

DRAFT Extended Abstract

Understanding the Functions of Animal Vision

What Are We Trying To Do:
How Do Logic And Probability
Fit Into The Bigger Picture?

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

January 30, 2008

Abstract

As I said when I received the original invitation I don't have expertise regarding *probabilistic* approaches. It seems to me that insofar as manipulation of probabilities has a role in connection with uncertainty due to noise, poor resolution, occlusion, aperture problems, etc. we have no hope of producing good mechanisms unless we have very clear and effective ideas about what needs to be represented when there is NO uncertainty and how that information can be represented, transformed, and used.

Putting in probabilistic mechanisms too soon is like building a repair kit for an engine before you have designed the engine.

As far as the use of logic is concerned, I think that is merely one kind of representation, which is very useful because of its generality, but for many problems involving spatial structures, processes and causal interactions it can be more useful to use spatial (geometric and topological) representations, though not necessarily isomorphic with what they represent – as pointed out in my IJCAI 1971 discussion of the importance of both Fregean and analogical representations, now online here: <http://www.cs.bham.ac.uk/research/projects/cogaff/04.html#200407>

However it has proved very difficult to design computer based virtual machines with the required properties. Perhaps that is because we are still not clear enough about the requirements. My work is mostly about requirements, but I have some sketchy design ideas.

Keywords:

analogical & Fregean representation animal vision, causation, geometry, processes, proto-affordances, representation, structures, topology

Some examples to think about

Here are some things you can use vision to do.

When you can see part of a room through an open doorway, what actions can you perform in order to see more of the contents of the room, and how do you know that those actions will help? What are the differences between moving forwards to see more and moving left and right to see more? How do you know?

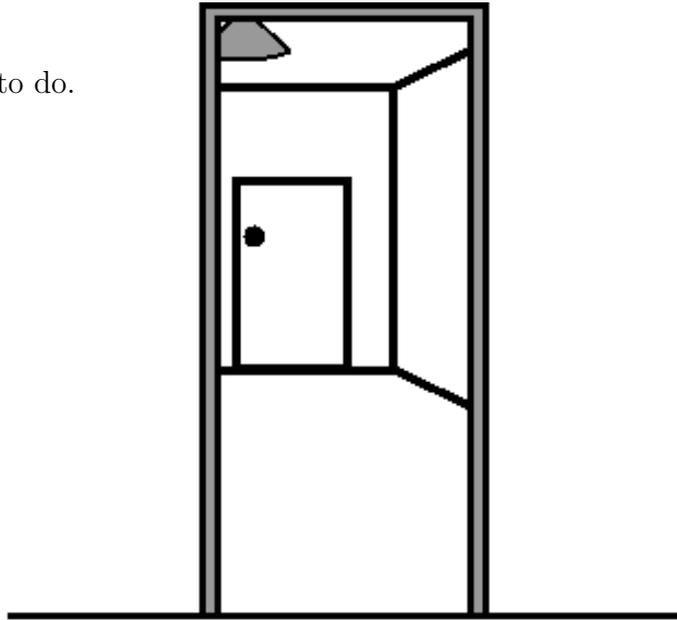


Figure 1

If the chair is too wide to be pushed through the doorway, what other options are available to get it through?

If a door opens when you push it away from you, why is no handle needed? Why do you need a handle to shut it? What happens if you try to shut it without using a handle? (Will it hurt?)

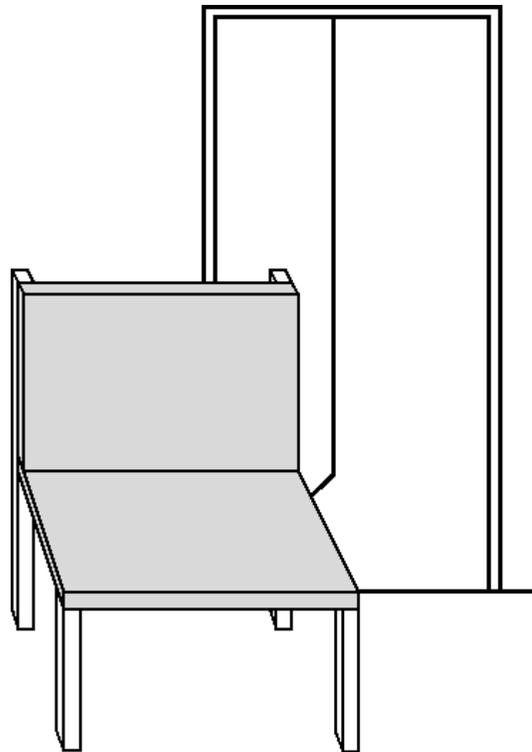


Figure 2

When you look at meshed gear wheels how are you able to work out which way one of the wheels will rotate when you move the other?

What are you assuming about the materials of the wheels? What has to change in a newborn infant to enable it to understand this? How many others can? What would enable a robot to understand it? (Compare Kant and Hume on causation.)

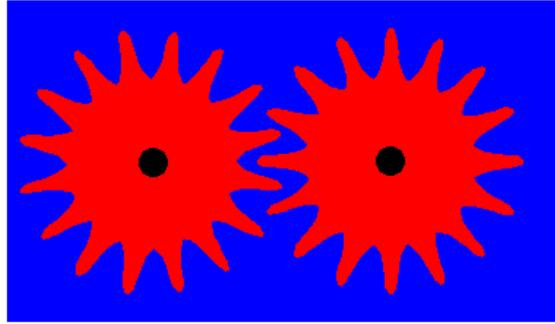


Figure 3

Possible topics for discussion

Depending on the opportunities and the interests of others, I could talk about a collected of related issues that interest me:

- I am not interested in solving specific engineering problems using TV cameras, though I am interested in trying to understand what the functions of animal vision (including human vision) are and what sorts of information-processing mechanisms and architectures can implement those functions.
- Vision is not *primarily* about recognition, since you cannot learn to recognise anything that you cannot already see. Seeing objects is prior to recognising them, though you may need to recognise types of object-fragments and relationships (e.g. types of surface fragment, and relationships between surface fragments) in order to see something.
- Recognition is a secondary function, and there are many different kinds of recognition which differ both in the content of the visual information (e.g. perception of processes, functions, causal interactions, affordances, dangers, etc. See below.) and in the uses to which the information is put (e.g. continuous servoing, testing generalisations, explaining, answering questions, generating new goals, predicting, designing, communicating,)
- Vision has been widely interpreted as being concerned with acquiring information about spatial structures in the environment, whereas, for organisms, perception of static structures is a special case of perception of *processes*, i.e. static structures are processes where nothing is changing.

(Structures are also parts of processes. And structures in the environment can be inferred from perceived processes.)

- J.J.Gibson's revolution (a bit like the Copernican revolution) was to point out that the role of animal vision is not to provide (viewer independent, though not viewpoint independent)

information about what is ‘out there’ in the world (e.g. as assumed by Marr and many others) but rather to acquire information about what the viewer can and cannot do and with what consequences: These are positive and negative affordances, using an ontology that depends on the possible actions and possible goals of the animal.

J. J. Gibson, *The Ecological Approach to Visual Perception*, Houghton Mifflin, Boston, MA, 1979

- I think this may be all that *some* species can do, and in many of them the ability to acquire and use information about affordances is ‘compiled in’ by evolution, whereas some animals (e.g. humans, hunting mammals, primates, nest-building birds, animals that have to dismember prey in order to eat) seem to have to acquire visual competences that are more general and allow novel situations to be dealt with: I suggest that in those cases perception of affordances depends on the ability to perceive lower level structures: proto-affordances.

- The ability to perceive not only processes that are occurring in the environment but the possibilities of processes that *could* occur, and constraints on those possibilities is a key function of vision in intelligent animals, and is the foundation for the perception of affordances. (See “Actual Possibilities” in KR’96

<http://www.cs.bham.ac.uk/research/projects/cogaff/96-99.html#15>)

- Perceiving proto-affordances involves perceiving what processes are *possible*, or *constrained* by a particular situation, independently of whether those processes are produced by or useful for the viewer or produced by or useful to any other agent. E.g. seeing that a twig can fit into an opening, or that a surface of one object could move closer to or further from a surface of another object, or that a flexible or articulated object has a shape that could change in certain ways.

- The ability to see proto-affordances may be common to the ability to see affordances for the viewer, and also affordances for others: *vicarious* affordances. The others could be predators, prey, or conspecifics, e.g. infants needing help or protection. (There is much confusion about mirror neurones).

- The ability to perceive processes requires the ability to *represent* processes of many kinds, including multi-strand processes in which multiple relationships (including metrical, topological, causal, functional relationships) at different levels of abstraction change concurrently, though not all admit temporal description at the same level of grain.

- As far as I know, there are as yet no good proposals regarding how such processes can be represented, both while they are being perceived, and when the information about what has been perceived is used later, except in very simple cases (e.g. visual servoing of simple actions).

[E.g. is a perceived process such as someone walking across the room. recorded as a repeated process, like auditory memories that use rehearsal, or as a description that can generate a process when required (e.g. a specification for simulations), or as a static information

structure that can be put to various uses, or ...?]

- The ability to perceive empty spaces is part of the ability to perceive possibilities. [How is empty space represented by a visual system? How do you see a blank sheet of paper? What would Picasso have seen there?]

- The ability to perceive and use information about actual and possible processes (and the structures involved in those processes) is one of the foundations of human mathematical capabilities. So any theory of vision that does not contribute to the explanation of those mathematical capabilities, e.g. reasoning in Euclidean geometry and topology, is a mistaken or at best a partial theory of vision. [This is connected with perceiving Kantian structure-based causal relationships, as opposed to probabilistic Humean causal relationships.]

<http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#math-robot>

(How could child robot grow up to be a mathematician or a philosopher. PDF presentation.)

- Vision can work together with other forms of perception, including auditory, haptic, proprioceptive and vestibular information processing: and we need to understand what forms of representation facilitate the integration of information from those various sources. [E.g. contrast the economy of a-modal exosomatic representations and ontologies with vs multi-modal somatic representations and ontologies, where the latter are only concerned with relations between sensory and motor signals at various levels of abstraction, and the latter are concerned with what exists outside and independently of the agent.]

- One of the facts about vision seems to be that people born blind can still make use of important brain mechanisms that evolved in part to serve the functions of vision. A good theory of vision should explain how that works. (E.g. can haptic information processing in congenitally blind people make use of visual forms of representation, and processing mechanisms.)

- Likewise there are things to be explained about how people with several physical abnormalities or deformities (e.g. people born limbless) use visual and other brain mechanisms that evolved to support normal bodies.

- One of the under-rated functions of vision is provision of information not about the physical environment, or about affordances for action, but about *epistemic affordances*: e.g. information about what information is available and what you have to do to get it, etc.

- Closely related is the role of vision in predicting how affordances (both action affordances and epistemic affordances can change if various possibilities are realised.) [Aspect graphs capture a special case of this.]

<http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0702>

Predicting Affordance Changes (HTML)

- Any good theory about the use of probabilities or any other form of representation for

handling uncertainty in visual processing should, as a special case, determine good ways to deal with visual information when there is NO uncertainty and everything is deterministic (e.g. seeing how an old-fashioned clock works, when everything is very clearly visible, etc.)

- In humans (and probably some other altricial species) a visual system does not have a fixed set of functions but can extend itself to cope with new ontologies that are not definable in terms of the initial ontologies, and new functions – e.g. learning to see new kinds of causal relations, learning to see new kinds of functional roles, learning to read music, learning to see intentions in actions, learning to see threats in board games, learning to see dances, learning to see computer program structures, including learning to detect bugs. Understanding how that architectural growth can occur seems to be one of the major unsolved problems of vision.

- The speed with which many visual functions can be performed, along with the complexity of the information processed and the variety of cases that an individual can handle from one moment to the next, suggest that none of the currently available AI mechanisms for processing information can do the job. None of the neural theories I have heard presented provide adequate explanations either.

<http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#compmod07>

Architectural and representational requirements for seeing processes and affordances (PDF)

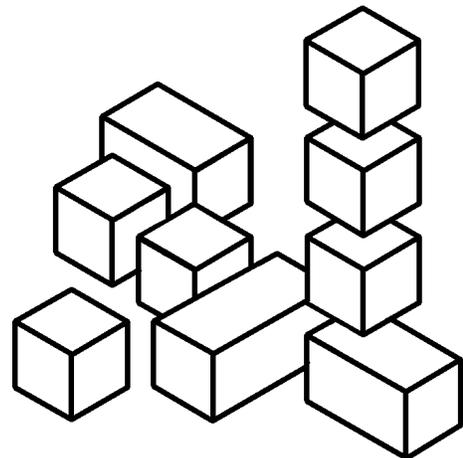
- How brains do it probably cannot be determined by bottom up neuroscientific investigations: some deep new theory may be required to direct research into what brains actually do.

I have more questions than answers, alas.

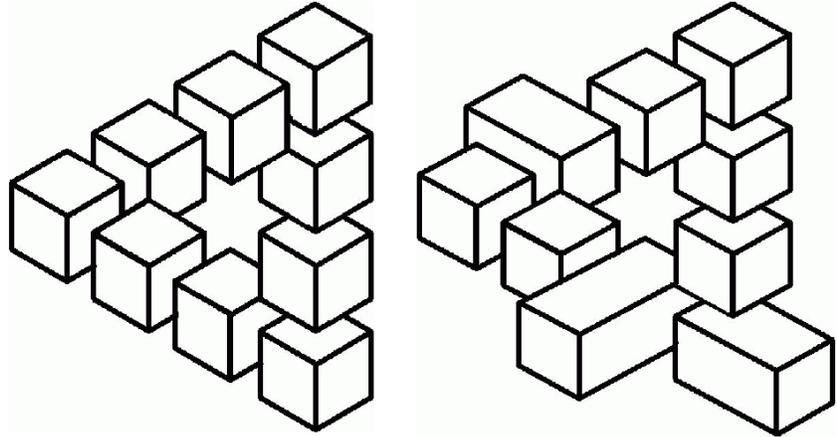
I have tried to be provocative.

A few final provocations follow:

Could you build something like this if I provided you with a set of rectangular blocks?



What about building these?



The above examples – inspired by Oscar Reutersvärd (1934)¹ – show that 3-D perception does not involve building internal objects that are *isomorphic* with the things seen.



No AI vision program comes close to seeing the affordances and proto-affordances you can see in these pictures (despite the low resolution and noise). Why is getting machines to do this so hard? The main problem is finding the kind of representation of what is in the scene. I suspect that we can generalise generalised aspect graphs to include collections of proto-affordances in registration with various parts of the optic array in some representation that preserves important relationships without necessarily having all the detailed metrical structure.

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

There is lot more related material in presentations and papers here:

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

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

Last modified: January 30, 2008

¹<http://www.sandlotscience.com/EyeonIllusions/Reutersvard.htm>