http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#talk89`
Genomes for self-constructing, self-modifying information-processing architectures

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. in humans: 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) also 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. (Compare (Natsoulas, 1984))
- 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?)

- Functions of vision for other organisms may not be obvious:
  e.g. Portia Spider. (Tarsitano, 2006) `http://dx.doi.org/10.1016/j.anbehav.2006.05.007`

Beyond JJG: Generalised Gibsonianism (GG)

Less constrained analysis of the many functions of vision,

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

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 only to provide information about what objectively exists in the environment.

In particular, don't 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 changes in features of 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 shareable between different affordances: "proto-affordances".

Information about what changes are possible in a situation, and what constraints there are relating changes – independently of particular actions. (Sloman, 1996)

E.g. these two surface fragments could move closer; this surface fragment will obstruct that motion, that rolling ball cannot fit through that gap....

Beyond behavioural expertise

Karmiloff-Smith has drawn attention to the fact that once behavioural expertise in some domain has been achieved (e.g. ability to move at varying speeds, keeping upright on rough ground, avoiding obstacles, going through gaps, keeping up with others, etc.) there are sometimes new competences to be acquired, which she describes as involving “Representational Redescription”.

Example: after learning how to run with the herd some animals may develop the ability to remember what they did and did not do, and why they failed to do certain things. They might also be able not only to behave but also to reason about possible behaviours and their consequences prior to producing those behaviours (and often a necessary prerequisite for producing sensible behaviour).

See (Karmiloff-Smith, 1992) for more on varieties of representational redescription and their consequences.

I suspect that in some cases it’s a change in architecture rather than a change in representation that’s important.

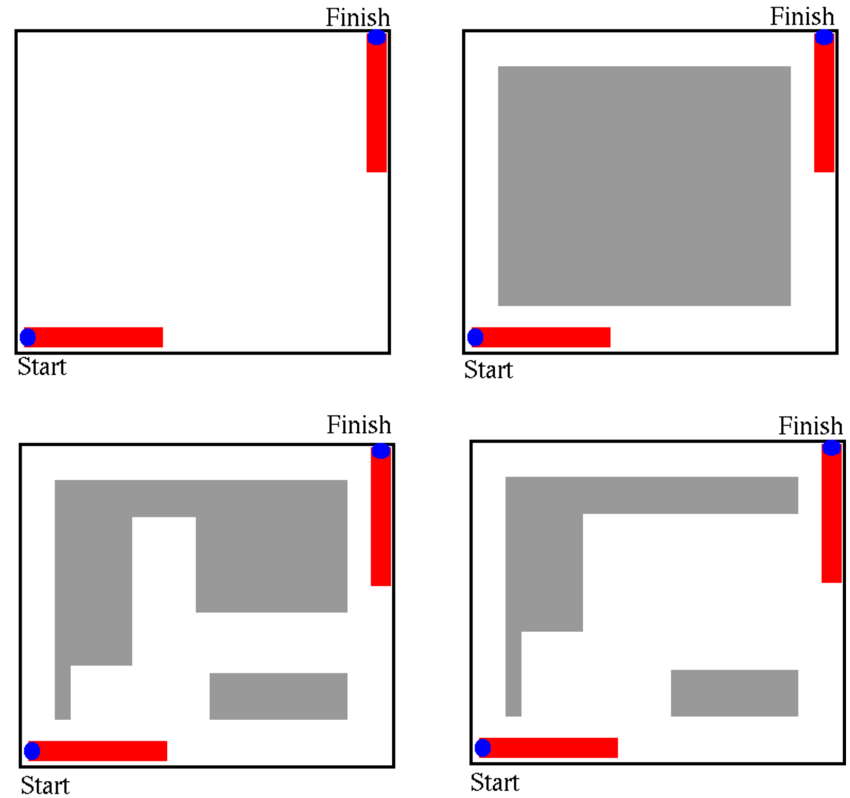
I have tried to bring out some of the implications of the book here
`http://www.cs.bham.ac.uk/research/projects/cogaff/misc/beyond-modularity.html`
(Work in progress.)

Many researchers in robotics and cognition, especially **embodied** cognition, ignore differences between behavioural expertise and additional requirements such as knowing why you did **not** do something, and what would have happened if you had done it.

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

If it is possible, how many significantly different ways of doing it are there?

(Based on a similar idea by Jeremy Wyatt)

This task uses the ability to detect the possibility of movements that are not happening,

and the constraints limiting such movements,

and to visualise combinations of such movements, while inspecting them for consequences:

Using what brain mechanisms?

Connections with evolution of mathematical competences

Long before there were mathematics teachers, our ancestors must have begun to notice facts about geometrical shapes that were later codified in Euclidean geometry.

Long before forms of logical reasoning powerful enough to serve the purposes of mathematicians were discovered/created by Frege, Russell and others in the 19th century, mathematicians were making discoveries and proving them, using their ability to notice possibilities for change, and invariants across change, in geometrical configurations.

I've discussed examples in (Sloman, 1968/9, 1971, 1978, 1996, 2010a)

If my generalisations of Gibson's notion of "perception of affordance" to cover a very much wider variety of perceptions of possibilities for change and constraints on change than he mentioned, are correct, then we can see how some of the roots of mathematical cognition as demonstrated in the discovery and proof of theorems in Euclidean geometry and topology may have developed from ancient animal abilities to perceive and reason about affordances required for selecting complex goals and plans.

Some fragments of Euclidean geometry concerned with possibilities are discussed here:
`http://www.cs.bham.ac.uk/research/projects/cogaff/misc/triangle-theorem.html`
`http://www.cs.bham.ac.uk/research/projects/cogaff/misc/triangle-sum.html`
Such animal capabilities were discussed by Kenneth Craik in (Craik, 1943).

I have tried to illustrate these capabilities in the processes of discovery of "toddler theorems" in young children, in (Sloman, 2008d).

For more information about toddler theorems and their relationships to Annette Karmiloff-Smith's ideas about "Representational Redescription" (Karmiloff-Smith, 1992) see

`http://tinyurl.com/BhamCog/misc/toddler-theorems.html`

Multi-strand relationships and processes

I discussed some of the representational problems arising from complexities of perception of processes at a conference in 1982 (Sloman, 1983).

In particular:

- Perceived objects whose parts move continuously can change their shapes, increasing or decreasing the numbers of features – e.g. a line growing a blob, a blob growing an appendage, two shapes merging, a moving string acquiring new changes of curvature, new inflection points, new contact points, or new junctions or knots.
- When two complex objects with parts are perceived together (e.g. two hands), not only are the wholes related but also the parts: they are related both within and across objects, e.g. a part of one object aligning with a part of another, and the relations can be of different sorts – metrical, topological, causal, etc.: "Multi-strand relationships".
- When objects exhibiting multi-strand relationships move or change shape, several of the pre-existing relationships can change in parallel: "Multi-strand processes".
- Perceiving and understanding multi-strand relationships and processes is required for many physical actions, e.g. washing up cups, saucers, bowls and cutlery, in a bowl of water using a dishmop, dressing or un-dressing a child, reversing a car into a narrow parking space, and many more.
- Description logics can be used to express static multi-strand relationships, but it is not clear whether they are useful for perception of complex multi-strand processes in which collections of relationships alter continuously or go in and out of existence (e.g. in a ballet performance, or rapids in a river).

Constraints on mechanisms
The problem of speed

Some experiments demonstrated here

`http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#talk88`

show the speed with which processing of an entirely new image can propagate to high level perceptual mechanisms – in humans.

The speed is hard to explain.

A possible explanation might refer to interactions between various sorts of dynamical system, processing interpretations at different levels of abstraction in parallel – as illustrated briefly in the following slides.

I conjecture that much visual learning (in humans and other species) involves growing new dynamical systems that lie dormant much of the time but can be very rapidly activated by information propagated upward, sideways or downward in the architecture –

in some cases triggering rapid actions (e.g. innate or learnt reflexes).

Competition and cooperation within and between such systems can cause rapid convergence to a particular combination of active sub-systems (multiple cooperating dynamical systems), whose continued behaviour is then driven partly by current changing perceptual inputs – when looking at processes in complex, changing scenes.

Sometimes there is much slower convergence, or non-convergence, in difficult situations: e.g. poor light, things moving too fast, blurred vision.

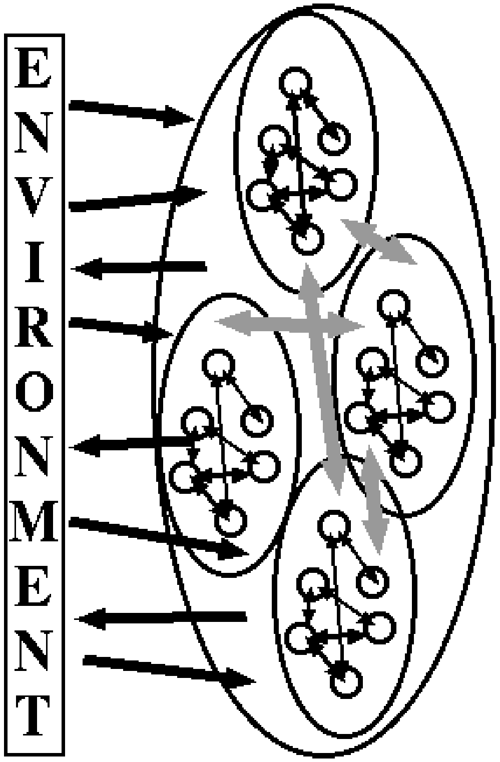
This extends the ideas demonstrated in the POPEYE program reported in (Sloman, 1978, Chapter 9).

Dynamical systems I

Many researchers who emphasise the importance of embodiment, also emphasise dynamical systems —

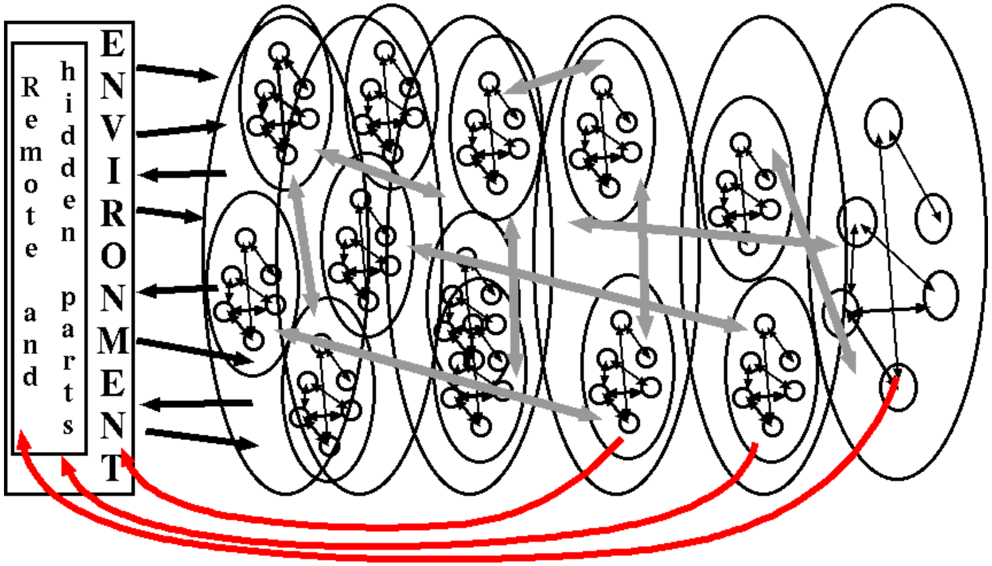
Especially dynamical systems closely coupled with the environment –

Where the nature of the coupling depends on the agent's morphology and sensory motor systems.



Dynamical systems II

A more complex dynamical system may involve large numbers of sub-systems many of which are dormant most of the time but can be re-activated as needed.



E.g. Perceiving, adaptation, learning, acting, self-monitoring can all involve information processed at multiple levels of abstraction.

Hypothetical reasoning: Science, mathematics, philosophy...

Some of the more abstract processes may be totally decoupled from the environment – but may produce results that are stored in case they are useful...

Which species can do what? – What are intermediate stages:
– in evolution?
– in individual development?

Do not believe symbol-grounding theory: use **theory-tethering** instead.

`http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#models`
(Label proposed by Jackie Chappell).

Further Issues That Need To Be Addressed

There's lots more to be said about the following issues, some discussed in online papers:

- Partial list of requirements for visual systems in intelligent animals and machines (many still not met): `http://www.cs.bham.ac.uk/research/projects/cogaff/misc/vision`
- Alternatives to use of global metrical coordinate systems for spatial structures and relationships: e.g. use networks of partial orderings (Contrast (Johansson, 2008))
- Alternatives to the use of probability distributions to cope with noise and uncertainty,

  e.g. use more abstract representations that capture only what is certain about structures, relationships and processes, e.g. X is getting closer to Y.

  `http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0702`
  Predicting affordance changes.
- Perception of processes of many kinds including not just spatial change, but also causal and epistemic changes, and intentional actions.
- The importance of perception of multi-strand relationships (not only objects have perceivable relationships, but also parts and fragments of those objects).
- Perception of multi-strand processes – processes in which there are multiple concurrent sub-processes, e.g. changing relationships.
- Why the idea that symbols need "grounding" is totally misconceived

  `http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#models`
- using a perceptual ontology that includes kinds of stuff.

  `http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#babystuff`
- How changes in ontologies, forms of representation, architectures, and uses of visual information, occur during evolution, development and learning.

Philosophy is one of the relevant disciplines

For example

- Conceptual analysis
- Metaphysics
- Kinds of meaning – semantics, ontologies
- What sorts of things can biological evolution do (in principle)?
- What is a form of representation?

  Donald Peterson, Ed., *Forms of representation: an interdisciplinary theme for cognitive science* Intellect Books, 1996,
- What forms of representation can a genome use?
- Kinds of machinery : e.g. physical, virtual
- Kinds of causation: e.g. in virtual machinery

NOTE: The ideas presented here are partly summarised and combined with a discussion of evolution of language in

`http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#talk111`

Talk 111: Two Related Themes (intertwined)

What are the functions of vision?

How did human language evolve?

(Languages are needed for internal information processing, including visual processing)

Summary of Main Points

Much vision research, unlike Gibson's, ignores functions of vision in on-line control and perception of continuous processes.

Work in robotics has begun to address small subsets of this.

Much vision research, including Gibson's, ignores

- **meta-cognitive perception (requiring meta-semantic competences)**

  (e.g. perceiving intentions, mood, how others feel, what they are trying to do, what they are trying to communicate, etc., where meta-semantic concepts are part of the ontology required)

- **perception of possibilities and constraints on possibilities, including many kinds of affordance besides affordances for immediate action by the perceiver,**

  including proto-affordances, vicarious affordances, auto-vicarious affordances, epistemic affordances, multi-step branching affordances, geometric and topological constraints, ...

- **the associated role of vision in reasoning and doing mathematics,**

  e.g. using actual or imagined diagrams, or visualising possible changes and their consequences, in a perceived machine or other complex structure.

- **The visual mechanisms that allow cognitive contents to be kept in registration with the optic array ((Trehub, 1991) attempts to address this)**

Not all organisms with visual systems have all these capabilities: identifying subsets and how they evolved is a major research problem – part of the Meta-Morphogensis project:
`http://www.cs.bham.ac.uk/research/projects/cogaff/misc/meta-morphogenesis.html`
Even young humans develop them piece-meal in ways that depend in part on their physical and social environment. See (Karmiloff-Smith, 1992; Sloman, 2008d; Chappell & Sloman, 2007)

NOTE: Proto-affordances were also recognized in (Siegel, 2014), and the same label proposed.

Related work (a small subset)

References

Barrow, H., & Tenenbaum, J. (1978). Recovering intrinsic scene characteristics from images. In A. Hanson & E. Riseman (Eds.), *Computer vision systems* (pp. 3–26). New York: Academic Press.

Boden, M. A. (2006). *Mind As Machine: A history of Cognitive Science (Vols 1–2)*. Oxford: Oxford University Press.

Carnap, R. (1947). *Meaning and necessity: a study in semantics and modal logic*. Chicago: Chicago University Press.

Chappell, J., & Sloman, A. (2007). Natural and artificial meta-configured altricial information-processing systems. *International Journal of Unconventional Computing*, *3*(3), 211–239. Available from `http://www.cs.bham.ac.uk/research/projects/cogaff/07.html#717`

Craik, K. (1943). *The nature of explanation*. London, New York: Cambridge University Press.

Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems*. Boston: Houghton Mifflin.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin.

Johansson, I. (2008). Functions and Shapes in the Light of the International System of Units. *Int Ontology Metaphysics*, *9*, 93–117. Available from `DOI10.1007/s12133-008-0025-z`

Kant, I. (1781). *Critique of pure reason*. London: Macmillan. (Translated (1929) by Norman Kemp Smith)

Karmiloff-Smith, A. (1992). *Beyond Modularity: A Developmental Perspective on Cognitive Science*. Cambridge, MA: MIT Press.

Natsoulas, T. (1984, July). Towards the Improvement of Gibsonian Perception Theory. *J. for the Theory of Social Behaviour*, *14*(2), 231–258. Available from `DOI:10.1111/j.1468-5914.1984.tb00496.x`

Siegel, S. (2014). Affordances and the Contents of Perception. In B. Brogaard (Ed.), *Does Perception Have Content?* (pp. 39–76). USA: OUP. Available from `http://philpapers.org/rec/SIEAAT`

Simon, H. A. (1969). *The sciences of the artificial*. Cambridge, MA: MIT Press. ((Second edition 1981))

Sloman, A. (1968/9). Explaining Logical Necessity. *Proceedings of the Aristotelian Society*, *69*, 33–50. Available from `http://www.cs.bham.ac.uk/research/projects/cogaff/62-80.html#1968-01`

Sloman, A. (1971). Interactions between philosophy and AI: The role of intuition and non-logical reasoning in intelligence. In *Proc 2nd ijcai* (pp. 209–226). London: William Kaufmann. Available from `http://www.cs.bham.ac.uk/research/cogaff/62-80.html#1971-02` (Reprinted in *Artificial Intelligence*, vol 2, 3-4, pp 209-225, 1971)

Sloman, A. (1978). *The computer revolution in philosophy*. Hassocks, Sussex: Harvester Press (and Humanities Press). Available from `http://www.cs.bham.ac.uk/research/cogaff/62-80.html#crp`

Sloman, A. (1983). Image interpretation: The way ahead? In O. Braddick & A. Sleigh. (Eds.), *Physical and Biological Processing of Images (Proceedings of an international symposium organised by The Rank Prize Funds, London, 1982.)* (pp. 380–401). Berlin: Springer-Verlag. Available from `http://www.cs.bham.ac.uk/research/projects/cogaff/06.html#0604`

Sloman, A. (1989). On designing a visual system (towards a gibsonian computational model of vision). *Journal of Experimental and Theoretical AI*, *1*(4), 289–337. Available from `http://www.cs.bham.ac.uk/research/projects/cogaff/81-95.html#7`

Sloman, A. (1993). The mind as a control system. In C. Hookway & D. Peterson (Eds.), *Philosophy and the cognitive sciences* (pp. 69–110). Cambridge, UK: Cambridge University Press. Available from `http://www.cs.bham.ac.uk/research/projects/cogaff/81-95.html#18`

Sloman, A. (1994). How to design a visual system – Gibson remembered. In D.Vernon (Ed.), *Computer vision: Craft, engineering and science* (pp. 80–99). Berlin: Springer Verlag. Available from `info:vKA5pOB9aq8J:scholar.google.com`

Sloman, A. (1996). Actual possibilities. In L. Aiello & S. Shapiro (Eds.), *Principles of knowledge representation and reasoning: Proc. 5th int. conf. (KR '96)* (pp. 627–638). Boston, MA: Morgan Kaufmann Publishers. Available from `http://www.cs.bham.ac.uk/research/cogaff/96-99.html#15`

Sloman, A. (2001). Evolvable biologically plausible visual architectures. In T. Cootes & C. Taylor (Eds.), *Proceedings of British Machine Vision Conference* (pp. 313–322). Manchester: BMVA. Available from `http://www.cs.bham.ac.uk/research/projects/cogaff/00-02.html#76`

Sloman, A. (2002). Diagrams in the mind. In M. Anderson, B. Meyer, & P. Olivier (Eds.), *Diagrammatic representation and reasoning* (pp. 7–28). Berlin: Springer-Verlag. Available from `http://www.cs.bham.ac.uk/research/projects/cogaff/00-02.html#58`

Sloman, A. (2005). *Perception of structure: Anyone Interested?* (Research Note No. COSY-PR-0507). Birmingham, UK. Available from http://www.cs.bham.ac.uk/research/projects/cosy/papers/#pr0507

Sloman, A. (2007). *What evolved first and develops first in children: Languages for communicating? or Languages for thinking? (Generalised Languages: GLs)* (No. COSY-PR-0702). Birmingham, UK. (Presentation given to Birmingham Psychology department. http://www.cs.bham.ac.uk/research/projects/cosy/papers/#pr0702)

Sloman, A. (2008a). Architectural and representational requirements for seeing processes, proto-affordances and affordances. In A. G. Cohn, D. C. Hogg, R. Möller, & B. Neumann (Eds.), *Logic and probability for scene interpretation.* Dagstuhl, Germany: Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany. Available from http://drops.dagstuhl.de/opus/volltexte/2008/1656

Sloman, A. (2008b). *Evolution of minds and languages. What evolved first and develops first in children: Languages for communicating, or languages for thinking (Generalised Languages: GLs)?* (Research Note No. COSY-PR-0702). Birmingham, UK. Available from http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#glang

Sloman, A. (2008c). *A Multi-picture Challenge for Theories of Vision* (Research Note No. COSY-PR-0801). Available from http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#talk88

Sloman, A. (2008d). *A New Approach to Philosophy of Mathematics: Design a young explorer, able to discover "toddler theorems".* Available from http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#toddler

Sloman, A. (2009a). Architectural and representational requirements for seeing processes and affordances. In D. Heinke & E. Mavritsaki (Eds.), *Computational Modelling in Behavioural Neuroscience: Closing the gap between neurophysiology and behaviour.* (pp. 303–331). London: Psychology Press. (http://www.cs.bham.ac.uk/research/projects/cosy/papers#tr0801)

Sloman, A. (2009b). *Ontologies for baby animals and robots. From "baby stuff" to the world of adult science: Developmental AI from a Kantian viewpoint.* University of Birmingham, School of Computer Science. Available from http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#brown (Online tutorial presentation)

Sloman, A. (2009c). Some Requirements for Human-like Robots: Why the recent over-emphasis on embodiment has held up progress. In B. Sendhoff, E. Koerner, O. Sporns, H. Ritter, & K. Doya (Eds.), *Creating Brain-like Intelligence* (pp. 248–277). Berlin: Springer-Verlag. Available from http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0804

Sloman, A. (2010a). If Learning Maths Requires a Teacher, Where did the First Teachers Come From? In A. Pease, M. Guhe, & A. Smaill (Eds.), *Proc. Int. Symp. on Mathematical Practice and Cognition, AISB 2010 Convention* (pp. 30–39). De Montfort University, Leicester. Available from http://www.cs.bham.ac.uk/research/projects/cogaff/10.html#1001

Sloman, A. (2010b). Phenomenal and Access Consciousness and the "Hard" Problem: A View from the Designer Stance. *Int. J. Of Machine Consciousness*, *2*(1), 117–169. Available from http://www.cs.bham.ac.uk/research/projects/cogaff/09.html#906

Sloman, A. (2011). What's information, for an organism or intelligent machine? How can a machine or organism mean? In G. Dodig-Crnkovic & M. Burgin (Eds.), *Information and Computation* (pp. 393–438). New Jersey: World Scientific. Available from http://www.cs.bham.ac.uk/research/projects/cogaff/09.html#905

Sloman, A., & CoSy project members of the. (2006, April). *Aiming for More Realistic Vision Systems* (Research Note: Comments and criticisms welcome No. COSY-TR-0603). School of Computer Science, University of Birmingham. (http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0603)

Tarsitano, M. (2006, December). Route selection by a jumping spider (Portia labiata) during the locomotory phase of a detour. *Animal Behaviour*, *72, Issue 6*, 1437–1442. Available from http://dx.doi.org/10.1016/j.anbehav.2006.05.007

Trehub, A. (1991). *The Cognitive Brain.* Cambridge, MA: MIT Press. Available from http://people.umass.edu/trehub/