School of Computer Science
University of Birmingham UK
CN - CR
Every intelligent ghost must contain a machine
an information-processing machine

Simplicity and Ontologies
The trade-off between simplicity of theories and sophistication of ontologies
(and the need for architectural changes during development)

Aaron Sloman
DRAFT: COMMENTS WELCOME
(This is work-in-progress and is likely to remain so for some time. Help needed.)

Installed: 21 Sep 2010
Last updated:
25-5 Jul 2014 (Further major revision and reorganisation)
18-23 Jul 2014 (fixed broken link, architectural development, reformatted, reorganised, clarified)
7 Jul 2011; 29 Jul 2011; 19 Sep 2011;
22 Sep 2010;23 Sep 2010; 26 Sep 2010; 12 Dec 2010; 21 Apr 2011; 5 Nov 2011

This paper is
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/simplicity-ontology.html
From time to time a derived PDF version will be created automatically:
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/simplicity-ontology.pdf

A partial index of discussion notes, including related work, is in
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/AREADME.html

---

Abstract [Added 27 Jul 2014]

Many researchers work on systems that learn. Different sorts of things can be learnt, including abilities of many kinds, and applications of those abilities. Learnable abilities include (in no significant order): abilities to label, abilities to describe, abilities to predict, abilities to manipulate directly or to control indirectly, abilities to perform actions of varying complexity and difficulty, abilities to classify, abilities to explain, abilities to evaluate, abilities to appreciate, abilities to plan, abilities to carry out plans, including modifying or extending them during execution, abilities to design things, abilities to make things to satisfy a need, abilities to prove things, make inferences, calculate or reason, abilities to discover things, abilities to communicate (using language or other media), abilities to understand communications, abilities to teach or assist others, abilities to collaborate as leader or subordinate or equal, abilities to empathise, abilities to resolve conflicts, within oneself or between individuals, and many more related abilities.

All those abilities are generic and may have sub-cases that have to be learnt separately, and in some cases the learning can include increasing speed, fluency, reliability and accuracy of performance.

Intelligent abilities require use of knowledge about types of things that can exist or happen, i.e. knowledge of an ontology. A simple homeostatic controller, e.g. a thermostat, may use a very simple ontology perhaps including contents of a form of sensory input (e.g. temperatures) and contents of a form of output e.g. ’raise’ or ’lower’ signals to a heater or cooler.

Organisms (and future robots) with multiple modes of sensing and acting on a complex independently existing environment need ontologies that straddle modes of perception and action, for instance the ability to express where something is or how it is moving, irrespective of how the object’s location is sensed or changed.

If we had a better understanding of how various ontologies used for various purposes in organisms evolved, and how they develop in individuals, we might be better able to design machines with intelligence that matches those of animals, including humans. Instead we can now only produce machines with very shallow and restricted abilities, that often turn out to be very brittle when dealing with novelty.

Theories and their ontologies

```
Albert Einstein once wrote:

    "It can scarcely be denied that the supreme goal of all theory is to
    make the irreducible basic elements as simple and as few as possible
    without having to surrender the adequate representation of a single
    datum of experience."
    In: On the Method of Theoretical Physics
       Philosophy of Science, Vol. 1, No. 2 (Apr., 1934), pp. 163-169
       http://www.jstor.org/stable/184387
```

Often, when people discuss the role of simplicity in science, they do not notice the trade-off between simplicity of ontology and simplicity of theory using an ontology. Einstein appears to have been emphasising simplicity of ontology (basic elements), though he might have included theory as well (basic axioms/assumptions).

The ontology used by a theory is determined in part by

```
(a) the syntax of the formalism that it uses,
    and
(b) the variety of ’atomic’ components  of the formalism that
    are not explicitly defined as abbreviations for something
    expressible using other components.
(The notion of atomic component is expanded below.)
```

The atomic components may have some semantic content -- referring to possible types of entity, process, event, property, material, relation, disposition, constraint, causal interaction, function, or whatever, in the portion of the world that the theory is about.

The atomic components may be **aspects** rather than **parts** of a form of representation. For example, in a class of maps different colours or textures may be used for otherwise similar parts to indicate some feature of whatever is represented by whatever has the colour or texture. For example, colours might be used to distinguish rivers, roads, footpaths, railway lines, and routes travelled by various migrating herds. Or they could be used to distinguish different connected subsets in a map with a complex topology, as in the London Tube map. In our discussion of a machine or organism learning to interpret visual sensory input we'll ignore most of these possibilities, though in actual biological systems and in robots they can be important.

Another sort of atomic component is a type of relationship between parts. For example in a map or visual sensory array, neighbourhood relationships in different directions may be interpreted as depicting various spatial relationships in the source of the visual information, possibly in a context-sensitive way -- i.e. what a spatial relationship means depends on the context. For example, when a 2-D image represents a 3-D structure, the local and global context can determined whether a vertical line in the image represents a horizontal crack in the floor, or a vertical edge of a far wall, or a cord hanging vertically from the ceiling, or a horizontal crack in the ceiling. (See Sloman(1971) for examples and an explanation of the difference between Fregean and Analogical forms of representation (among others).

The need for mechanism
Often, when people talk about how something is represented they are thinking of things humans create outside themselves in order to store or manipulate information, for instance in pictures, maps, sentences, equations, computer programs, blueprints, tables, graphs, street signs, direction indicators at road junctions, and many different types of display on machines humans interact with, including the driver's displays on the fascia of a car (automobile), aeroplane or a machine controlled in a factory.

Unfortunately, some researchers seem to think cognition and perception are about **objects**, and ignore all the semantic contents and perceptual contents that are not about objects, including aspects or fragments of objects (e.g. surface fragments relevant to some manipulation task), relations, processes, causal interactions, opportunities, constraints, goals, values, theories, or explanations, for example. Relations can be static, e.g. proximity, alignment, containment, obscuring, or dynamic, e.g. approaching, rotating around, obscuring more of, and many more.

Some theories have components that use ontologies at different levels of abstraction, as I shall illustrate below.

Main Thesis:

> A theory can be extended by adding levels of abstraction that are not **definable** in terms of previous descriptive mechanisms. This can add to the complexity of a theory's basic ontology, while either

(a) simplifying the structure of the theory and its deployment in explaining and predicting specific phenomena, or

(b) significantly extending the explanatory and predictive scope of the theory.

Or both!

(We'll later see an example where specifying a 3-D process and a projection process -- producing a 2-D shadow -- enormously simplifies the description of a 2-D process.)

In particular, adding new indefinables can alter search spaces relevant to solving problems, finding explanations, making plans, or learning useful generalisations.

Example from perception

If things perceived are not directly coupled with the sensors perceiving them then all sorts of factors other than changes in a perceived object can produce sensory changes, for example, a light going on or off or changing its intensity, or another object blocking the view.

In such cases, ideas from signal detection theory, and standard analyses of sensors as partly noisy or probabilistic measuring devices become irrelevant to understanding some of the most important features of perceptual processes and mechanisms. For example, this point seems to be missed in the otherwise very interesting six lectures on information theory and perception by Prof. William Bialek available on Youtube, starting here:
http://www.youtube.com/watch?v=naRbEddGTMY&list=PLoxv42WBtfCAY8icy7uChz_kpBXpWoMwk
(Closely related to "Signal Detection Theory"
http://en.wikipedia.org/wiki/Detection_theory)

Such theories may be very useful in understanding ancient organisms able to sense and respond to chemicals in contact with their enclosing membranes, internal and external pressures (e.g. osmotic pressure), temperature, illumination, direction of gravitational force, rate of flow of liquid with which it is in contact, concentrations of various chemicals in its immediate environment, or its own interior, and perhaps other measurable states of itself or the environment.

But the more

Examples of theory extension based on ontology extension

Development of the theory of chemical structures and interactions, which required extending our ontology to describe atoms, molecules, their properties, and types of chemical bonds (sometimes summarised in chemical formulae), provided vast new explanatory power compared with the previously available ways of describing observed phenomena when substances interact. The main advance of the theory was to explain the possibility of far more types of matter and types of interactions of matter than had previously been discovered or imagined. The ability to explain possibilities need not provide the ability to predict when the possibilities will be realised: that requires extensions to the theory.

The importance for science of abilities to represent and explain how things could or might exist, i.e. the ability of a theory to **explain possibilities**, was discussed and illustrated in chapter 2 of Sloman(1978), where it was argued that the ability to formulate **predictive laws** is of secondary importance.

In the case of chemistry, further power was gained by extending the ontology to include *sub-atomic* particles, and their properties, interrelationships and interactions.

Below, I'll describe a hypothetical organism with a discrete 2-D sensory array, starting with an ontology of changing 2-D bit patterns, can gain explanatory power by extending its ontology (a) so as to allow for continuous lines moving continuously in 2-D, and (b) by adding a third dimension, allowing 3-D continuous structures to move in 3-D trajectories. In both cases, there are two stages, first enriching the ontology to allow for more possible structures and entities to exist, which cannot be sensed and whose descriptions require concepts that cannot be defined in terms of what can be sensed, and secondly using the ontology to formulate conjectures about structures and processes that cannot be sensed, but which allow predictions about sensory data to be made because the theory allows continuous 2-D projections of continuous 3-D structures and processes, and allows a discrete sensory array to sample 2-D continuous projections. This amounts to a two-stage projection, from 3-D to 2-D, and from continuous space to a discrete space.

Creating the more complex enriched ontology allows an organism or machine to use a powerful and (relatively) simple theory to explain complex changing 2-D sensory data.

This idea goes back to Kant(1781). Demonstrating that this is possible constitutes a refutation of "Symbol Grounding" theory (known to philosophers as "Concept Empiricism"), which claims that all concepts must be definable in terms of contents of experiences.

Kantian theory extension can be contrasted with many AI learning mechanisms that are trained by a teacher to attach labels to collections of 2-D sensory data, or which use statistical analysis of image data to identify and label recurring clusters of sensory data at increasing levels of abstraction without ever postulating the existence of entities that can exist without being sensed, as proposed by Kant. It is not yet clear to me whether the non-reductive (Kantian) ontology extension capabilities described below are found in any of the "Deep Learning" systems described in Schmidhuber (2014)

A challenge for learning theories in psychology, neuroscience, AI and robotics, will be presented, and implications discussed, starting with a visual learning example, based on changing contents of a 2-D rectangular array of bits (e.g. 0s and 1s, or any other pair of distinguishable items). This kind of learning can use either an exhaustive search through a space of possible explanatory theories, generated by an apriori theory about possible theories, or innate mechanisms, produced by natural selection, or a human designer, to drive the choice of proposed explanatory ontologies and theories on the basis of what has been found to work in previous generations, or previous designs.

The exhaustive search is more general, but also more likely to be intractable, and too slow for individuals that need to feed, grow, escape predators, and reproduce. On the other hand, the use of (Kantian) innate meta-theories can lead to a failure to find good explanations because of limitations of the meta-theories.

However, in organisms, natural selection can complement the use of innate mechanisms by individuals with a richer, more thorough, and very much slower, search across generations. And in some organisms that process of natural selection can lead to a process of social/cultural evolution that works much faster than natural selection can change genomes, though it requires individual organisms to have novel teaching and learning capabilities, including use of language and creative educational abilities (not found in all human teachers!).

NOTE: I am not claiming that animal visual systems use anything remotely like the form of input in the examples below: 2-D arrays of bit patterns. The structure of a typical biological retina is totally different, and that suggests that biological retinas perform quite different functions from electronic frame-grabbers. For now, those differences will be ignored. For more information about the human retina, including numbers of foveal and non-foveal receptors see:
http://webvision.med.utah.edu/book/part-xiii-facts-and-figures-concerning-the-human-retina/
E.g. Total number of cones in fovea: Approximately 200,000. There are 17,500 cones/degree^2. Approximately 17,500 cones in the central rod-free fovea.
Total number of cones in the retina. 6,400,000
Total number of rods in the retina. 110,000,000 to 125,000,000

Example: the power of ontology extension from 2D to 3D

Consider a computer program with access to a 450x450 2-D black and white display in which pixels are continually changing (in synchrony) where the task of the program is to find a way of predicting what is going to happen next. Note that this number of pixels is much smaller than the number of receptors in a human retina.

According to the above web site, a human fovea has approximately 200,000 cones, not far off 450x450 (i.e. 202,500), though retinal cells are not arranged in a square array, so relationships in a video image derived from cartesian coordinates, e.g. forming a horizontal vertical, or diagonal line, may have no analogue for retinal input. Moreover, whereas a computer can calculate the relationship of direction and distance between an image point and the centre of the image, the corresponding measure for an organism might be expressed in terms which saccade, or which head movements, or which limb movments could bring a point of interest to the centre of a fovea, or both foveas simultaneously (using vergence), and what positive or negative magnification effects those changes would have on image fragments.

Figure Lines below is an example of a subset of a snapshot of such a digital display (much smaller than 450x450):

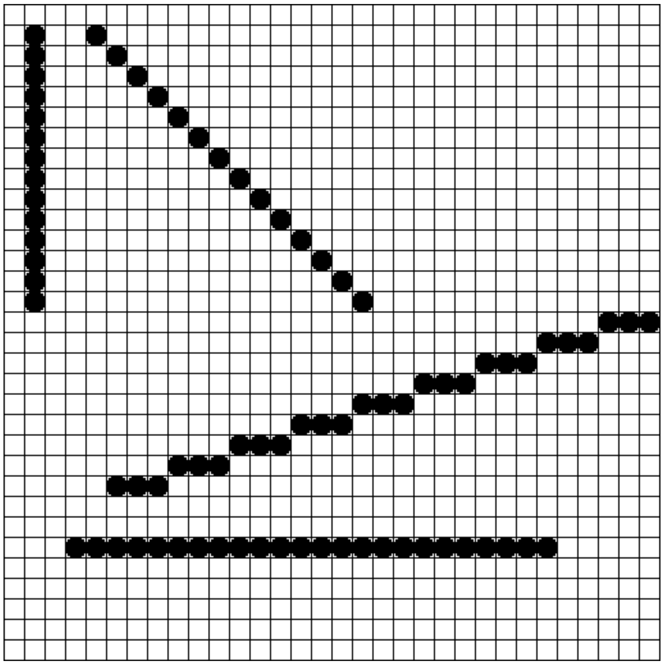


Figure Lines

At the next time-step that portion of the image could look the same, could look slightly different (with only a few pixels changing), or could look completely different, and likewise during successive time steps. Finding a way to describe the changes observed over some number of time steps in order to be able to predict future time-steps is, in general, an impossible task, since the pixel contents might be generated by some random process.

But if there is some structured mechanism, outside the perceiver, generating the changing images, the task has a solution, even if it is hard to find.

The program may have to search for a form of description summarising how the the pixel values, their 2-D coordinates and the time are related. E.g. it may attempt to construct some sort of law of the form:

$$C(x, y, t) = F(x, y, t)$$

where F may be a complex formula or algorithm possibly referring to the initial state at a particular time, t0, as well as more recent times, e.g. t-1, t-2, etc.

Alternatively, the program may attempt to find a continuous function or some collection of differential equations, treating the discrete values as a sample from a set of continuously varying black white and grey patterns distributed over the 2-D plane, with a thresholding operation determining whether each pixel is recorded as black (0) or white (1).

Or it might search for a generative model or simulation that depicts both an external structure and a projection process, where the model supports reasoning about how the projected images will change.

I’ll discuss some of the complexities of detecting and describing patterns of change in a 2-D rectangular array, and later show how the complexity can be reduced if the perceiver is able to generate a hypothesis about the changing

2-D image being a projection (shadow) of a rotating 3-D object, a wire frame cube.

Searching for the best way to predict (and explain) sensed changes

A machine with an attached fixed camera repeatedly generating a 450x450 2-D array of binary (0 or 1) values, might attempt to start searching for a pattern in the changing 2-D array as soon as it is switched on, or it might wait for a time, store a sequence of such 2-D patterns and then attempt data-mining in the recorded collections of pixel values at different locations at different times.

> ***Notes on complexity of the task***
> *If the pixel-states change 10 times a second, and information is collected for an hour, then the program would have 36000x450x450 (over 7 billion) records of the form [colour, x, y, time]. Looking for relationships between various subsets of the pixels at different times, possibly including relationships between patterns separated by several time steps, could require the machine to examine a very large set of subsets of the pixels, if it starts with no prior knowledge about which patterns are likely to exist.*
>
> *The number of subsets of pixels in a 450x450 array is far larger than the number of atoms in the universe, estimated as about 10^{80} in*
> *http://en.wikipedia.org/wiki/Observable_universe*
> *The number of subsets expressed as a decimal number has over 600,000 digits. So there is no possibility of searching through the set of possible subsets seeking patterns of association between images. Any such learning will have to be based on selections of much smaller image fragments. How can a learning system that knows nothing about the mechanism generating image sequence decide on good ways to select image fragments to compare across time-steps, or across a single image if seeking regular 2-D patterns?*

If all the changes are completely random there will be no way of simplifying all that information. But suppose it is not random: will a visual learning system have to make use of some innate biases, or can some totally general learning mechanisms find the underlying patterns -- the laws relating the pixel changes over time?

However general a learning mechanism is, it cannot check all possible hypotheses about the causes of its sensory input in parallel. It will have to order the hypotheses in some way. If the learner is a product of biological evolution on this planet, it may be predisposed to try hypotheses in an order that the ancestors of the learner found useful, if the order is encoded in the genome. But what worked well for ancestors may not work well in a new context, e.g. attempting to use a device connected to the internet to gain knowledge.

Obviously it will depend in part on what sort of pattern of changes is being displayed and whether the available feature detectors find features that are parts of important patterns of change.

Suppose that in the world of our hypothetical learner, all that happens is that the pixels are either all black or all white and they alternate at regular intervals.

That pattern of behaviour may appear to be fairly easy to detect. There will be only two states of the whole display, between which the display switches (all black and all white), and that would be obvious to humans). But some change-detectors may be incapable of recognising a global binary alternating process, especially if they start by trying to detect static or moving edges and building on the results, as many AI vision systems do!

But let’s consider more complex cases, including, later on, images caused by projecting the shadow of a rotating 3-D object, e.g. a wire frame cube, onto the grid. My conjecture is that there are many types of image sequence (or video stream) that include regularities whose discovery would require either strong initial ("innate") hypotheses as to their contents or else "astronomically" large searches through the space of possible explanatory mechanisms.

Searching by enumerating possible designs

One such search mechanism would enumerate all possible Turing machines capable of generating sequences of 450x450 bit patterns, and trying the machines in increasing order of size, using an "alphabetic" ordering for machines of the same size.

As the length L of the machine description increases, the number of descriptions of length L increases explosively. For example, the number of possible descriptions with 1000 ordered symbols with at least K options for each symbol, is K^{1000} -- far larger than the number of atoms in the universe if K > 1.
(This web site suggests that the number atoms is at most around 10^{80}:
http://en.wikipedia.org/wiki/Observable_universe#Matter_content_.E2.80.94_number_of_atoms)

This approach is intractable, even if steps are taken to reduce the explosion by eliminating redundant options (e.g. equivalent descriptions using different symbols). The approach may appear to work if tested on ’toy’ problems, however.

Although I have not yet understood it fully, a more sophisticated approach is proposed in the "Powerplay" system described by Juergen Schmidhuber, which not only searches for a good explanatory model to generate the sensory data, but simultaneously searches for good learning algorithms. I don’t know if it overcomes the combinatorial explosion outside of ’toy’ test situations where the only sensory input is a single bit stream. (Of course a single bit stream could represent a sequence of 450x450 binary retinal images, but with no prior information about the order in which the bits are fed into the stream, the complexity of the search task will be significantly increased.)

```
POWERPLAY: Training an Increasingly General Problem Solver by
 Continually Searching for the Simplest Still Unsolvable Problem.
Frontiers in Cognitive Science, 2013. ArXiv preprint (2011)
http://arxiv.org/abs/1112.5309
```

Special-purpose learners, tailored to the environment

An alternative is not to attempt an exhaustive combinatorial search, but to make use of some powerful assumptions about the nature of the environment, as discussed by John McCarthy in

```
    http://www-formal.stanford.edu/jmc/child/
    John McCarthy, "The Well-Designed Child", unpublished 1996
    Also published in the Artificial Intelligence Journal in 2008.

He states:
    "Evolution solved a different problem than that of starting a baby
    with no a priori assumptions."
```

Various kinds of non-randomness in the sensory input may be fairly easily detectable, if the machine starts off with an appropriate set of built in (or "innate") mechanisms, or if it has a learning system that is biased towards developing such mechanisms. For example, evolution seems to have biased many animal visual systems towards developing detectors for edge features of various scales in various orientations.

For example, if no pixel ever changes its colour, what would enable that constancy to be detected? If the system starts with built in change-detectors at every pixel, all feeding into another 2D array recording all the last recorded changes at each pixel, then a third layer with a mechanism that scans the second layer looking for any deviation from a constant value might repeatedly report that no change is being detected anywhere.

But detecting image constancy quickly would require pre-built mechanisms. If the system has to *learn* which algorithms operating on the input values produce useful results, it may take some time to learn to detect total constancy! The ’no change’ condition might be detected only after various tests for interesting types of change have failed.

[NOTE: Many people are puzzled by ’change blindness’ demonstrations of Kevin O’Reagan and others, and ask for explanations of why change is not detected. This is really silly: we need explanations of how change **is** detected, not why it sometimes is not detected.

Having found an explanatory mechanism (which requires an understanding of the computational problems), we can then ask under what conditions the mechanism might fail to produce the right result.]

Detecting global motion

Another relatively simple pattern of change would also be quickly detectable if all the required algorithms are pre-built, namely global motion from left to right -- i.e. in the direction of increasing x value, if each pixel has coordinates (x,y). A detector associated with each pixel that has a memory for a previous neighbouring pixel value could easily check whether every pixel value at location x, y at time t+1 is exactly the same as the pixel value x-1, y at time t-1, apart from the left-most pixel in each row (e.g. x=1).

Using those pre-built detectors for similarity between left and right neighbours across a time step, an additional mechanism could check whether then all the similarity detectors had a positive result at every time step. That would be a relatively simple mechanism to hard-wire into an electronic retina, though if it had to be discovered by searching through a space of algorithms that search could take some time.

A slightly more complex challenge would be recognition that the pixel value at location (1000,y) at time t is always the same as the pixel value at location (1,y) at time t+1: i.e. the horizontal motion to the right "wraps around" to the left, so that an unchanging pattern is constantly cycling through the system.

There are many other such patterns (e.g. horizontal flow in the opposite direction, or vertical flow upward or downward, with or without ’wrapping’, or combined horizontal and vertical flow (i.e. diagonal flow) in any one of four directions. More complex stepwise diagonal flows could have everything constantly moving two steps right then one step up. All of those ’global’ patterns could easily be detected if the designer of the system had pre-installed suitable checking algorithms (or hardware equivalents, in some cases, like the ’optical flow’ detectors that seem to be used in animal visual systems).

Your visual system would very quickly spot a simple movement across the display from left to right, though it’s not clear whether that’s because all humans learn to use such a motion detector or whether there is something in the genome that ensures its presence in normal brains.

Detecting that the pattern’s motion is ’wrapped’ round the vertical edges of the display, could be a result of noticing that some patterns keep repeating, starting at the edge and moving to the right. That will be easier to do if the dots do not form a random array but have some clearly visible large scale structure, e.g. a 10x10 array of large squares moving across the screen. In that case,detecting that as a square moves off the screen to the right, an exactly similar square moves onto the screen on the left, should be feasible if the search for process structures is designed to look for moving vertical edges, instead of searching among all possible patterns of pixel combinations.

In such cases the formulation of a ’theory’ that describes what is going on and allows predictions to be made about what will happen next, can use the same set of concepts as was required for the initial data (i.e. pixel locations and their contents), plus some additional concepts defined in terms of the concepts used to define the data (e.g. 50x50 array of pixels, 10x10 array of 50x50 pixel arrays, etc.).

I do not know whether Schmidhuber’s ’Powerplay’ system referenced above would easily discover this sort of description in a 1000x1000 array.

(There are minor complications about defining the concept of a complete display moving horizontally, that need not be discussed here.)

Extending the ontology to include continuous lines

Not all patterns of change will be so simple. For example consider an array of pixels which are mostly white, but contain what we would see as a number of black lines at various orientations, moving linearly across the display in different directions. A snap-shot might look like Figure Lines above.

In that case the program would do better if it were able to extend its ontology to include the concept of a continuous 2-D line **projected** onto the 2-D discrete array. Each such line would then be represented approximately by a nearly co-linear set of black pixels. The description of a line moving continuously across the screen will be much simpler than the description of a jagged collection of squares moving, and possibly slightly changing it’s shape because of the different adjustments required to fit a portion of the image exactly into a square.

A learning program that from "birth" includes the notion of a "line-segment" as a movable entity that can be manifested or represented by a changing set of pixels in a display, might be able to detect indicators of such lines and discover **that** they move and **how** they move. Without a suitable set of innate concepts, searching among all possible configurations of pixel patterns for useful invariants across time intervals could be at least extremely slow and possibly also completely intractable.

NOTE ADDED: 29 Jul 2011
Social evolution and cultural transmission could change this: if structures found to be useful by members of a community are not encoded in the genome but recorded in the culture and passed on to young learners to constrain their searches for useful features. That form of guidance is one of the factors that enables each generation to learn more than previous generations, as discussed below in connection with the influence of a teacher.

With appropriate initial concepts available, the program might find "maximal lines", i.e. lines that are not parts of larger lines, by scanning outwards from linear-fragments and merging adjoining nearly collinear fragments, as is typically done by computer vision programs.

(The concept of a continuous line segment, with arbitrary orientation, moving continuously in continuous space in an arbitrary direction, while producing a projection in a discrete 2-D array is non-trivial but I shall not expand on requirements for possession of such a concept here. That concept certainly cannot be **defined** in terms of experiences in a changing discrete grid.
But from the concept in an appropriate theory it is possible to **derive** criteria for detecting the projection of such a continuously moving line in a 2-D discrete grid, and human vision researchers have explored a variety of such line detectors. Since they all work on rectangular grids of pixel values, whereas animal retinas do not have that structure, quite different detectors will be needed by anyone trying to model or understand natural vision systems.)

Could a totally general learning process, without relying on any "innate" concepts for building explanatory theories, discover this way of explaining the sensed patterns, e.g. using learning based only on mechanisms for information compression, without any built-in biases in favour of particular ontologies or forms of representation to use for compression -- and without any initial bias towards using increased dimensionality to achieve reduction in complexity? Or would some sort of innate disposition to use the concept of a continuous straight line be required to make the learning feasible in the lifetime of an organism?

Things get more complex if the lines can change their 2-D orientation.

Suppose the display includes changing collections of black pixels that can be taken as evidence for a small number of continuous lines projected onto the display, including some lines that are neither horizontal nor vertical, each moving linearly without rotating or bending, then describing the lines could produce a considerable reduction in the complexity of the perceived process, compared with a full description of the changing pixel values.

The extra complexity of the latter description would arise out of the need to continually specify which pixels are black and which white (0, or 1). There is no unique way to do this: it will depend on the assumed thickness of the line relative to the pixel size, and the orientation of the line relative to the two major axes of the pixel grid.

Pixel projections of four such continuous lines are shown in Figure Lines above. Notice that the relationship between length of line and number of black pixels required to draw it depends on orientation, as shown by the vertical and diagonal lines of different lengths, but with the same number of pixels. The concept of a line is not the same as the concept of a set of black points, though the latter can be taken as providing information about (i.e. representing) the former.

So, instead of having to predict behaviours of a million discrete pixels changing colour in synchrony, such a program can use a richer ontology providing a way of predicting behaviours of a relatively small number of continuous lines moving continuously, but **sampled** discretely at discrete times. In this case the concept of a continuous line and the concept of continuous motion are not something given as part of the domain of the original sensory data, but creative extensions of that original ontology.

For more complex examples, including multiple layers of representation, using several different ontologies, see the description of POPEYE, the image interpretation program in Chapter 9 of Sloman(1978) http://www.cs.bham.ac.uk/research/projects/cogaff/crp/chap9.html

The pictures interpreted by POPEYE depicted words made of cut-out capital letters some overlapping others, with additional positive and negative noise, as in the examples in the Figure POPEYE below and Figure Noise, below.

Figure POPEYE

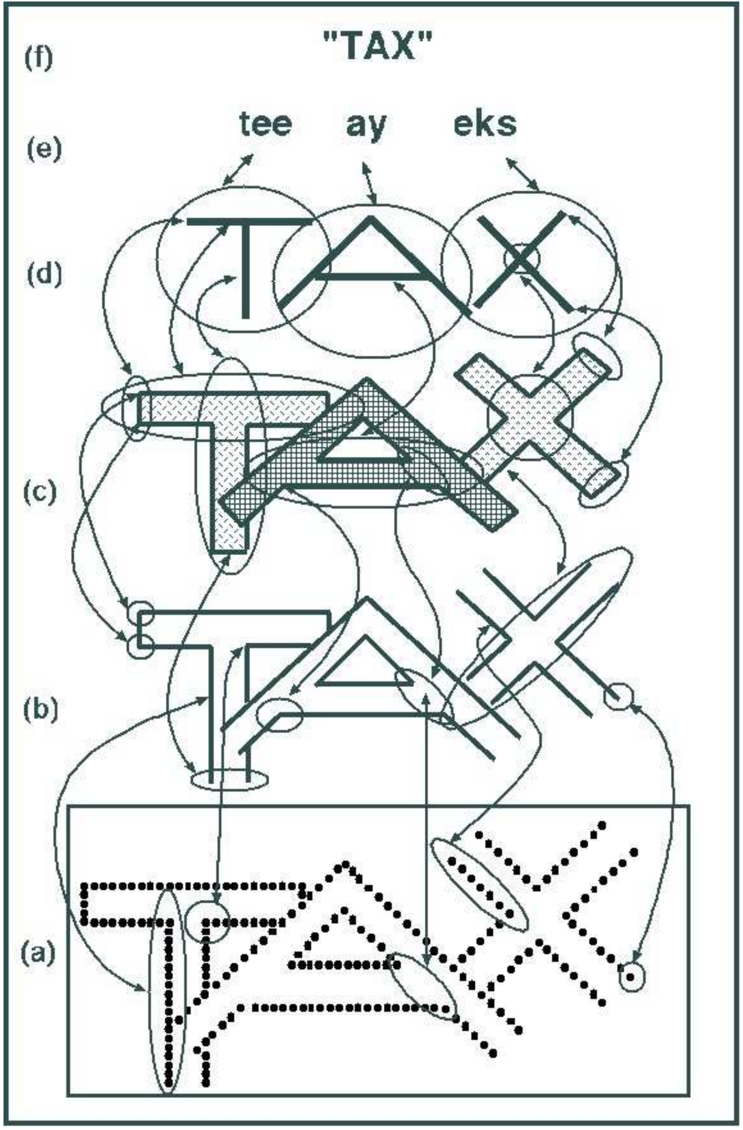
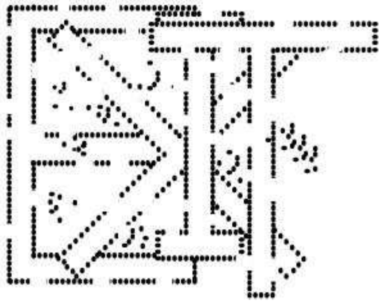


Figure Noise



The program was able, in many cases, to recognise the word before all the picture fragments had been found because the interpretation proceeded in parallel at all the levels shown in Figure POPEYE. Gaps or noise at a particular level level could be compensated for by information acquired at other levels, using bottom up or top down or middle-out inferences.

This program did not do any learning. Adding the ability to invent the ontologies required for all the intermediate layers would have been a challenging task, but the project was terminated for lack of funding. Doing the learning would have required the program to discover the need for the intermediate layers when presented with increasingly complex images. Then given a new image, like Figure Noise, it should be able to generate a suitable interpretation autonomously, by combining separately learnt ontologies.

Adding rotations

When normal adult humans are presented with displays of moving linear configurations of fixed lengths their visual systems naturally interpret the display in terms of 2-D objects moving continuously in a plane surface, despite the discreteness of the display.

However, Johansson and others (see below) have shown that under some conditions, if the lengths of lines change while their orientation in the plane changes this may be seen as 3-D motion of an object of fixed length. For example a 2-D line segment rotating about an end that is fixed while the other end moves on an ellipse with centre at the fixed end, will often be seen as a line of fixed length rotating in a plane that is not parallel to the display plane. That interpretation requires a 3-D ontology and the ability to interpret a sensed 2-D process as a projection of a 3-D process.

In Johansson’s demonstrations, more complex moving patterns, with lines changing their orientations, their lengths and the angles at which they meet, are often interpreted as moving non-rigid 3-D objects, made of rigid fixed-length components linked at joints, possibly with motions characteristic of living organisms, e.g. walking.

A 2-D line-segment is a four dimensional entity insofar as four different items of information are needed to specify each line. They could be two cartesian co-ordinates for each end, or a pair of co-ordinates for one end plus a length and a direction to the other end (polar coordinates for the second end), or a pair of polar co-ordinates for the first end plus a length and direction (polar co-ordinates) for the second end. It requires a substantial ontological extension to switch to representing 3-D line segments, which need six items of information to identify them. However the switch is much more than merely a matter of increasing the size of a vector: the set of relations, structures and processes that can occur in a 3-D space is very much richer, including projections of structures and processes from 3-D to 2-D.

Algebraic representations

A more algebraic form of representation for the line could take the form of an algebraic expression involving some variables, representing a class of lines, plus some numbers to select an instance from that class. Depending on the algebraic expression used we might be dealing with more than four dimensions, e.g. if not only straight lines are considered. The space of algebraic expressions that could be used to characterise subsets of a 2-D space would not have any well-defined dimensionality, since the structures of algebraic expressions can vary infinitely in complexity. But let’s ignore that for now.

> The concept of a continuous (Euclidean) line moving continuously could not be explicitly defined in terms of the appearance of its projection into the discrete array. So in that sense the concept of continuity cannot be **grounded** in the sensory-motor information available to a machine of the sort described.
>
> The notion of "grounding" is a source of confusion for cognitive scientists and philosophers, as argued here:
> http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#models

Although the ontology extensions described above can simplify the description of a complex pixel array (e.g. "It’s the word ’EXIT’"), it complicates the ind of universe the program needs to contemplate. The advantage is considerable simplification of the description of the changing sensory (pixel) data originally given.

Further simplifications

Later, the program may find that the pattern is repeated every 1000 cycles, so all it has to do is describe the trajectories of each line during 1000 steps.

It may find that a good way to do that is break the trajectory of each line into sections during which its motion is in roughly the same direction and then find some algebraic formula that approximates the motion in each section. (Animal brains would not use algebraic formulae, but more likely qualitative descriptions of the forms of motion, e.g. going left, decelerating, changing direction, going right, etc.)

If the machine is able to look for patterns that are not in the original data, but in its derived descriptions, it may then discover that there are groups of lines that share the same motion patterns, allowing further simplification. For example several of the lines might change their direction of motion simultaneously -- moving from left to right then from right to left, with similar accelerations and decelerations, while also altering their vertical locations in the picture so that their ends move smoothly, in roughly elliptical paths (depicted of course by jagged discrete sequences of black pixels in the display).

So, instead of considering a million pixels of which many but not all change colour at each time-step, it can consider 12 lines each of which has a small number of continuous trajectories, where the lines can be grouped perhaps into three sets of lines with similar types of trajectory in each set.

By now, the ontology used by the machine has been enriched with continuous trajectories of lines, directions and speeds of movement, and accelerations, of lines and ends of lines. On that basis the machine may define a concept of change of direction, and identify times at which such direction-changes occur for different lines or line-ends. This will enable it to represent groups of lines and groups of line-ends with similar patterns of movement and change of movement.

The form of representation found could then be used, if suitable mechanisms are available, to predict what will happen over various future time-intervals. Depending on the ontology used the predictions may be precise and metrical or imprecise and qualitative.

In principle, this form of representation could also explain some sensed pixel patterns where a group of black dots shrinks in size as the group approaches an edge of the display, then later starts expanding. An economical description of such a process might be a line partly moving beyond the end of the pixel array, and then moving back.

It would even be possible to find evidence for some lines moving in a circular pattern, disappearing completely and then reappearing at a different part of the edge of the display.

[It would be good to have little videos illustrating these possibilities. Offers gratefully received. Other videos are referenced later.]

NOTE (Added 5 Nov 2011)
John McCarthy has a web page making a similar point -- except that he uses a rather obscure puzzle that humans don't all find easy -- to make the point that there's a difference between appearance and reality. See
http://www-formal.stanford.edu/jmc/appearance.html

The benefits of dimensionality extension

A great deal of research on machine perception and machine learning has been concerned with techniques for reducing dimensionality of information, e.g. by projecting high-dimensional data into lower dimensional spaces and searching for patterns in the reduced data instead of the original.

I have been discussing different techniques above, namely moving to a different space from the original data-space, where the different space may be richer (e.g. continuous instead of discrete) but easier to reason about.

A really clever learning system (unlike any so far produced in AI that I know of) might go even further and invent the notion of 3-D space containing rigid structures that can move and rotate in that space, as described above: but that would require something more than a completely general learning system.

For example, the learner might start off with the knowledge that, instead of having only 2-D spatial coordinates, simple bits of stuff can have 3-D coordinates, and instead of motions involving changes in 2-D they can be 3-D changes, including changes of distance from the perceiver, and also changes of orientation and direction of motion in space, if the objects rotate.

In that case, a learning system presented with the data described above may be able, in some cases to achieve a further simplification of its description of what is going on by describing it as a rotating 3-D wire-frame cube (for example) projected onto the 2-D pixel display, like a shadow projected onto a translucent screen.

There are some examples of online demonstrations of 2-D projections of 3-D rotating cubes here, along with further discussion of requirements for being able to make this discovery:
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/nature-nurture-cube.html
If a 3-D wire-frame cube is rotating about a fixed axis passing through it then its twelve edges will project onto a pixel display as twelve moving groups of

black pixels of the sorts described above. Each approximately linear group will move in a manner that depends on the size of the corresponding cube edge and its distance from and orientation relative to the axis of rotation of the cube. In terms of changing black and white pixel patterns the projection will be quite complex to describe and the behaviour of the pixels hard to predict. But if the sensed patterns are conjectured as to be shadows (2-D projections) of a 3-D rotating wife-frame cube then a single changing angle of rotation can be used to explain/predict all the sensed projected data: a very great simplification based on considerable ontological sophistication.

All the sensed processes can be summarised by an initial state of the cube, and an angular velocity for the rotation, plus the current time. For each time the 3-D configuration can be computed (including the 3-D linear velocities of all components) and the 2-D projection derived. The encoding of that specification of an unending sequence of pixel displays could be much smaller even than the explicit encoding of a single state of the display.

Note that in this case if part of the rotating shadow does not fall within the bounds of the pixel display the theory that assumes the edges continue to exist, whether their shadows are sensed or not, will allow reappearance of the projections to be predicted and explained.

My conjecture is that humans and many other animals are innately provided with mechanisms that attempt to interpret visual, haptic and auditory percepts in terms of an ontology of 3-D mobile entities some of which are other humans. How that process works and how it evolved are topics for further research, as is the problem of getting machines to learn and perceive in a similar way.

Clearly the animals that walk, suckle and run with the herd almost immediately after birth don’t have time to learn to see and respond to the complex 3-D structures and processes they cope with. So evolution can provide very powerful biases (as McCarthy noted).

I am not claiming that such highly specialised perceptual mechanisms are always present at birth: biological evolution has produced some species whose specific competences develop through interaction with environment, though the development is constrained by and partly driven by genetic influences, as McCarthy suggested in The Well-Designed Child (mentioned above)

The best known arguments for innate knowledge are concerned with human language learning, which is not matched by any other species on earth. Here, it is not the particular language learnt that is innately specified but something more general that can learn a very wide variety of languages.

I suggest there are far more examples of innate generic competences that can be instantiated in many ways as a result of interaction with a specific environment after birth, most of which have not yet been noticed.
Similar ideas are in Karmiloff-Smith’s outstanding survey of the issues **Beyond Modularity** (1992).

Some sketchy ideas about genetically influenced, staggered/layered, developmental processes are presented in Chappell & Sloman (2007)
The ideas I have been presenting can be taken as a development of Kant’s idea that in addition to concepts of things as they are experienced, an individual perceiving and acting in a world that exists independently of that individual’s percepts and actions would have to have a notion of a "thing-in-itself" ("ding an sich") whose existence has nothing to do with the existence of any perceiver. In more modern terminology we can express the conjecture that biological evolution produced some organisms that have innate dispositions to create concepts that are a-modal (not necessarily directly tied to any sensory or motor modality) and exosomatic (refer to things outside the skin of the

organism).

A conjecture about the evolution of generalised languages required for internal purposes by pre-verbal humans and also by many non-human animals interacting intelligently with a complex world, which might have developed later into a language for communication is presented here:
http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#glang

Does adding a teacher help?

The example could be elaborated by postulating the existence of a teacher who somehow poses problems for the learner to solve and provides positive and negative rewards, or comments, on the basis of evaluating the learner’s responses to the problems.

In that case we would have a learning system that is a combination of teacher and learner and the prior knowledge of the teacher used in setting questions and providing answers would be part of the total learning mechanism.

In the case of many animals, and also much of what goes on in very young humans there is a lot of learning that goes on without any teaching. In fact that must have happened in human evolution before there were adults with enough knowledge to do explicit teaching. So we need to explain what sorts of mechanisms, and what sorts of prior knowledge (including meta-knowledge of various kinds) are capable of generating different sorts of learning.

It’s a mistake to look for the one **right** learning system if we want scientific understanding as opposed to (or in addition to) mere engineering success.

(We already know how to build wonderful learning systems -- human babies: it doesn’t follow that we understand how they learn.)

Note on theory revision

A theory is an information structure (often a set of sentences in some language, along with techniques for manipulating and using them, and in some cases additional mathematical, logical, and diagrammatic forms of representation suited to the theory -- e.g. maps in geology) that usefully summarises a large number of empirical observations (and possibly also previous theories) but goes beyond the observations and can be used for a number of purposes:

- Explaining observed phenomena
  typically by showing how they could have been predicted if missing information had been available. The missing information is part of the explanation: the theory provides the rest. This can include explaining why something did not happen, e.g. why an action did not achieve its goal.

- Predicting what is going to happen
  on the basis of what has already been observed (plus known theories).

- Counterfactual and conditional prediction:
  Predicting what will happen, or will become possible, or will not happen IF something changes in the present situation.

- Future conditional prediction
  Predicting what would happen, or would become possible, or would not happen IF something were to change in a possible future situation (which may or may not occur).

- Past conditional prediction
  Predicting/retrodicting what would have happened, or would have become possible, or would have happened IF something had changed in a previous situation.

More or less complex variants of these abilities are also important for humans and other animals learning to interact with a complex environment, including understanding why some actions (done by themselves or by others) succeed and why others fail, and learning to find ways of improving their ways of planning, reasoning, and acting.

The theory can include undefined symbols, expressing new concepts that are part of the theory, and which get their meaning from their role in the theory. The system is usually associated with methods and mechanisms of observation, measurement, experiment, manipulation which play the role of "theory tethering" as explained in
http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#talk14
Getting Meaning Off The Ground:
Symbol Grounding Vs Symbol Attachment/Tethering

These additions do not define the terms of the theory: they can change while the theory endures -- e.g. adopting a new more accurate method of measuring the charge on an electron does not redefine ’electron’ or ’charge’, if most of the structure of the theory is unchanged, including the roles of those two symbols in the theory.

Often AI theories of learning do not take account of all of these applications of theories that are results of learning.

Revision of a theory can be motivated by either dissatisfaction with the structure/complexity of existing theories or problems of explaining or otherwise accommodating or reconciling the theory with new empirical information (or new theories).

The process of theory revision can include any or all of:

- Ontology extension, which usually requires addition of new undefined symbols in the theory, or modification of a generative system of symbols e.g. a grammar.

- Ontology modification, e.g. reorganisation of existing concepts, and the corresponding notations, e.g. subdivision of existing concepts into new cases or re-allocation of some concepts to different parts of the theory (e.g. whales are mammals, no longer thought of as big fish)

- Revision/extension of the propositions/formulae using the symbols of the theory so as to enable new predictions, explanations, and considerations of possibilities.

- Revision or extension of the modes of reasoning, including use of new types of mathematics, new diagrams or other forms of representation, or new computer models.

In some cases, the ability to make full use of new symbols (or more generally, new forms of representation) and new ways of generating propositions, ontologies, formulae, experiments, or searches for new data, may require changes in the physical mechanisms available, e.g. new sensors, stronger or smaller, or remotely controlled, manipulators, and also changes in the user’s information-processing architecture. For example if the changes are occurring inside the head of a learner, the learner may need to develop new ways of combining information from different sources, including different sensors, different kinds of memory, and new ways of processing the combined information (e.g. searching in a space of possibilities to find good explanations or predictions).

[Many of these changes have been experienced by the Artificial Intelligence community as a whole. Even more complex changes are going on within biology.]

All this amounts to an unusual way of looking at the process often labelled "Abduction", namely going from old information to new information that does not use deductive or statistical reasoning.

Moreover, the possibility of changing the undefined symbols of the theory makes the search space for such abductive processes potentially intractable. So heuristic constraints can on the search will be required. Those constraints may come from a culture, from individual preferences or hunches, or, in the case of some processes of individual development, from the genome, as implied in Chappell & Sloman (2007).

I.e. some of the abductions done by humans and other organisms use constraints provided by evolution, often in very complex and unobvious ways. The constraints are not all explicit at birth, but emerge at various stages through interactions between the genome and the environment Chappell & Sloman (2007).

Towards an implementation of these ideas

So far I have said nothing about how one might build a machine that has visual perception mechanisms that use retinal input as a basis for seeing the world. I shall offer some sketchy ideas which are close to ideas that others have proposed and some have implemented in the past, though nowadays it is not clear to me that such designs are being used.

The key idea is to abandon any notion that seeing happens either at the retina or in the lowest level processing mechanisms driven by retinal input (such as area V1 in the primate visual cortex). Instead the retina and processing elements that are retinotopically mapped should be thought of as together forming a peripheral sensor device for **sampling** what J.J.Gibson referred to as "the optic array": a cone of information streaming into the location where the eye is, from all the surfaces that are in view from the eye’s location. Only a subset of that information will enter the eye at any time, depending on the direction of gaze. In animals the sampling of the optic array is non-uniform, with a small area of high resolution sampling (the fovea) surrounded by lower resolution areas. For now we can ignore the variation in resolution and just talk about a retina that can be directed at different subsets of the cone of incoming information, to pick up samples. Some of the sampled information may be used instantaneously, while others will mainly be used to extend information structures built up over extended periods of time, of varying lengths.

This retina requires a large collection of processing units to find important information fragments in the optic array, including fragments of ’edges’, texture fragments, optical flow fragments, evidence for highlights, and many others. These fragments are automatically categorised, and where appropriate grouped into slightly larger fragments (where grouping decisions may be context sensitive), and the results of that processing are fed to various other subsystems that have different uses for the information, e.g. collision avoidance, posture control, detection of faces and other objects, detection of various processes in retinal patterns, description of various structures and processes in the environment, fine grained control of action (e.g. monitoring grasping processes), constructions of dynamically changing histograms that are useful for coarse-grained categorisation of the current scene, and also building up longer term records of what has been seen, where things are, what they are doing etc.

The longer term information will include things that are temporarily out of sight because the sampling has been moved to a different part of the optic array, or out of sight because they have been temporarily occluded by something closer to the viewer.

All the various information structures need to be kept available for use in various tasks (including controlling actions, avoiding bumping into things, answering questions, finding lost objects, following moving objects, catching things, making plans for future actions, etc.). Bringing items back into use will require mechanisms for re-instating links with the retinal array as needed, after such links have been removed because another region of the optic array is being sampled, and the information fed into another part of the more enduring information about the environment.

Added 18 Jul 2014
An implication of the above discussion is that learning (under the influence of both the genome and the environment, as described in Chappell & Sloman (2007) produces not just stored facts and new algorithms, but also a changed architecture, possibly including new subsystems with both acquired factual information and new control information, and new communication channels linking them.

The new architectures in some cases, instead of requiring new components and new physical connections may make use of new virtual machines composed of interacting sub-systems that are also virtual machines, as described in http://www.cs.bham.ac.uk/research/projects/cogaff/misc/vm-functionalism.html

NOTE:
some of the ideas described here are closely related to the retinoid mechanism proposed in this book by Arnold Trehub, online here:

> The Cognitive Brain, MIT Press, 1991,
> http://www.people.umass.edu/trehub/
> One feature that the author did not intend when he constructed his model was that it should explain why the retinal blind spot does not enter into consciousness as some sort of information gap. This follows from the fact that the blind spot is just an aspect of a sampling device feeding information into a more integrated and enduring information structure whose contents are more closely related the contents of introspection. The retinoid will maintain records of information received, not records of information not received.
>
> However I do not believe the details of the retinoid model are adequate to meet the requirements of all aspects of human and animal intelligence. That is a topic for another time.

Other examples

1. The example of the rotating 3-D wire-frame cube projected onto a 2-D retina can be varied in a number of ways, including, for example, allowing the axis of rotation of the cube to rotate e.g. in the surface of a cone, allowing the cube to expand or contract, or pulsate in size, or change its location. All of these will complicate the patterns of 2-D motion of points in the projection into the discrete retina described above.

   If some of the changes, e.g. orientation of the axis of rotation, velocity of rotation, size change, are under the control of output devices managed by the learning machine, that may make it easier for the explanatory theory to be developed, by partitioning the learning task into various sub-tasks.

2. **P-geometry:** Euclid discovered a feature common to all triangles: the interior angles sum to a straight line. The normal proof uses Euclid’s parallel axiom, and axioms about equality of angles formed when a straight line (a transversal) crosses two parallel lines. A former sussex student who became a mathematics teacher, Mary Pardoe, found a proof that was easier to remember, but very different. It involves aligning an arrow with one side then rotating the arrow around each of the vertices in turn, through the internal angles, aligning it with the next side. After being rotated through each angle the arrow ends up on the original side pointing in the opposite direction. Euclid’s axioms, and very many proofs based on them are products of human learning, which is clearly triggered by exploring structures and processes in space, but must make use of competences that are somehow products of the human genome, though they are not available at birth.

   I have been attempting to construct a new axiomatisation of what may be an extension of a subset of Euclidean geometry, which does not explicitly assume the parallel axiom, but does involve types of motion (translation and rotation) of line segments, initially just in a plane. I call this P-geometry (Pardoe-geometry) and have an incomplete discussion paper here:
   http://www.cs.bham.ac.uk/research/projects/cogaff/misc/p-geometry.html
   This was motivated by a desire to make explicit the assumptions of Pardoe’s proof and to test a conjecture that its assumptions do not include the parallel axiom, though its ontology does include motion, unlike Euclid’s axioms (though motion may be implicit in some of the theorems, e.g. about loci of sets of points satisfying a constraint).

   The process of searching for such an axiomatisation, which would, like Euclid’s axiomatisation, enormously compress a vast amount of information about spatial structures and also processes in the case of P-geometry, does not feel like a process using a totally general information-compression engine: rather it depends heavily on the specialised ability to represent, manipulate, and reason about spatial structures in an abstract way that does not depend on precise locations, sizes, orientations, etc.

   It would not be at all surprising to discover that there are evolved features of the human genome that support the development of such mathematical abilities, and without them humans might be unable to learn what they do learn in a normal lifetime. (Compare the features of the human genome that seem to be required to support language learning.)

3. Another example, mentioned to me by Alexandre Borovik in conversation is a small ball moving in a path the shape of a cylindrical spiral coil. Depending on the angle of view (or projection) the 2-D path displayed on the retina will be appear to have very different appearances, in some of which there are discontinuities not present in others, even though the motion of the ball is always continuous, with no discontinuous changes of direction. Invoking a 3-D structure producing these visible paths produces a simple uniform explanation of a lot of messy and complex 2-D trajectories.

   A similar comment can be made about a 3-D wire coil in the shape of such a cylindrical path. Its 2-D projections will be very different from different angles (including a 2-D circle as one of the special cases, and a zigzag linear shape as another) despite the common simple 3-D structure projected.

4. **Note on the History of Mathematics:**
   It has often happened in the history of mathematics that puzzles arising in some domain (e.g. natural numbers [1,2,3,4...], integers, real numbers) can be dealt with more simply by embedding that domain in a richer, more complex domain.

   Examples include adding negative numbers and 0 to the natural numbers, adding fractions (rational numbers) to the line of positive and negative integers, adding so-called irrational and transcendental numbers to the rational numbers to produce the so-called real numbers, and adding imaginary numbers (square root of -1) to the real numbers. For an excellent discussion of this listen to the episode of ’In our time’, chaired by Melvyn Bragg, broadcast on 23 Sep 2010
   http://www.bbc.co.uk/programmes/b00tt6b2
   http://www.bbc.co.uk/radio4/features/in-our-time/

5. The work of Gunnar Johansson on perception of moving points of light provides many additional examples. He produced movies made by attaching lights to joints on humans filmed in the dark, walking, dancing, fighting, doing push-ups, climbing a ladder, and performing other actions. In each case where a still snapshot was seen merely as an inexplicable collection of points, the movies were all instantly perceived as 3-D movements of one or more humans. Similar effects were produce with light points attached to simulated 3-D biological organisms of various morphologies moving. There were additional experiments involving just two points that could be seen either as ends of a rotating rod or as moving in various 2-D patterns. Generally the simplest 3-D interpretation was preferred.
   http://www.questia.com/library/book/perceiving-events-and-objects-by-gunnar-johansson-sten-sture-bergstrom-willi
   For more details see
   Perceiving Events and Objects
   by Gunnar Johansson, Sten Sture Bergström, William Epstein, Gunnar Jansson

Lawrence Erlbaum Associates, 1994

Note on motivation and learning

It is often assumed that all motivation must be related to some sort of reward that can be experienced by the individual that has the motivation.

This assumption underestimates the power of biological evolution, which is capable of producing many kinds of reflex response. Some of them are externally visible behavioural responses to situations that can cause damage -- e.g. blinking reflexes and withdrawal reflexes. Many such reflexes work without the individual having any anticipation of reward to be obtained or punishment to be avoided, even though the response may have been selected by evolution because it tends to enhance long term reproductive success. Individual animals do not need to know that having a damaged eye can be a serious disadvantage in order to have reflex behaviours that avoid damage.

I have argued in this paper:
    http://www.cs.bham.ac.uk/research/projects/cogaff/09.html#907
    Architecture-Based Motivation vs Reward-Based Motivation,

that in addition to external behavioural reflexes there can be, and are, internal reflexes that produce not behaviours but powerful motives to achieve or avoid some state, and the mere existence of such a motive can, in many situations, trigger planning processes and action process tending to fulfil the motive. These may be as biologically beneficial as external behavioural reflexes but far more flexible because they allow the precise behaviour to achieve the newly triggered motive to be a result of learning.

I suspect the irresistible urge to find proofs in mathematics, to improve elegance or efficiency of computer programs, to find a unified explanation of a range of observed phenomena in science, and to produce works of art all depend primarily on architecture-based motivation.

A learning system that has just one long term goal, namely to compress as much as possible of information received, might have only one architecture-based motive that drives all others.

TO BE EXPANDED: COMMENTS, CRITICISM, SUGGESTIONS WELCOME

The ideas proposed here are intended not to form a definitive explanatory theory, but to be part of a long term "progressive" research programme, of the type defined by Imre Lakatos, in

    Falsification and the methodology of scientific research programmes,
    Philosophical papers, Vol I, Eds. J. Worrall and G. Currie,
    Cambridge University Press, 1980, pp. 8--101, CUP

See also his Open University Broadcast:
    Science and Pseudoscience (1973)
    http://www.lse.ac.uk/collections/lakatos/scienceAndPseudoscience.htm

Further reading

- Jackie Chappell and Aaron Sloman, 2007,
  Natural and artificial meta-configured altricial information-processing systems,
  International Journal of Unconventional Computing 3, 3, pp. 211--239,
  http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0609,

- G. Johansson, (1973)
  Visual perception of biological motion and a model for its analysis,
  *Perception and Psychophysics,*vol 14, pp. 201--211.

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

- Juergen Schmidhuber (2013)
  PowerPlay: training an increasingly general problem solver by continually searching for the simplest still unsolvable problem
  Frontiers in Psychology, Cognitive Science June 2013, Vol 4, Article 313
  http://arxiv.org/abs/1112.5309

- Juergen Schmidhuber, (2014)
  Deep Learning in Neural Networks: An Overview,
  Technical Report IDSIA-03-14,
  http://arxiv.org/abs/1404.7828

- Arnold Trehub,
  Evolution’s Gift: Subjectivity and the Phenomenal World,
  Journal of Cosmology, In Press, 2011,
  http://journalofcosmology.com/Consciousness130.html,

- A. Sloman et. al.
  The Cognition and Affect Project, papers and presentations:
  http://www.cs.bham.ac.uk/research/projects/cogaff/
  http://www.cs.bham.ac.uk/research/projects/cogaff/talks/

- **Interactions between philosophy and AI: The role of intuition and non-logical reasoning in intelligence,**
  Proc 2nd IJCAI, 1971, London, pp. 209--226,
  (Reprinted in AI Journal 1971 and in chapter 7 of Sloman(1978))
  http://www.cs.bham.ac.uk/research/cogaff/04.html#200407

- Aaron Sloman,
  CRP: **The Computer Revolution in Philosophy: Philosophy, Science and Models of Mind,**
  Harvester Press (and Humanities Press), 1978,
  http://www.cs.bham.ac.uk/research/cogaff/62-80.html#crp

Maintained by Aaron Sloman
School of Computer Science
The University of Birmingham