



Simplicity and Ontologies

The trade-off between simplicity of theories and sophistication of ontologies

Aaron Sloman

DRAFT: COMMENTS WELCOME

(This is work-in-progress and is likely to remain so for some time. Help needed.)

Last updated: 22 Sep 2010; 23 Sep 2010; 26 Sep 2010; 12 Dec 2010; 7 Apr 2011; 21 Apr 2011; 5 Nov 2011

7 Jul 2011; 29 Jul 2011; 19 Sep 2011

Installed: 21 Sep 2010

This file 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>

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 the 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 which are not explicitly defined as abbreviations for something expressible using other

components.

The atomic components will 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. (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, but about contextually identified object-parts (e.g. surface fragments relevant to some manipulation task), relations, processes, opportunities, constraints, goals, values, theories, or explanations, for example.)

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

Main Thesis:

Adding levels of abstraction that are not **definable** in terms of previous levels 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.

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

This will be illustrated by the example of invoking inaccessible 3-D structures and processes to explain sensed 2-D structures and processes (changing pixels on a retina). The more complex 3-D **ontology** allows a relatively simple **theory** to explain the sensory data.

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) that usefully summarises a large number of empirical observations (and possibly also previous theories) 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.

The theory can be 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
- ontology modification, e.g. reorganisation of existing concepts, and the corresponding notations
- revision of the propositions/formulae using the symbols of the theory so as to enable new predictions, explanations, and considerations of possibilities.

This is a way of looking at the process often labelled "Abduction". The possibility of changing the undefined symbols of the theory makes the search space for abduction potentially totally intractable. So some heuristic constraints can be very valuable.

This paper suggests that the abductions done by humans and other organisms use constraints provided by evolution, often in very complex and unobvious ways.

Examples of the power of ontology extension - in a 2D domain

If a computer program is given access to a 1000x1000 2-D black and white display in which pixels are continually changing (in synchrony) and the task of the program is to find a way of predicting what is going to happen next, it may struggle to find some pattern in the relationships between the colours of the pixels, their 2-D coordinates and the time. E.g. it may attempt to derive some sort of law of the form:

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

where F may be a complex formula or algorithm and may refer to the initial state at a particular time, t_0 , 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 the continuously varying values.

The task of finding such a formula could involve data-mining in a large collection of records of colours of pixels at different locations at different times.

E.g. if the pixel-states change 10 times a second and information is collected

for an hour, then the program would have 36000x1000x1000 (36 American billion) records of the form [colour, x, y, time]. Looking for relationships between subsets of the pixels at different times, possibly including relationships between patterns separated by several time steps, could require examining very much larger sets of subsets of the pixels. The number of subsets of pixels in a 1000x1000 array is considerably larger than the number of electrons in the universe.

If all the changes are completely random there will be no way of simplifying all that information.

The usefulness of innate biases (e.g. produced by evolution)

Various kinds of non-randomness may be fairly easily detectable, if the learning system is biased towards looking for them.

For example, if no pixel ever changes its colour, or if each colour at location x, y at time t is the same as the previous colour of the pixel at the left, i.e. the colour at x-1, y, t-1 (with 'wrapping' of the value of x at edges), then such a pattern might be fairly easily detected. Certainly your visual system would very quickly spot a simple movement across the display from left to right, though it may be harder to detect that the pattern's motion is 'wrapped' round the vertical edges of the display.

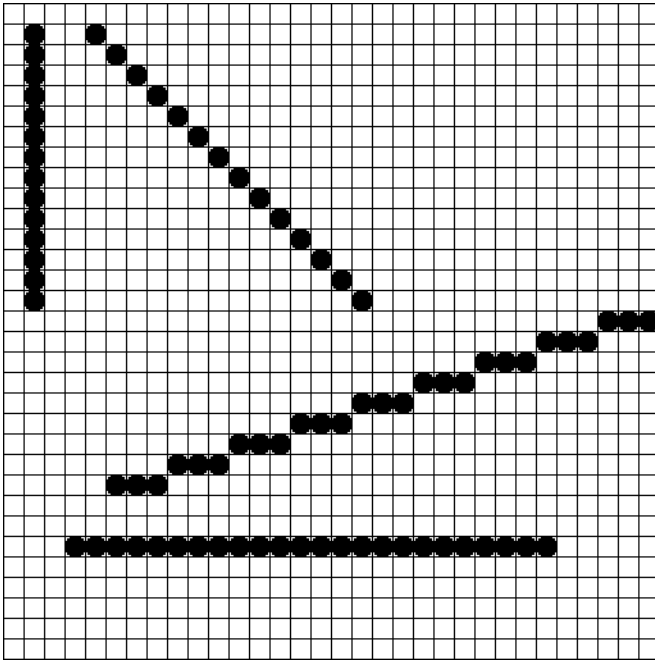
It will be easier if the dots do not form a random array but have some clearly visible large scale structure, e.g. a 10x10 array of large squares moving across the screen. In that case, detecting that as a square moves off the screen to the right, an exactly similar square moves onto the screen on the 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.).

(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 this:



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.

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 invariants across time intervals could be 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. However for humans that may be coming to an end, as suggested in The Singularity of Cognitive Catchup (SOCC):
<http://www.cs.bham.ac.uk/research/projects/cogaff/misc/another-singularity.html>

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. It certainly cannot be **defined** in terms of experiences in a changing discrete grid.)

Could a totally general learning process, discover this way of explaining the sensed patterns, e.g. 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 -- including increasing dimensionality to achieve reduction in complexity?

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 small number of such lines, including some 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 pixel projections of four such lines are shown in the picture above. Notice that the relationship between length of line and number of black pixels 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. For more complex examples, including multiple layers of representation, using several different ontologies, see chapter 9 of The Computer Revolution in Philosophy (1978)

<http://www.cs.bham.ac.uk/research/projects/cogaff/crp/chap9.html>

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 much smaller 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.

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 will be seen as 3-D motion of an object of fixed length. For example a line segment rotating about an end that is fixed while the other end moves on an ellipse with centre at the fixed end, will 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.

There are many reasons why 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/coaff/talks/#models>

Although that creative piece of ontology extension complicates the kind of universe the program is contemplating, it can considerably simplify the description of the changing sensory (pixel) data originally given, as explained in more detail below.

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 used 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.ezproxyd.bham.ac.uk/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 wire-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

<http://www-formal.stanford.edu/jmc/child/>
John McCarthy, "The Well-Designed Child"
Also published in the AI Journal 2008.

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 this invited journal article:

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

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/coqaff/talks/#qlang>

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.)

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.

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

The Cognitive Brain, MIT Press, 1991,
<http://www.people.umass.edu/trehub/>

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. 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 to the contents of introspection. The retinoid will maintain records of information received, not records of information not received.

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 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 produced 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-william-epstein>

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

<http://www.cs.bham.ac.uk/research/projects/cogaff/>
The Cognition and Affect Project
<http://www.cs.bham.ac.uk/research/projects/cogaff/talks/>
Online presentations since about 2000

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

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

Maintained by [Aaron Sloman](#)
[School of Computer Science](#)
[The University of Birmingham](#)