

The biological bases of mathematical competences: a challenge for AGI

Invited talk for The Fourth Conference on Artificial General Intelligence
Google, Mountain View, California, USA, August 3-6 (Wed-Sat) 2011
<http://agi-conf.org/2011/>

This conference was held shortly before the AAI 2011 Conference in San Francisco

<http://www.aaai.org/Conferences/AAAI/aaai11.php>

Where I presented a closely related tutorial on Philosophy and AI, on Monday 8th Aug. 2-6 pm:

<http://tiny.cc/AIphiltut> expanding on some of the ideas below.

Downloadable video of presentation

<http://www.cs.bham.ac.uk/research/projects/cogaff/movies/#agi-2011>

Youtube version: <http://www.youtube.com/watch?v=58DdZ8EkOik> (Talks by Sloman and Boyden).

This file is <http://www.cs.bham.ac.uk/research/projects/cogaff/misc/agi-2011-talk-sloman.html>

Abbreviated link: <http://tinyurl.com/aaronagi11>

A messy PDF version (may be out of date) will be generated occasionally using html2ps and ps2pdf

<http://www.cs.bham.ac.uk/research/projects/cogaff/misc/agi-2011-talk-sloman.pdf>

Aaron Sloman

School of Computer Science, University of Birmingham.

Last updated: 14 Feb 2011; 30 Apr 2011; 8 Jul 2011; 26 Dec 2011

Installed: 14 Feb 2011

Abstract

Evolution produced many species whose members are pre-programmed with almost all the competences and knowledge they will ever need. Others *appear* to start with very little and learn what they need, but appearances can deceive. I conjecture that evolution produced powerful innate meta-knowledge about a class of environments containing 3-D structures and processes involving materials of many kinds. In humans and several other species these innate learning mechanisms seem initially to use exploration techniques to capture a variety of useful generalisations after which there is a "phase transition" in which learnt generalisations are displaced by a new generative architecture that allows novel situations and problems to be dealt with by reasoning -- a pre-cursor to explicit mathematical theorem proving in topology, geometry, arithmetic, and kinematics. This process seems to occur in some non-human animals and in pre-verbal human toddlers, but is clearest in the switch from pattern-based to syntax-based

language use. The discovery of non-linguistic toddler theorems has largely gone unnoticed, though Piaget investigated some of the phenomena, and creative problem solving in some other animals also provides clues. A later evolutionary development seems to have enabled humans to cope with domains that involve both regularities and exceptions, explaining "U-shaped" language learning. Only humans appear to be able to develop meta-meta-competences needed for teaching learnt "theorems" and their proofs. I'll sketch a speculative theory, present examples, and propose a research programme, reducing the 'G' in AGI, while promising increased power in return.

Extended Abstract

FACT:

Human children seem first to learn to talk using lots of learnt verbal patterns with which they get along quite well.

Then they usually change, and apparently acquire a syntax-based competence that is far more powerful because it can generate entirely new utterances and enable them to understand novel utterances (in combination with compositional semantics).

But when that transition occurs, some of their learnt patterns are **wrongly** over-ridden by the new mechanism so they say "He hitted me", "I runned home", "She caught the ball", whereas previously they would have said "He hit me" etc.

No amount of parental explanation, rebuking, repetition helps to correct the error, at that stage.

After a while, children spontaneously change and start coping with the exceptions alongside the grammatical rules. (e.g. "ran" not "runned").

I assume that takes time because extending the newly constructed rule-based architecture to cope with exceptions is a non-trivial change (not easy even for a programmer to implement) whereas a purely pattern-based learning system doesn't have rules that can have exceptions -- there are just lots of learnt associations with different priorities.

CONJECTURE:

The first two parts of that process (learning re-usable patterns/associations, then replacing them with something more axiomatic and generative) occurs in many NON-linguistic competences in many species (humans, monkeys, cats, squirrels, crows, ...).

WHY?

Because, as Kenneth Craik noted in 1943, it's a requirement for coping with a world in which not all dangers can be discovered by trial and error (e.g. because error = death or serious injury) and where it is advantageous to be able to *work out* good things to do in novel situations instead of *always* having to use associative (trial and error) learning, which typically takes much longer, unless the learner is improbably lucky (or rigidly steered by a trainer.)

That is what I think is the biological basis of mathematical competence.

The vast majority of organisms, if they can learn at all, merely adapt by modifying parameters. A smaller subset of organisms can do statistical learning: acquiring new empirical generalisations that can be used either for forming expectations or for selecting or avoiding actions on the basis of what has previously worked or failed. This form of learning uses domain-neutral mechanisms,

which, up to a point are very useful, and if combined with a suitably abstract ontology, can allow generalisations to be learnt that go beyond the evidence used, e.g. by interpolating or extrapolating (not too much) to new types of instance.

But a tiny subset of species, including humans and many other primates, squirrels, corvids, elephants, cetaceans (e.g. whales, dolphins), and octopuses seem to be able to switch from being able to apply empirical generalisations to be able to *work things out*. That requires an architectural change, to a system that makes use of forms of representation with generative powers and compositional semantics -- though not necessarily systems that look like human languages.

As explained in <http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#glang>

Evolution of minds and languages.

What evolved first and develops first in children: Languages for communicating, or languages for thinking?

(Generalised Languages: GLs)

This requires a genome that is able to express itself in stages, possibly delaying some later stages until enough material has been acquired by the exercises of earlier competence to provide the basis for a major reorganisation without too much error that would require later correction.

That's important because correcting erroneous associations is a simpler matter than correcting some of the deeper (axiom-like) components of a generative system, because of the amount of compression that will have gone into that system, making it difficult to unravel errors without breaking too much.

Of course there have been many proposals for learning systems that depend on compression (e.g. by [Juergen Schmidhuber](#) among others).

However, I am not saying, as they do, that this is a *general purpose* compression mechanism. Rather, as hinted by John McCarthy, it may have been specially tailored to what can usefully be hypothesised in environments of particular sorts (e.g. physical processes of various kinds in a 3-D space). At this stage I don't know how those mechanisms work.

Compare John McCarthy, in "[The Well Designed Child](#)" (Also in AIJ 2008):

Evolution solved a different problem than that of starting a baby with no a priori assumptions.

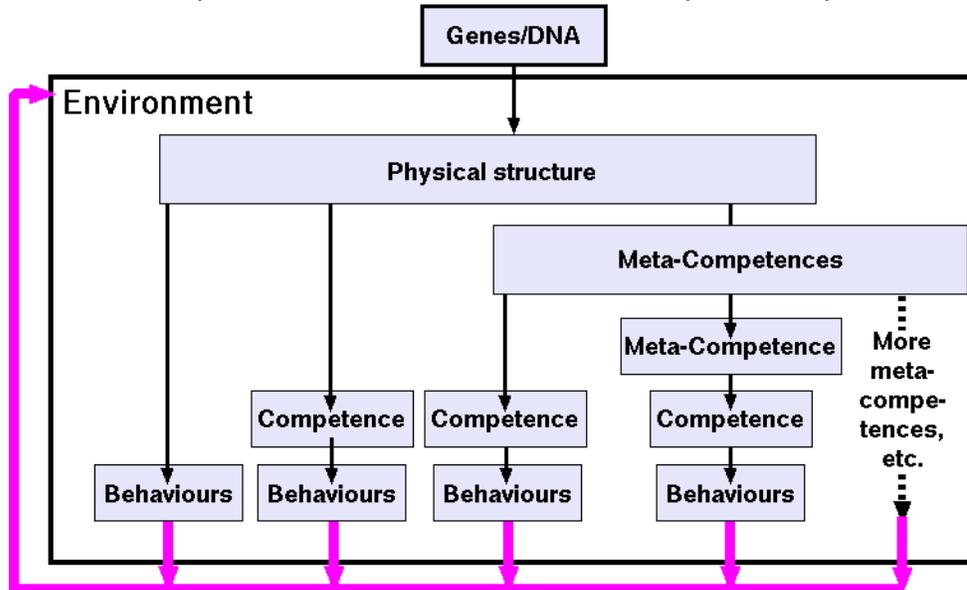
.....

Instead of building babies as Cartesian philosophers taking nothing but their sensations for granted, evolution produced babies with innate prejudices that correspond to facts about the world and babies' positions in it. Learning starts from these prejudices. What is the world like, and what are these instinctive prejudices?

Jackie Chappell and I have tried to characterise a system that partly fits this description but also extends its learning capabilities on the basis of what it has previously learnt, as summarised (roughly) in this diagram:

Multiple routes from genome to behaviours

(Environment affects all embedded processes)



I later discovered that some closely related ideas had been expressed by Annette Karmiloff-Smith in

Beyond Modularity: A Developmental Perspective on Cognitive Science
MIT Press 1992

<http://www.cs.bham.ac.uk/research/projects/cogaff/misc/beyond-modularity.html>

CONJECTURE

The "U-shaped" language learning phenomena observed in human children are a special subset of a more general, largely unnoticed, form of learning that happens in pre-verbal children and other species. But the more general mechanism shared with other species does not in all cases involve the additional step of modifying the new architecture to cope with large numbers of exceptions, and although it allows novel goal-directed behaviours of varying complexity to be produced, is not specifically concerned with communicative behaviours.

(It may be that production of bird song makes use of a similar mechanism, but its communicative function, if there is any, does not seem to have semantic specificity of human language.)

[Relevance to later evolution of communicative behaviour is discussed here:

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

CONJECTURE

In most animals, and in young children, when the transition happens from using only learnt generalisations to development of a new generative mechanism capable of working out solutions to novel problems, or predictions in novel situations, the individuals have no idea that it is happening, any more than language learners do.

However, young children can express a partial awareness of differences between empirical generalisations and derivable "theorems" by using words and phrases like "must", "has to", "cannot", and by expressing contempt or very confident denial when others propose contrary predictions or solutions to problems.

(Piaget's last book, on Necessity reports investigations charting the gradual and erratic growth of such competence.)

CONJECTURE

In the case of humans, various further stages can occur (though whether they do or do occur not may depend both on individual differences and also cultural differences), e.g.: some individuals can explicitly (self-consciously) begin to talk about and communicate to others what they have learnt.

Some of them may go further and develop publicly communicable formalisations of the content, as happened a long time ago with Euclidean geometry, then more recently with other branches of mathematics, mechanics, chemistry, theoretical physics, logic, linguistics, computer science, and other fields.

This may be one effect of a more general development of meta-semantic competences and further architectural growth that allows self-monitoring of certain kinds of thought processes. For example, it is sometimes useful, when a plan fails, to be able to think back over the steps taken previously, looking for a step that might have been different, leading to a different conclusion.

[Compare Sussman's HACKER: circa 1974]

Of course, that requires not only the architectural change allowing self-monitoring and subsequent recall, but also development of a theory of what the planning/reasoning process is and how it can work well or go wrong.

I suspect no other species can do this (though Irene Pepperberg's Parrot Alex sometimes appeared to).

Alas, I have conclude, on the basis of many years teaching experience, that not all humans can do it, or do it equally well -- at least not in all domains.

Because of this capability it is also possible for some individuals to communicate to others what they have discovered, not just about the world, but also about their forms of reasoning. Those individuals may also be able to form meta-cognitive hypotheses about the forms of reasoning (and planning) used by other individuals and offer them advice on how to improve their thinking, just as they can give themselves advice.

[Later, social processes can lead to all this knowledge being organised systematically and taught systematically, eventually leading to development of courses and books on mathematics, and departments of mathematics in learned institutions, though debates about the nature of the activity can rage unchecked by good theories.]

A subset of the explicit meta-knowledge will be concerned with reasoning about spatial structures and processes, which when formalised can produce theories of topology and geometry.

I have a challenge for **general** learning theorists here:

<http://www.cs.bham.ac.uk/research/projects/cogaff/misc/simplicity-ontology.html>

Simplicity and Ontologies:

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

Other subsets of explicit meta-knowledge can be concerned with more general and abstract structures and processes, such as one to one mappings and operations of various kinds involving sequences and that can lead to teaching and learning of arithmetic.

[Perceptual subitizing is a red herring and has attracted too much attention.]

Other subsets of the new meta-knowledge can occur when what has been previously learnt is applied both to itself and to new kinds of problem. E.g. developing general notions of length, area, volume as quantities that can systematically be mapped onto numbers in ways that have many practical uses.

This can lead to applied mathematics including physics, engineering, architecture, etc.

It can also lead to *unjustified* over-generalisation that is later corrected, e.g. by theories of measurement, and the development of integral calculus.

Some of this agrees with Karmiloff-Smith's ideas in *Beyond Modularity*, while some of this goes further in postulating a wider variety of types of mechanism required to support the various domain-specific forms of learning.

I also think her emphasis on **Representational** change is too narrow: there are also architectural changes, changes in ontology, and changes in mechanisms (algorithms).

UNDERSTANDING THE PROCESSES

We don't know much about the processes, and mechanisms, conjectured above, or how the genome encodes and produces them, nor what can be done to help or hinder the processes.

I suggest progress can be made by collecting many examples of domains in which not only children but also adults can explore collections of structures and processes discovering various previously unfamiliar generalisations and constraints empirically at first then later coming to see how some of those discoveries are actually "theorems" about a portion of the world that allows a systematic, generative characterisation.

(E.g. there's a domain of 2-D structures involving rubber bands and pins that is both unfamiliar to most people and easily explored.)

Some of those discoveries will then be recognized as "toddler theorems" discovered by many young children, as shown by their problem solving behaviour, but mostly unnoticed because nobody has been looking.

(There may be many examples buried among the junk typed into systems like the Open Mind project -- though I have not looked closely.)

There are probably examples in other intelligent species including corvids, orangutans, squirrels, elephants and some domestic pets.

[Betty the hook-making New Caledonian Crow studied in Oxford around 2001-2005 raises many unanswered questions.]

I have not yet seen anything like this in robots, though that may be because either people have not been trying (like all the roboticists who emphasise embodied cognition as based on statistical learning) or they have trying in a manner that is far too general, instead of looking for learning mechanisms that might have evolved in particular types of learners in particular types of

environment.

I have several examples in talks on this subject here:

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

Why (and how) did biological evolution produce mathematicians?

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

Piaget (and collaborators) on Possibility and Necessity

And the relevance of/to AI/Robotics/mathematics (in biological evolution and development)

A test example is here:

<http://www.cs.bham.ac.uk/research/projects/cogaff/misc/simplicity-ontology.htm>

I suspect the time is ripe for a collaborative effort to explore a variety of domains in which such learning can occur and attempt to develop systems (robots, or simulated explorers that learn) that can play and build up empirical knowledge then after a while trigger mechanisms that produce the reorganisation postulated above.

Some of the harder domains will include not only spatial structures but also properties of different kinds of material, as discussed in this discussion of requirements for learning about kinds of "stuff"

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

(Betty's understanding of how hooks can be made of wire in several different ways is unexplained. Her species did not evolve in junk-yards.)

CONJECTURE

The processes that come after the collection of re-usable patterns, associations, and generalisations discovered empirically, and which produce novel explanatory theories, can include:

- Creation of new syntax
I.e. new forms of representation with new sorts of complexity and variability. The use of Fregean syntax (based on application of functions to arguments) was probably a late development in biological evolution.
- Creation of new forms of composition (of more complex from less complex structures and processes)
Physics and chemistry provide a huge variety of forms of changeable, manipulable structure. In order to be usable for information processing, the structures must be inspectable, comparable, searchable, and also manipulable -- by other information processing mechanisms.

I suspect that human thinking about forms of representation and types of syntactic manipulation is too much constrained by what we have explicitly learnt to do externally (e.g. in communicating, or keeping records, making maps, etc.) on the basis of which we make guesses about what might be going on inside animals. We need a much richer repertoire of guesses to explore.

- Creation of novel ontologies (new concepts not definable in terms of pre-existing ones)
For example, the transition from pattern-based language use to syntax-based use requires development of an ontology of syntactic types, relations, and transformations (e.g. combinations). But there's also ontological creativity required for new semantic contents, referring to things previously incapable of being sensed, or thought about, e.g. invisible properties of matter, arrangements of structures and processes whose invariant features are not definable in terms of old concepts, types of non-physical, non-perceptual abstract entity such as infinite structures, sequences, and processes and limits of infinite sequences processes. (Perhaps only human ontologies, and not all human ontologies, include this.)

- Architectural extensions required to support new representations, new types of concurrent processing, new forms of meta-cognition, hypothetical reasoning, ...

All this strongly contradicts AI researchers, philosophers, psychologists, and neuroscientists who think ALL learning is statistical and ALL biological inference mechanisms are probabilistic. (I don't know what proportion of current researchers think that.)

It will also be seen to contradict philosophical theories of concept empiricism, or "symbol grounding" that require all concepts used to be definable in terms of patterns in sensory-motor signals (or experiences). (Mid-20th century philosophy of science showed such theories to be false, though Kant had previously provided strong counter-arguments *circa* 1781.)
<http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#grounding>

I suspect that if we get to the stage where our young robots are clearly reorganising what they have learnt into a generative system in which they make mathematical discoveries ("toddler theorems"), that will substantiate Kant's philosophy of mathematics against its main rivals. That's what first got me interested in AI about 40 years ago.

CONJECTURE: Motivation

It is often assumed that motivation needs to be based on expectation of reward (positive or negative). However, it can be argued that biological evolution produced mechanisms that generate motives in young learners (as a form of affective reflex) before they are capable of having any knowledge about the benefits of achieving those motives

Evolution may have discovered that very young learners with those motives, and mechanisms that (normally) react to the creation of such motives by attempting to achieve them, learn things that they otherwise would not learn.

I call this **Architecture-based** as opposed to **Reward-based** motivation. For more details see:
<http://www.cs.bham.ac.uk/research/projects/cogaff/09.html#907>

Architecture-Based Motivation vs Reward-Based Motivation

The key idea is that alongside innate physical reflexes (like blinking) and learnt physical reflexes required for walking, running, jumping, sporting and musical expertise, there are also innate and learnt cognitive (internal) reflexes that under certain conditions trigger changes in information structures or in modes of information processing.

A special type of cognitive reflex is generation of a motive: the innate motives are generated by perceived situations and opportunities, or by success and failure in various actions. Acquired motive-generating reflexes are of many sorts (including possibly addictions).

One of the ways in which biological evolution, or a robot designer, can prepare a species or a type of machine for learning and developing in a particular sort of environment is by providing a collection of innate motive-generating reflexes. (E.g. if something comes into view set up the motive of grabbing it. If some physical process is perceived, set up the motive of reversing it, and many, many, more.) In addition there seem to be innate mechanisms for acquiring new motive generating reflexes, which as a result of processes of learning and development are capable of triggering new motives. This is far more flexible than triggering physical actions directly, since the actions that work in one situation may fail in another, whereas generating a motive (or goal) that in turn generates processes of collecting information, planning, selecting actions, executing actions, can be successful in far more situations, including novel situations.

This is related to but different from the idea presented in this paper, namely that there are innate reward mechanisms selected by evolution:

S. Singh, R. L. Lewis, A. G. Barto, Where Do Rewards Come From?, in
Proc. 31st Annual Conference of the Cognitive Science Society,
Editors, N.A. Taatgen and H. van Rijn, pp. 2601--2606, 2009
<http://csjarchive.cogsci.rpi.edu/Proceedings/2009/papers/590/paper590.pdf>

Singularity of cognitive catchup

The amount learnt by successive generations of humans has increased over centuries, and this has helped to make it possible for some individuals to achieve major advances over previous work.

That process may be coming to an end

<http://www.cs.bham.ac.uk/research/projects/cogaff/misc/another-singularity.html>

The Singularity of Cognitive Catchup (SOCC)

Added: 8 Jul 2011

The work of Annette Karmiloff-Smith

I recently realised that Karmiloff-Smith's 1992 book **Beyond Modularity: A Developmental Perspective on Cognitive Science** presents a deep and important collection of ideas, closely related to what I have been investigating, but offering a new and better way to organise the evidence and theories. I have begun to summarise the connections in this draft document:

<http://www.cs.bham.ac.uk/research/projects/cogaff/misc/beyond-modularity.html>

TO BE REVISED AND EXTENDED

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