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

Invited pre-recorded talk

Sharif Spring School on AI
Philosophy, Ethics, and Society

http://www.en.sharif.edu/

Artificial and Natural Intelligence
Nature and Philosophical Debates

Especially:

"How can a physical universe produce mathematical minds?
And why are they so hard to replicate in current AI systems?"

April 2019

Aaron Sloman
http://www.cs.bham.ac.uk/~axs
Honorary professor of AI and Cognitive Science
School of Computer Science
University of Birmingham UK

(Retired but still doing full time research linking AI, Cognitive Science, Philosophy, Biology/Evolution, and aspects of several other fields, e.g. Psychology, Neuroscience, Education, Computer Science, Physics and Chemistry.)

This invited talk at the Sharif Spring School April-May 2019, is available in two formats:
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/sharif-talk.html
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/sharif-talk.pdf

This is linked to and used in a recorded 45 minute video presentation available on youtube, in this collection:
https://www.youtube.com/playlist?list=PLYC-dSilAaYa6Mk1g6hBGUyqCwrIvyOWB
Direct link: https://www.youtube.com/watch?v=mmW044dlujA
The video deals with only a subset of the material in this web page. It was unscripted and the presentation is somewhat disorganised. Parts of this web page were used in the talk, though this page was extended after the recording was made.

A closely related talk was given a few weeks earlier in Zurich on 11th March 2019:
How can a physical universe produce mathematical minds?
And why are they so hard to replicate in current AI systems?
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/zurich-talk.html

Related to this: a video recording of an unscripted 20 minute talk about squirrel, toddler, and adult human spatial reasoning given to a conference of artists in November 2018, is available here
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/sloman-eastside-2018.html

---

Contents (below)

ABSTRACT

Added: 22 Apr 2019 (Not in video)
Expanded: 7 May 2019

This talk offers an unusual attempt to bring together a collection of topics usually studied separately, in different disciplines, namely: attempts to understand human minds and brains (e.g. philosophy, psychology and neuroscience); research on human and non-human spatial reasoning competences; how biological spatial reasoning relates to ancient mathematical discoveries in geometry; Immanuel Kant's characterisation of such mathematical discoveries as synthetic (not merely definitional), non-empirical (i.e. not derived from experience), though based on mechanisms that Kant said were "awakened by experience", and essentially concerned with necessary truths or impossibilities.

In particular, Kant noticed deep connections between causal cognition and mathematical cognition, especially in ancient mathematical discoveries in geometry and topology. These include causes of impossibility, and necessity.

E.g. how does the structure of two interlocking rings made of impermeable material cause impossibility of separating them? What causes the impossibility of two straight lines bounding a finite area on a plane surface (one of Kant's examples), or the impossibility of three planes bounding a finite volume? If you slice a vertex off a convex polyhedron with a planar cut, that causes changes to the number of vertices, faces and edges of the remaining polyhedron. What changes? How do you know?

Such ancient kinds of mathematical cognition are connected with intelligence of pre-verbal toddlers such as this 17.5 month child reasoning about 3D topology (4.5 minute video, also shown near the beginning of the recorded lecture):
http://www.cs.bham.ac.uk/research/projects/cogaff/movies/ijcai-17/small-pencil-vid.webm
I have a large collection of examples, some involving other intelligent animals, e.g. squirrels, elephants, crows, weaver-birds, as well as ancient mathematicians and their precursors...

Instead of postulating that animals or robots have to start with a combination of sensory mechanisms and general purpose learning mechanisms, I suggest that evolution discovered different kinds of increasingly abstract and increasingly powerful forms of representation with associated learning and reasoning mechanisms, not all of which are active from birth in intelligent species. Instead "meta-configured" genomes produce different, increasingly abstract, competences each of which builds on and extends results of previously developed forms of learning.

This allows the same genome to support cognitive development in very different environments. In humans this was also used in evolution of multi-layered linguistic competences and mathematical competences. Both humans and some other intelligent species develop spatial intelligence using multi-layered learning. In humans there are additional layers that allow discovery that some of what is learnt involves impossibilities and necessary connections. I suggest that currently popular theories and mechanisms proposed in AI, psychology and neuroscience are inadequate and need to be modified to take account of the powers of a meta-configured (multi-layered) genome, otherwise AI systems will continue to lack human linguistic competences, language-based thinking competences, and spatial and mathematical reasoning competences required for intelligent

behaviour. So they will not be able to replicate or understand the discoveries made by ancient mathematicians, e.g. Archimedes, Euclid, Zeno and many others. Neither will they replicate the intelligence of pre-verbal human toddlers. However, full replication of human intelligence will face the enormous challenge of replicating the computational powers of sub-neural chemical mechanisms, combining discrete and continuous forms of information processing, unlike digital computers. That will not happen soon.

CONTENTS

Background

This web page is associated with a recorded video presentation mentioned above, making use of parts of this page. Some parts will be skipped in the presentation, but viewers can come back here to read them.

I'll start with some general comments on the aims of AI and Philosophy and then describe some deep features of natural intelligence, in humans, including very young humans, and also other intelligent animals, that have not been replicated in AI systems, or explained in neural and psychological theories.

Most AI researchers, psychologists and neuroscientists don't even notice that they need to be explained, although the philosopher, Immanuel Kant, drew attention to some of them in 1781, namely an important variety of abilities to perceive, reason about, and make use of spatial structures and relationships, including detecting that some tasks are impossible, and that some states and processes have necessary consequences. These competences are based on deep, generally unrecognised, aspects of spatial perception that are closely related to mathematical cognition.

Later I'll compare Kant's ideas with ideas of David Hume. Much research in psychology, neuroscience and AI unwittingly follows Hume's ideas and as a result the theories and AI models proposed by most researchers have the flaws Kant pointed out.

> **Introductory challenge:**
> Try to think of examples of spatial structures or processes that you can describe, or depict, but which are impossible. How do you know they are impossible? Try also to think of features of a situation that necessarily imply the presence of other features. Again, how do you recognise that necessary connection.
> If you can't think of any examples of these two types, I have a messy collection here:
> http://www.cs.bham.ac.uk/research/projects/cogaff/misc/impossible.html

These abilities to detect impossibilities and necessary connections are closely related to the deep discoveries regarding spatial structures and relations (geometry and topology) by ancient mathematicians collected in Euclid's *Elements*, available here:
http://www.gutenberg.org/ebooks/21076.

That used to be taught in all good schools when I was a child, but Euclidean geometry is now often ignored by educators, partly because of the mistaken view that the "proper" way to teach mathematics, including geometry, is to use logic-based axioms and reasoning methods. I'll return to those later. I'll also point out their limitations. A result of this educational disaster is that many people are blind to deep limitations of current AI, psychology and neuroscience.

Alternative Topics

There are many other aspects of AI that I could have chosen for this talk, including the study of complete architectures for human-like minds and other possible minds. A complete architecture can explain not only central cognitive capabilities, such as perceiving, learning, making and using mathematical discoveries, planning, predicting, explaining and communicating, but also affective mechanisms, states and processes, including wanting, enjoying, disliking, fearing, being curious about, being proud of or about something, approving, disapproving, loving, hating, being infatuated, feeling guilty or ashamed, and many more.

Those topics are closely connected with the study of varieties of complete information-processing architectures that are possible for humans, other animals, and intelligent machines, why they evolved, how they develop, and why they are important.

In contrast, the information-processing capabilities of a single-celled organism that lives in a chemical soup, or in your digestive system, are much simpler than those required for a crawling or flying insect. Insects have much simpler information processing mechanisms than vertebrate species such as birds, hunting mammals, elephants, apes and dolphins.

But very simple organisms were precursors of more complex biological organisms, and studying limitations of the simple ones may give us a better understanding of the benefits of additional information processing capabilities.

More complex architectures may include mechanisms for dealing with factual information, sensory evidence, motive selection plan formation, control of actions, cooperating or competing with others, and also growing in size and also information concerned with mating and reproduction.

For fairly obvious reasons research in AI and robotics has not attempted to address all those problems -- although there are now people selling robots to serve as sex mates. Use of a search engine will reveal examples. But their responses are all fakes. This talk is about whether we can build machines that really do perceive, reason, choose what to do and make mathematical discoveries.

My answer will be that there are still deep gaps in the achievements of those aspects of AI, especially as concerns perceiving and interacting with our physical environment, which you might think is one of the easiest problems for robot designers to solve.

A web site collecting our past work on such topics, in the CogAff (Cognition and Affect) project is in preparation here, though still incomplete and disorganised.
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/emotions-affect.html
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/emotions-affect.pdf

There is also a great deal of such work done by other researchers, including researchers in AI. Many of them have not had the broad cross-disciplinary education required for such research and they often therefore produce very shallow theories about motivation, emotions, and other forms of affect. In fact they often ignore major categories, such as long term grief or long term passionate support for a political movement, or a long term love of mathematics.

A subset of relevant references to such work can be found here:
https://en.wikipedia.org/wiki/Affective_computing

Bad theories of motivation

A particular problem is that many theorists in philosophy, psychology, neuroscience, social science, economics and AI believe that all motivation must be based on expectation of positive or negative rewards following choices (expected rewards). A paper arguing that that theory is shallow and of limited explanatory power can be found here:
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/architecture-based-motivation.html

It proposes that biological evolution produced an additional powerful form of motivation that has nothing to do with expected rewards, but which plays important roles in biological evolution: Architecture-based motivation.

When a child sees something new and wants to pick it up, or kick it or put it in its mouth, it doesn't necessary have any expectation that it will derive some benefit from the action. Instead evolution has produced "internal reflexes" of various kinds that can be triggered by things perceived or thought about, and when activated they produce a new motive.

Organisms with that kind of mechanism may acquire a great deal of information that somewhat later turns out to be useful in ways that they could not possibly hypothesise in advance.

A fairly obvious example is babies being triggered to try to mimic sounds made by other individuals. The child cannot possibly know that that activity will develop abilities that are later useful in speaking the language spoken by the people imitated. Moreover, that benefit may not be achieved until much later. So having that kind of instinctive motivation to imitate illustrates the concept of architecture-based motivation, as opposed to reward-based motivation.

I don't have time to say more about that today, but it illustrates an important point: Biological evolution can achieve amazing results in very indirect ways.

The problem of education

There is a vast amount of literature already available related to my topics, in a wide variety of disciplines, and information keeps growing as new theories, techniques, applications and prejudices (!) emerge.

Unfortunately it is very difficult for most people to learn about more than a tiny subset of all these topics. It is somewhat easier for retired people like me, though even for us keeping up with new publications has become impossible. So it is possible that this lecture has some important gaps or errors. If you find some, please let me know.

So anyone writing anything about AI risks getting something wrong as a result of being unaware of something important and relevant that has been discovered elsewhere, and even published.

For reasons like this I strongly advise all students to be highly skeptical of all claims to have a complete, or final, or even a good description or explanation of some class of biological phenomena concerning cognition or motivation, or other forms of *affect*.

Some of my own history

I was introduced to AI by a researcher on machine vision, Max Clowes, around 1969 when I was still a lecturer in philosophy at Sussex University. There's more information about some of his inspiring ideas here:
http://www.cs.bham.ac.uk/research/projects/cogaff/81-95.html#6

He inspired me to learn to program and gave me papers written by leading AI researchers to read, especially McCarthy and Hayes(1969).

Since I did not agree with everything they said, including their claim that human like intelligent machines can do all their reasoning, perceiving, learning, decision making, etc. using a logic-based form of representation, I decided try to find ways of making advances using other forms of representation, including reasoning about spatial structures and processes by making use of spatial structures and processes, as explained in Sloman(1971) also elaborated in my 1978 book. which is now freely available online and is updated from time to time:

The Computer Revolution in Philosophy, Philosophy, Science and Models of Mind,
Originally published in 1978, now available in html and pdf:
http://www.cs.bham.ac.uk/research/cogaff/crp/
http://www.cs.bham.ac.uk/research/cogaff/crp/crp.pdf

Unlike most researchers in the field of AI, my primary interest is not in developing new useful machines. Rather I am trying answer philosophical and scientific questions about what minds are (including human and non-human minds), how they work, how they develop, how biological evolution produced them starting from a planet with no minds, and what the physical universe must be like in order to make those evolutionary achievements possible.

New answers to such questions about evolved forms of natural intelligence may provide new clues regarding designs for artificial forms of intelligence, or more generally new designs for functional artificial minds.

Such work can also reveal gaps and errors in existing other popular theories, e.g. theories in philosophy, psychology, and neuroscience, about the nature of mathematical discovery (or mathematical consciousness) and stimulate development of new theories and models.

In particular, I think there is good evidence, presented below, that some important ancient mathematical competences, used in the discovery of concepts and theorems in Euclidean geometry makes use of forms of intelligence that humans share with other intelligent animals, and which have proved very difficult to replicate in computers.

This led me to speculate that in order to replicate what the brains of ancient mathematicians and other intelligent animals do we may need new kinds of computer that operate not only on discrete bit-patterns (binary codes) but also on continuously deformable structures, as might be used if you imagine the process of peeling a banana, or drawing a circle, for example.

After all, human brain cells include very complex processes involving more or less complex molecules doing things in huge numbers at very high speeds, such as forming or releasing chemical bonds (discrete processes) and also twisting, folding, moving together and apart (continuous processes). If all that sub-neural chemical processing plays an essential role in the spatial perception and reasoning done by humans and other intelligent animals then replication of natural intelligence in machines is still a very long way off.

Serious gaps in our thinking about computation?

Those ideas may point to very serious gaps in our current ideas about computation, now generally assumed to be concerned with manipulation of discrete structures (e.g. symbols in logical, algebraic, or numerical formulae) which can be implemented in mechanisms based on bit-patterns i.e. on digital computers. I'll return to that below.

Working on such designs, and "debugging" theories that turn out not to be practically useful, can also help to shed light on old philosophical, psychological, and biological questions about natural intelligence, including questions about consciousness of various sorts, motivation, emotions and other kinds of "affect" in humans and other animals, forms of learning and discovery, forms of cooperation, and many more.

But making that sort of progress requires learning how to design, build, test, analyse and modify working systems, which is not yet a normal part of the education of philosophers or psychologists.

It is important not to try to understand (or model) only human minds, since there are many other animals whose forms of intelligence overlap with human intelligence (e.g. making possible collaboration between a shepherd and a sheepdog), and that requires some partial overlap in the types of information they use. For more information on investigating "The space of possible minds" see Sloman(1984). (It really needs to be updated when I find time!)

The task of understanding these aspects of cognition in humans requires a great deal of multi-disciplinary collaboration, and development of new forms of research for collecting requirements for explanatory mechanisms.

Human and non-human abilities to understand and make use of spatial structures and processes, have mostly existed for a very long time. Human abilities overlap with abilities of other animals with high spatial intelligence which they use in building nests, hunting, foraging, looking after their young, and other activities.

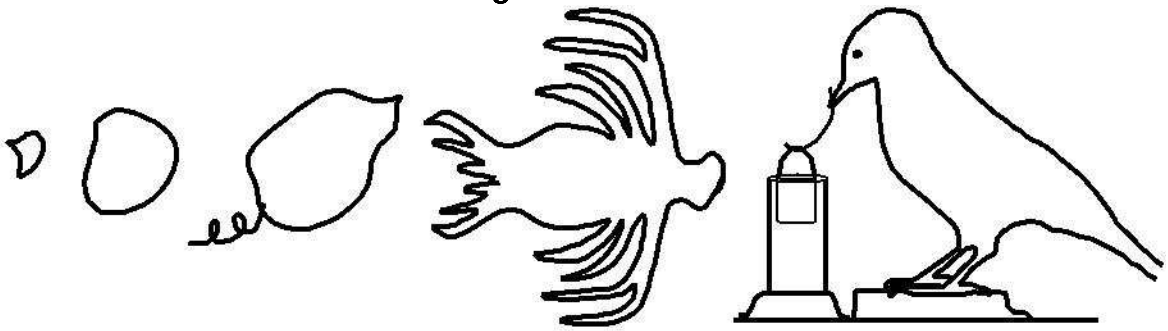
This include abilities to perceive and make use of or avoid various objects or other features of the 3-D environment.

Examples include perceiving and consuming, and in some animals hiding/storing, food, choosing or building shelters or nests, cooperating with or competing with other members of the same species, including caring for offspring (or eggs: unhatched offspring), avoiding, or pursuing, or fighting with other animals for various reasons, e.g. to eat or avoid being eaten, to protect offspring, to win a mate, to build a nest or shelter, or to find nests of other organisms as a source of food.

I shall try to present some ideas about how these abilities develop in intelligent organisms, which, if correct, have profound implications about the problems of replicating human intelligence in robots. The ideas were developed in collaboration with Dr Jackie Chappell, a biologist who studies animal cognition.
https://www.birmingham.ac.uk/staff/profiles/biosciences/chappell-jackie.aspx

Some evolutionary transitions to think about:
Changing relationships between organism and environment

Fig: Transitions



Types of environment with different information-processing requirements

- Microbes in chemical soup, can be wholly dependent on nutrients in their neighbourhood.
- Soup with detectable gradients: offers opportunities to improve location if motion can be controlled.
- Soup plus some stable structures, e.g. places with good stuff, bad stuff, obstacles, supports, shelters: offers advantages for organisms that can build up long term spatial memories, and can plan routes.
- Many different kinds of matter in the environment, with different nutritional, toxic, or other properties, not all immediately distinguishable from external appearance. Requires notion of "kind of stuff" related to potential behaviours and/or responses to prodding, pulling, twisting, hitting, pouring, etc. (What philosophers have called dispositional properties.)
- Things that have to be manipulated to be eaten (e.g. disassembled) need new forms of process perception and process control, including ontologies that include kinds of ‘‘stuff’’ with different properties.
- Controllable manipulators: mouth, hands, feet, require different uses of information.
- Things that try to eat you -- lead to forms of control for escaping, hiding, defending, etc.
- Food that tries to escape -- lead to forms of control for chasing, trapping, lying in wait, heading off, etc.
- Mates with preferences, lead to forms of control of behaviours for attracting attention and winning favour.
- Competitors for food and mates lead to deception, fighting, warding off, defending territory, etc.
- Collaboration with others requires action controlled to aid collaboration, and possibly communication.
- and so on (don’t forget information-processing in plants) .....

NB: This list illustrates only a **tiny** subset of the diversity of requirements and designs for information-processing functions and mechanisms in products of biological evolution. There may have been thousands of important transitions in information processing functions and designs in our evolutionary ancestors (some more important than others).

Some of the solutions seem to have been "compiled" into genomes for species that have survived for a long time. For example, soon after birth a human infant needs to be able to control a complex collection of muscles in and around the mouth that control sucking milk from a nipple. That ability is shared with the young of other mammals, some of which are born with even more sophisticated abilities, such as new-born foals that find their own way to the mother's nipple because she can't pick them up to feed them.

Other control mechanisms produced by evolution are not so well specified at birth, but have instead been "meta-compiled" or "partially compiled" into parametrisable specifications that are instantiated over an extended period during development and can cope with novel environments like human-toddlers using mouse-pointer on a computer screen -- unlike any of their ancestors at that age. It is unlikely that evolution anticipated such products of evolution and built into the human genome specifications of how to use those products to help them survive.

genome does not include mechanisms for understanding the screen and deciding what to do then they have work something out.

How information-processing **requirements** change, depends on both features of the environment and features of the organism (products of previous evolution).

Contrast: how **designs** change.

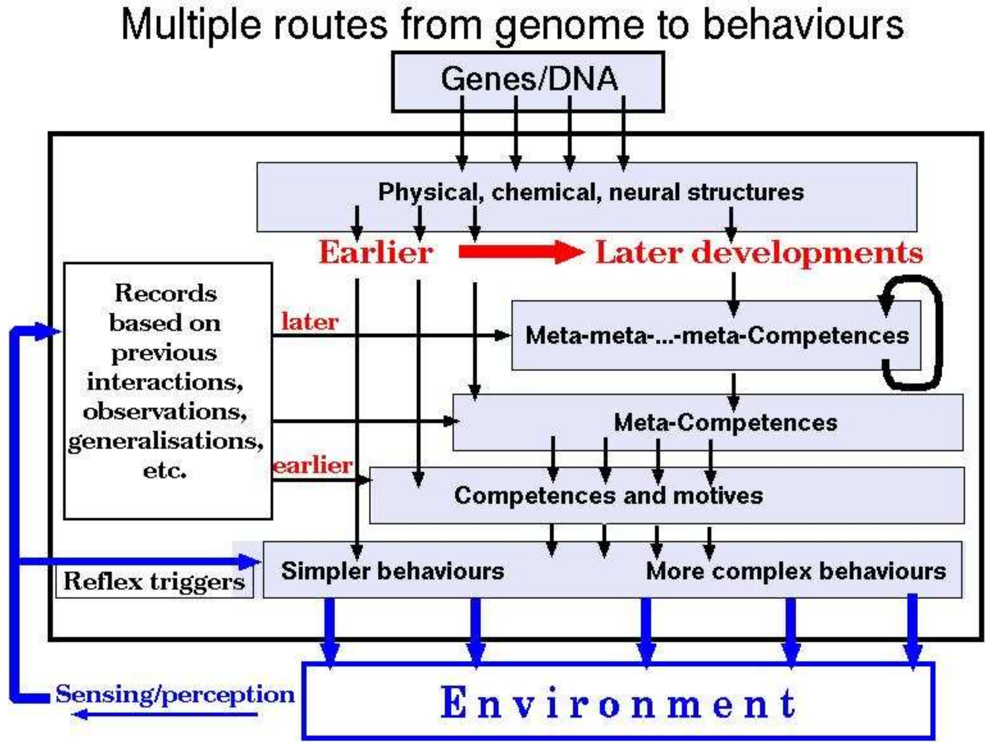
The meta-configured genome: Multi-layered gene expression and development

In more complex organisms, evolution discovered the power of layered epigenetic mechanisms that use different mathematical structures (including new combinations of simpler structures) at different stages of development. As a result, recently evolved sophisticated learners have multi-stage, "meta-configured" genomes [2], whose later stages of gene expression produce more complex, more recently evolved, more mathematically complex and abstract, discovery and learning mechanisms with different evolutionary histories. The best known example of this is the collection of different developmental transitions in (signed, spoken and written) language "learning" capabilities. (Strictly, these are language creation capabilities, which is why there are so many different human languages.)

> If humans had to learn languages, instead of being able to create them (cooperatively), then the thousands of known human languages could not have existed, because there were originally no language users to learn from, at least not on this planet.

A meta-configured genome repeatedly produces new learning powers within each developing individual. The "layered" developmental processes are depicted crudely in this diagram (joint work with Jackie Chappell):

Figure: Meta-Configured Genome



A video -- under 10 min. -- extracted from the lecture video explaining this diagram, is available on Youtube here: https://youtu.be/G8jNdBCAxVQ

A key feature is that the specific powers produced during development are not all specified in the genome: some of them, e.g. later acquisition of skills and knowledge, are absorbed from aspects of the environment influenced by ancestors, e.g. results of prior stages of cultural evolution, or changes in other species, as illustrated by evolved symbiosis.

E.g. in 2019, development in many young children is strongly influenced by computer based products of older generations who did not have access to those products when they were children.

Likewise, when I was a child (in the 1940s-1950s) after I had been deeply influenced by playing with a series of increasingly complex Meccano sets (construction kits with vast generative powers, yet highly constrained), my development was strongly influenced by others who had written books on Euclidean geometry, whereas a few thousand years earlier there were no teachers with knowledge of Euclidean geometry nor books on Euclidean geometry. The minds of humans in those prehistoric times had significant gaps, some of which were filled in later generations by the work of mathematicians who made new discoveries and produced teaching practices and materials passing on those discoveries. (There will always be unfilled gaps. This project helps to identify deep gaps in our knowledge of how minds work -- which may lead to future developments in which some, or all, of the gaps are filled.)
(Some of those ancient mathematical discoveries are no longer taught in all schools because of misguided educational decisions in the 20th Century. As a result there are now many AI researchers and knowledge engineers whose minds lack important relevant information that had been removed from their school syllabus!)

Some of the more recently evolved parts of the human genome, whose expression occurs relatively late during development, (e.g. in teen years, or later when working on research projects, artistic creation, engineering design laboratories, etc.), can produce enormously varied adult competences, including inventive powers that cannot result simply from a uniform learning mechanism presented with examples of invention.

We can think of competences produced by later stages of gene expression as being parametrised. Information in the relevant genes has gaps for parameters that are extracted from relevant subsets of the records derived from the environment during earlier gene expression (box on left). Since those records can be very different in different cultures, the genes expressed late may have very different effects in individuals born with identical genomes developing in different environments.

Nevertheless many aspects of the environment in which the individual genome is expressed include activities and products of other members of the same species, including products of previous generations as well as actions and other products produced by living conspecifics of different ages. In that sense, evolution has discovered ways of using parts of the environment, including parts produced by members of the same species, as extensions of the genome, or parts of the "Extended Genotype" (Sloman and Chrisley,2005).

Language development is an obvious example, but there are many more. Moreover, these processes can constantly modify the environment in which individuals develop, so that each individual learns and develops in a context that is different from the context available the parents of that individual.

If the parents and their parents all enrich the development environment of their offspring in novel ways, e.g. by taking them to novel locations, introducing them to novel games, toys, tools, techniques and manufacturing processes, then the offspring in each generation may push the achievements further.

In that way the same genome parts of which develop late, building on achievements of earlier parts, and repeatedly doing that in each new generation, can produce enormously varied late developing products.

This is evident in the history of mathematics, science, art, literature, engineering, architecture, etc. (I believe these ideas are related to but different from Annette Karmiloff-Smith's ideas about "representational redescription" in her book *Beyond Modularity* (1992).

All that diversity can therefore be attributed, ultimately, to the creativity of biological evolution, including both its ability to produce meta-configured genomes and its ability to produce constantly changing external products of roughly the same genomes.

In some ways the processes of biological evolution seem to behave like mathematicians who, when they have solved a problem, or identified some class of structures, often as: "Is this a special case of something more general, and if so are there other special cases that we have not yet encountered". One way of moving to "something more general" is to consider a part or aspect of the class of structures as having a location in the structures that might be occupied by something else. That leads to replacing something constant with a "variable" that can take different values: e.g. replacing multiplication and addition of numbers with other operations on other objects, leading from the particular case of arithmetic to something more abstract, which, when instantiated in a different way may have new features. An example would be replacing multiplication and addition of numbers with some other operations on another class of entities.

Depending on the ways in which genetic codes are structured and interpreted during development of organisms, it may be possible for this process of abstraction followed by novel instantiation to produce something new and biologically useful, e.g. a new class of physical structures, or new classes of behaviours. In that case the processes of abstraction followed by new instantiation may produce new species, partly similar to older species. Alternatively it may produce variations in later stages of development within a species. If, in addition a process of abstraction allows different forms of instantiation to occur relatively late in an organism's development that may produce novel behaviours without requiring major changes in physical forms. In that case the same species may develop different behaviours and other adaptations in different environments. Such mechanisms may make possible meta-configured genomes of the sort described above. In particular members of the same species developing in different environments may develop in very different ways, based on different parameters used.

This is very different from, and potentially a much richer source of variation, than, the commonly postulated uniform powerful learning mechanism (e.g. deep learning of some sort) that can be used in many different environments.

If evolution does produce changes in the manner just suggested the evolutionary process can be compared with a mathematician, or group of mathematicians who after acquiring deep knowledge of a particular class of mathematical structures explore a variety of alternative structures, including the previous class as a special case.

That would allow a form of cultural evolution to produce major useful changes far more quickly than standard Darwinian evolution. Moreover, information gained about how to cope with the new environment using the new genetic mechanisms may allow a form of cultural transmission that gives the appearance of Lamarckian inheritance, without waiting for random mutations to produce the benefit.
This section of this document summarises aspects of a separate document on the Meta-Configured Genome:
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/meta-configured-genome.html (also pdf)

Can AI catch up with natural intelligence?

It is obvious that in some ways computer-based machines out-perform humans, including storing and retrieving information with great accuracy and precision, and in performing computations with many intricate details at very high speeds, e.g. in designing printed circuit boards for the next generation of computers.

Despite impressive recent results, nothing in current AI matches or explains the evolved types of spatial intelligence found in squirrels, weaver birds, elephants, human toddlers and ancient mathematicians, e.g. Archimedes, Euclid, Zeno, and others. (That's why I'll refuse to use a self-drive car in cluttered busy urban environments.)

Current psychology and neuroscience also lack the required explanatory power.

These claims can be supported by a collection of examples, including spatial reasoning competences of pre-verbal toddlers, squirrels, nest-building birds, and others. (For a sample, see this discussion of "toddler theorems":
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/toddler-theorems.html (Also PDF)

Languages are created not learned

Artificial intelligence researchers and theoretical linguists have done a lot of work trying to explain produce working models of human language development and use. In the last few decades that has been dominated by attempts to train machines by giving them huge amounts of linguistic data from which they can find patterns. I think that work can be useful in limited contexts, but is doomed to failure in the long term for reasons that I don't have time to discuss in detail today, except to say that evidence from child development suggests that a great deal of human linguistic development is driven by very complex multi-layered genetic mechanism which produce language *creation*, not language *learning*. Some evidence, is presented, advertising a BBC video recording, related to deaf children in Nicaragua, who created a new sign language.
https://www.youtube.com/watch?v=pjtioIFuNf8

For this talk I shall not attempt to defend all these claims because there is not enough time, and I want to focus on an aspect of human and animal minds that is very important, and which has not been understood by psychologists or neuroscientists and has not been replicated in robots, although I think Immanuel Kant had some deep insights, largely ignored nowadays.

Hume and Kant on types of knowledge

Many philosophers since ancient times have tried to develop theories about what we know, whether we could be wrong because we are deceived by appearances, whether we can know anything about the future, whether we can know anything about things that are too small, too far away or for other reasons not perceivable by us, e.g. electromagnetic waves used for radio transmission.

Some philosophers have doubted whether we can know anything except the contents of our own minds. (You may have encountered Plato's fanciful idea that we are like people sitting in a cave looking at shadows on a wall produced by people outside the cave walking past the entrance. But we know nothing about what's outside the cave and think that the shadows are the full reality.

I leave it to you to work out for yourselves whether there is any way of designing intelligent animals or machines that are able to perceive, learn about and act in a complex environment that extends beyond what they have access to, without leaving them essentially stranded in Plato's cave -- having access only to their own sensory signals and thought processes.

A question you can ask is whether even that images on the walls of the cave are still not directly accessible: information has to come from them to retinal cells at the back of your eyes, or other sensors on and in your body.

But how do you get information about those sensory signals? Do they also have to be transmitted to other receivers, perhaps with a never ending chain of receivers transmitting to other receivers deeper inside the person, animal or robot that is fooled into thinking she/he/it has access to information about the environment.

That line of thought leads to an infinite regress. I think the actual mechanisms are currently only partly understood. If we wish to design intelligent machines, we'll have to find ways of ensuring that they are not subject to any infinite regress that prevents them perceiving their environment.

The solution can't be the same for all organisms. Evolution has produced many branching strands of development of increasingly complex and sophisticated information users, as discussed above.

Sometimes it is not easy to specify exactly what evolution has achieved. I'll now try to illustrate this using a disagreement between two great philosophers.

David Hume and Immanuel Kant (from Wikimedia)

Very crudely, David Hume, depicted above, on the left, claimed that there are only two kinds of real knowledge:
-- empirical knowledge that comes through sensory mechanisms, possibly aided by measuring devices of various kinds; which called knowledge of "matters of fact and real existence"
-- what he called "relations of ideas", which we can think of as things that are true by definition, such as "All bachelors are unmarried", and (if I’ve understood Hume rightly) all mathematical knowledge, for example knowledge of arithmetic.

He then famously claimed that if someone claims to know something that is neither simply derivable from definitional relations between ideas nor "based on experimental reasoning concerning matter of fact and existence" we should "Commit it then to the flames: for it can contain nothing but sophistry and illusion", which would have included much philosophical writing by metaphysicians, and theological writing.

In response to Hume, Immanuel Kant, also depicted above, on the right, claimed that there are some important kinds of knowledge that don’t fit into Hume’s two categories ("Hume’s fork"), for they are not mere matters of definition, or derivable from definitions purely by using logic, i.e. they are not analytic but synthetic, and they also are not merely derived from experience in such a way that future experiences might prove them false, as could happen with most supposed knowledge about the world around us. E.g. the fact that for millions of years sunset on this planet has always been followed by sunrise does not prove that that will always be the case, since, for example, after some future sunset a huge asteroid might collide with the earth, smashing it to pieces and preventing any future sunrise here.

But, says Kant, some kinds of mathematical knowledge do not fit into either of Hume’s two categories: since we can discover by means of special kinds of non-logical, non-empirical reasoning, that I don’t think he was able to explain, that "5+3=8" is a necessary truth, but not a mere matter of definition, nor derivable from definitions using only logic. He thought such mathematical discoveries in arithmetic, and discoveries in Euclidean geometry were synthetic, not analytic and also could not possibly be false, so they are necessary truths, and because they are not based on or subject to refutation by observations of how things are in the world, such knowledge is non-empirical, i.e. a priori.

Kant also made use of the Necessary/Contingent distinction. Some statements that are true could have been false, for example "The planet earth has only one moon", and some that are false could have been true e.g. "The planet earth is too hot to sustain human life", if the earth had been closer to the sun.

For a more careful and detailed, but fairly brief explanation of Kant’s three distinctions. apriori/empirical, analytic/synthetic and necessary/contingent, see http://www.cs.bham.ac.uk/research/projects/cogaff/misc/kant-maths.html

My 1962 DPhil thesis was an attempt to defend Kant against critics who thought he had been proved wrong because work by Frege, Russell and others had shown that arithmetical knowledge was reducible to logic, and therefore analytic, and claimed geometrical knowledge about the nature of space had been proven false by Einstein’s general theory of relativity, supported by evidence gained by Eddington and collaborators during the 1919 eclipse of the sun. https://en.wikipedia.org/wiki/Solar_eclipse_of_May_29,_1919

In 1962 I knew nothing about computing or AI, but after learning to program several years later I thought I might be able to prove that kant was right by showing how to design a "baby" robot that could grow up to be a mathematician making discoveries like Archimedes, Euclid and others. However, so far I don’t think anyone knows how to build such a robot. This may because we don’t really know how human brains work. Perhaps our failure to replicate their functioning so far is explained by the fact that chemical computations, using a mixture of discrete and continuous processes, play essential roles in brains, but cannot be replicated on computers, which are purely digital. Research on answering such questions requires a deep combination of philosophy and science, perhaps investigating forms of information processing required for ancient mathematical discoveries, that cannot be fully implemented on digital computers.

It may be that Alan Turing had related thoughts when he wrote, in his PhD thesis, published in 1938, that human mathematical intuition cannot be replicated by computers whereas human mathematical ingenuity can be. This is discussed further in: http://www.cs.bham.ac.uk/research/projects/cogaff/misc/turing-intuition.html

If he had a hunch that replicating mathematical intuition might require machines using a mixture of discrete and continuous reasoning mechanisms, that could help to explain why, in 1952, shortly before his death, he published Turing(1952), which was quite unlike anything he had written previously. But he died in 1954 so we’ll never know what he might have achieved later. I’ll return to this question below after discussing different sorts of reasoning.

Philosophy and Science

I disagree with most philosophers and scientists about the nature of philosophy and science and how they relate to one another. That is spelled out in Chapter 2 of The Computer Revolution in Philosophy. http://www.cs.bham.ac.uk/research/projects/cogaff/crp/#chap2

Philosophy, among other things, asks questions about how things are possible, and so does science, e.g. how is motion possible, how is it possible for minds to exist in a material world, and many more.

The difference is that after a lot of exploration of various kinds answers begin to become more precise and more useful, e.g. explaining how things are possible that were never previously thought of, such as wireless communication, or conversion of matter into energy, or creation of new kinds of minds by making new kinds of machines.

When the proposed answers become testable and useful and start suggesting new observations and experiments, we tend to say that that’s no longer philosophy: it’s science.

In the case of physics, chemistry, geology, biology, astronomy, and other sciences, those changes began a long time ago. But more recently we have begun for the first time to produce tentative answers to questions about how minds are possible in a physical universe and more specific questions such as how perception of the environment is possible, how finding scientific explanations is possible.

And the new answers are beginning to lead to designs of new kinds of machinery with capabilities that to most people were totally unexpected: machines apparently performing in human-like ways, such as making plans, answering questions, holding conversations and (if you don’t examine them too closely) even apparently feeling emotions.

My view is that although progress has been very real and there are many very useful applications, there are also deep gaps between what AI machines can do and what intelligent animals, including pre-verbal human toddlers, squirrels, nest building birds, elephants and others can do.

Show video of toddler topologist with pencil.
http://www.cs.bham.ac.uk/research/projects/cogaff/movies/ijcai-17/small-pencil-vid.webm

Show weaver birds on BBC video
https://www.youtube.com/watch?v=6svAIgEnFvw

(Propositional) Logical reasoning: easy for computers

A lot of work has been done on automated theorem proving using logic, arithmetic and algebra. This reasoning uses manipulation of notations allowing only discrete transitions, e.g. modifying a symbol, adding or removing a premise. The treatment of elementary propositional logic using truth-tables is similarly discrete. Elementary examples are concerned only with sets of possibilities in a finite space of possibilities. For example, in propositional logic, many students learn how operators related to the words, "not", "and", "or", can be defined using truth tables. Then the validity of inferences can be checked using truth tables, though the sizes of the truth tables grow exponentially with the number of propositions (or propositional variables) involved.

A toy example is reasoning in propositional calculus, illustrated here:

Figure Logic

Premises: P or Q not Q Conclusion: :- P
Premises: P or Q or R not Q Conclusion: :- P

By starting from the truth tables for "or" and "not" you can reason about which possible combinations of truth values for P and Q (and R in the second example) will make the premises true and ask whether the conclusion can be false in any of those cases.
If it is impossible for all the premises to be true while the conclusion is false, then the inference is valid. One of these is not valid.

Here the symbol "**:-**" can be read as "therefore", to indicate the conclusion of an inference to be evaluated.

In the first (upper) example, where only two propositions P and Q are involved there are only four possible combinations of truth values to be considered. And that makes it easy to discover that no combination makes both premises true and the conclusion false.

In the second case, for each of those four combinations R may be true or false, so the total number of possibilities is doubled. But it is still a finite discrete set and can be examined exhaustively to see whether it is possible for both premises to be true and the conclusion false. I assume the answer is obvious for anyone looking at this who understands "or" and "not". Checking for validity in propositional calculus involves exhaustive search through finite sets of discrete possibilities, so it is easy to program computers to do this. More generally, it is the discreteness of the structures and transitions and the fact that there are only finitely many elementary components that allows such reasoning to be modelled on Turing machines, and digital computers.

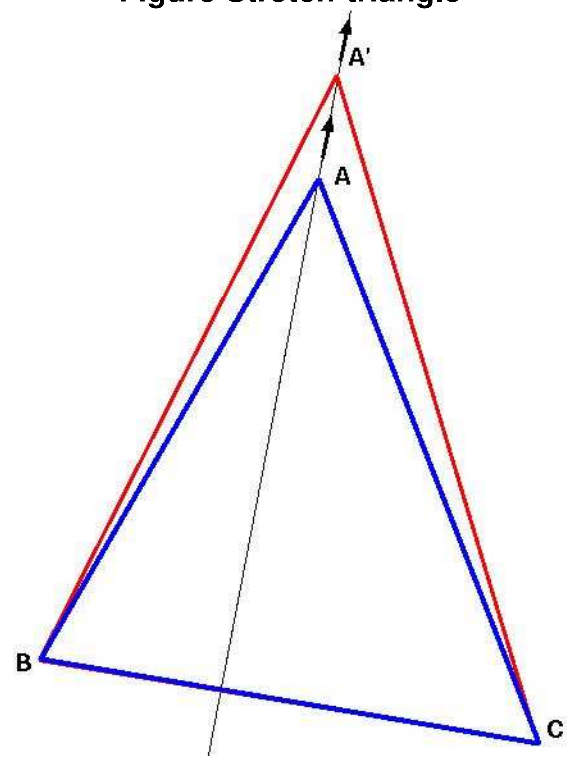
Things get more complex if the propositional variables, e.g. P, Q, etc. instead of being restricted to two possible truth values (T and F) can take other intermediate values, or even vary continuously. The simple discrete reasoning, based on truth-tables, using or and not etc., will have to be replaced by something mathematically much more complex.

Geometrical/topological reasoning is much harder for computers

It is not yet clear whether computers can replicate the discoveries made by ancient mathematicians, over 2000 years ago, long before the discoveries in modern logic, formal reasoning, set theory, and Descartes' "translation" of geometry into arithmetic using coordinates.

Ancient geometrical discoveries, were not concerned only with collections of discrete possibilities. Euclidean geometry (like much of real life perception and action) is concerned with smoothly varying sets of possibilities, including continuously changing shapes, sizes, orientations, curvature and relationships between structures. In some cases a smooth change in one relationship can produce discrete changes, for instance a point moving smoothly in a plane, from outside a circle to inside the circle. I'll start with a deceptively simple example, and later reveal hidden complexity, discussed in a separate document. Contrast the simple truth-table analysis above, involving only discrete options, with reasoning about continuous deformation of a triangle.

Figure Stretch-triangle



What will happen if you start with a triangle, like the blue triangle in Figure Stretch-triangle, and move the top vertex away from the opposite side along a line going through the opposite side? What happens to the size of the angle at the vertex, as the vertex A moves further from the base, BC, e.g. moving from location A to A'? Here the red triangle illustrates one of the possible new triangles that could result from moving the top vertex along the thin line through point A (which goes between B and C).

I usually find that even people who have never studied Euclidean geometry find the answer obvious after a while -- not quite as obvious as the answer to a topological question like this:

> Given that it is possible to remove a shoelace from a shoe by pulling one end, and equally possible to remove the shoelace by pulling the opposite end, shouldn't the lace come out twice as fast if you pull both ends at the same time, but not hard enough to break the lace?

What kinds of brain mechanisms allow a child to find it obvious that the answer is "No" because it is impossible to remove the lace by pulling both ends at the same time? Compare the impossibility of unlinking two linked rings made of solid unbreakable material. Although I won't defend this claim here, I believe there is nothing in current neuroscience or psychology that explains such an ability. It can't just be experience gained from much trying and failing, because that sort of evidence merely supports the claim that success is unlikely, not that it is impossible.

I suggest that there is something in brains that allows a continuously variable structure to be considered in imagination in such a way as to reveal impossibilities or entailments, though the mechanisms must be very different from those that allow consideration of an exhaustive set of discrete possibilities in a truth table to be used to establish impossibility. The latter is easy to implement on a computer. The former can't straightforwardly be implemented on a computer using any programming technique I know of.

There is an alternative way of attempting to prove that something is geometrically impossible, namely find some axioms (statements) about spatial structures that can somehow be seen to be obviously true in all possible situations (how?), then use pure logic to derive the desired conclusion. (I leave that as an exercise for mathematically experienced readers)

In principle that can be done, using logical theorem provers, but the resulting combination of complex axioms and powerful theorem prover does not obviously replicate or explain what's going on in a human mind when the above impossibility is recognized using spatial reasoning abilities.

I believe those ancient reasoning capabilities are connected with aspects of very rapid unconscious reasoning as we or things in our spatial environment move while other things remain fixed, or when tings move in different directions.

This video illustrates what I am talking about.
http://www.cs.bham.ac.uk/research/projects/cogaff/movies/chairs/two-chairs.mp4

There are some related videos here:
http://www.cs.bham.ac.uk/research/projects/cogaff/movies/chairs/
This link provides more videos and a partial textual description and analysis of what is going on.

For reasons that are too complex to discuss here, although it is possible to give robots video cameras that can acquire the kinds of "retinal" projection information that you get as you move around a crowded or cluttered scene, there is not yet, as far as I know any system that can simultaneously track changes in the projected retinal images and instantly interpret those changes as due to perception of 3D structures from a viewpoint that is constantly changing its location and direction of view, and simultaneously work out which changes in the projected image are due to object motion, not viewer motion, as in the video including the pot plant on one of the chairs.

These visual processes that require brains to recognise and reason about multiple spatial objects in a variety of locations and orientations, with some stationary and others moving in a variety of trajectories are probably shared with many intelligent animals including mammals and birds that move around in complex environments in which some of the other visible objects may also be moving, though not necessarily in fixed directions.

The perceptual processes and mechanisms are closely related to those postulated and discussed in connection with perception of affordances by James Gibson(1979). However he discussed only relatively simple cases where the optical flow patterns or texture gradients could be summarised in relatively simple mathematical formulae, as demonstrated by some AI vision research a few decades ago. For a computer to replicate what brains do with the much richer combinations of visual change when walking through a forest or approaching a tree, or even moving around a kitchen, as illustrated by my videos showing chairs, and using the changing flow patterns to derive information about 3D structures, relationships and processes very quickly in complex scenes will be much harder.

Future work

Future work should investigate possible new forms of computation, including mixtures of discrete and continuous processing, to see whether they can answer some of the hard problems about human visual perception and human spatial reasoning, and then finally show how to design robots capable of making the same discoveries as Archimedes, Euclid, Zeno, and many other great ancient mathematicians.

In that case we may also be able to explain the perceptual and problem solving capabilities of pre-verbal human toddlers, elephants, squirrels, crows, apes and many more.

But it may turn out that there is only one way to build the required sorts of computing machinery, namely the way it has been done for thousands of years making use of the great scientific and engineering discoveries of biological evolution.

That will require us to make use of the relevant concept of information, which is not Claude Shannon's concept Shannon(1948). but the concept of information used by the novelist Jane Austen, and many others long before her, as explained in
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/austen-info.html

One of the unobvious implications of the ideas about evolution presented here is that evolution must have produced rich internal languages for use in processes of perception, reasoning, planning, control of actions, long before there were external languages used for communication, as I have argued, e.g. in this online presentation, also suggesting that sign languages would have grown out of cooperative languages before the development of spoken languages. That may be why all humans are born able to develop sign language competence, even though most never have the need to do so: (http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#talk111)
If correct, this has deep implications for AI as well as psychology, neuroscience and education.

CONCLUSION

There is no separate conclusion, as it would merely repeat the abstract above!

A RELATED ONLINE LECTURE AUGUST 2017

In 2017 I was invited to present a recorded lecture, later discussed via the internet, to this IJCAI Workshop in Australia

> Architectures for Generality and Autonomy
> http://cadia.ru.is/workshops/aga2017/
> My talk was entitled:
> Why can't (current) machines reason like Euclid or even human toddlers?
> The online web page has links to several separately recorded sub-talks:
> http://www.cs.bham.ac.uk/research/projects/cogaff/misc/ijcai-2017-cog.html
> There is some overlap with this lecture, but some students may find the differences useful.

REFERENCES
(From a longer paper. To be pruned later)

M. A. Boden, *The Creative Mind: Myths and Mechanisms,* Weidenfeld & Nicolson, London, 1990. (Second edition, Routledge, 2004).

M. A. Boden, 2006, *Mind As Machine: A history of Cognitive Science (Vols 1--2),* OUP, Oxford
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/boden-mindasmachine.html

[Pythag] Alexander Bogomolny, (2017) Pythagorean Theorem and its many proofs from *Interactive Mathematics Miscellany and Puzzles*
http://www.cut-the-knot.org/pythagoras/index.shtml
Accessed 15 August 2017

Euclid and John Casey, *The First Six Books of the Elements of Euclid,* Project Gutenberg, Salt Lake City, Apr, 2007, http://www.gutenberg.org/ebooks/21076
Also see "The geometry applet"
http://aleph0.clarku.edu/~djoyce/java/elements/toc.html (HTML and PDF)

Jackie Chappell and Aaron Sloman, (2007) "Natural and artificial metaconfigured altricial information-processing systems", *International Journal of Unconventional Computing,* 3(3), 221-239.
(Includes the first presentation of the meta-configured genome idea -- with a simpler diagram than the one used here.)
http://www.cs.bham.ac.uk/research/projects/cogaff/07.html#717

N. Chomsky, 1965, *Aspects of the theory of syntax,* MIT Press, Cambridge, MA.

Shang-Ching Chou, Xiao-Shan Gao and Jing-Zhong Zhang, 1994, *Machine Proofs In Geometry: Automated Production of Readable Proofs for Geometry Theorems,* World Scientific, Singapore, http://www.mmrc.iss.ac.cn/~xgao/paper/book-area.pdf

Juliet C. Coates, Laura A. Moody, and Younousse Saidi, "Plants and the Earth system - past events and future challenges', *New Phytologist,* 189, 370-373, (2011).

Alan Turing - His Work and Impact, eds., S. B. Cooper and J. van Leeuwen, Elsevier, Amsterdam, 2013. (contents list).

Kenneth Craik, 1943, *The Nature of Explanation,* Cambridge University Press, London, New York,

Deaf Studies Trust, 'Deaf Children Beginning to Sign', A paper prepared for Frank Barnes School, London, 17th February 1995

D. C. Dennett, 1978 *Brainstorms: Philosophical Essays on Mind and Psychology*. MIT Press, Cambridge, MA.

D. C. Dennett, 1995, *Darwin's Dangerous Idea: Evolution and the Meanings of Life,* Penguin Press, London and New York,

D.C. Dennett, 1996
Kinds of minds: towards an understanding of consciousness,
Weidenfeld and Nicholson, London, 1996,
http://www.amazon.com/Kinds-Minds-Understanding-Consciousness-Science/dp/0465073514

T. Froese, N. Virgo, and T. Ikegami, Motility at the origin of life: Its characterization and a model', *Artificial Life,* 20(1), 55-76, (2014).

ibor Ganti, 2003 *The Principles of Life,* OUP, New York, Eds. Eors Szathmary & James Griesemer, Translation of the 1971 Hungarian edition.

Gallistel, C.R. & Matzel, L.D., 2012(Epub), The neuroscience of learning: beyond the Hebbian synapse, *Annual Revue of Psychology,* Vol 64, pp. 169--200, https://doi.org/10.1146/annurev-psych-113011-143807

H. Gelernter, 1964, Realization of a geometry-theorem proving machine, in *Computers and Thought,* Eds. Feigenbaum, Edward A. and Feldman, Julian, pp. 134-152, McGraw-Hill, New York, Re-published 1995 (ISBN 0-262-56092-5),
http://dl.acm.org/citation.cfm?id=216408.216418
Also at
https://pdfs.semanticscholar.org/2edc/8083073837564306943aab77d6dcc19d0cdc.pdf
[Note: Like many geometry theorem provers this starts with a logical formulation of some of Euclid's axioms. Such program cannot see or reason about diagrams. They uses only logic, algebra and arithmetic. That's not how the original discoveries were made.]

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

Seth G.N. Grant, 2010, Computing behaviour in complex synapses: Synapse proteome complexity and the evolution of behaviour and disease, *Biochemist* 32, pp. 6-9, http://www.biochemist.org/bio/default.htm?VOL=32&ISSUE=2

M. M. Hanczyc and T. Ikegami, 'Chemical basis for minimal cognition', *Artificial Life,* 16, 233-243, (2010).

Immanuel Kant, *Critique of Pure Reason,* Macmillan, London, 1781. Translated (1929) by Norman Kemp Smith.
Various online versions are also available now.

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

Stuart Kauffman, 1995 *At home in the universe: The search for laws of complexity,* Penguin Books, London.

M.W. Kirschner and J.C. Gerhart, *The Plausibility of Life: Resolving Darwin's Dilemma,* Yale University Press, Princeton, 2005.

D. Kirsh, "Today the earwig, tomorrow man?', *Artificial Intelligence,* 47(1), 161-184, (1991).

I. Lakatos, 1976, *Proofs and Refutations,* Cambridge University Press, Cambridge, UK,

John McCarthy and Patrick J. Hayes, 1969, "Some philosophical problems from the standpoint of AI", *Machine Intelligence 4,* Eds. B. Meltzer and D. Michie, pp. 463--502, Edinburgh University Press,
http://www-formal.stanford.edu/jmc/mcchay69/mcchay69.html

John McCarthy, "The well-designed child', *Artificial Intelligence,* 172(18), 2003-2014, (2008). (Written in 1996).

Tom McClelland, (2017) AI and affordances for mental action, in Computing and Philosophy Symposium, *Proceedings of the AISB Annual Convention 2017* pp. 372-379. April 2017. http://wrap.warwick.ac.uk/87246

Tom McClelland, 2017 The Mental Affordance Hypothesis" MindsOnline 2017 http://mindsonline.philosophyofbrains.com/2017/2017-session-1/the-mental-affordance-hypothesis/ Video presentation https://www.youtube.com/watch?v=zBqGC4THzqg

Kenneth Manders (1998) The Euclidean Diagram, reprinted 2008 in *The Philosophy of Mathematical Practice*, OUP, pp.80--133 *Ed* Paolo Mancosu, http://doi.org/10.1093/acprof:oso/9780199296453.003.0005

Kenneth Manders (2008) "Diagram-Based Geometric Practice", In Paolo Mancosu (ed.), *The Philosophy of Mathematical Practice.* OUP, pp.65--79 http://doi.org/10.1093/acprof:oso/9780199296453.003.0004

Nathaniel Miller, 2007, *Euclid and His Twentieth Century Rivals: Diagrams in the Logic of Euclidean Geometry,* Center for the Study of Language and Information, Stanford Studies in the Theory and Applications of Diagrams, https://web.stanford.edu/group/cslipublications/cslipublications/site/9781575865072.shtml

David Mumford(Blog), Grammar isn't merely part of language, Oct, 2016, Online Blog, http://www.dam.brown.edu/people/mumford/blog/2016/grammar.html

Jean Piaget, 1952, *The Child's Conception of Number,* Routledge \& Kegan Paul, London.

Jean Piaget, *Possibility and Necessity* (1981/1983)

Vol 1. The role of possibility in cognitive development (1981)
Vol 2. The role of necessity in cognitive development (1983)
Tr. by Helga Feider from French in 1987 (Published after Piaget's death.)

Michael Rescorla, (2015) The Computational Theory of Mind, in *The Stanford Encyclopedia of Philosophy,* Ed. E. N. Zalta, Winter 2015, http://plato.stanford.edu/archives/win2015/entries/computational-mind/

Philippe Rochat, 2001, *The Infant's World,*
Harvard University Press, Cambridge, MA,

C. Rutz, S. Sugasawa, J E M van der Wal, B C Klump, & J St Clair, 2016, 'Tool bending in New Caledonian crows' in *Royal Society Open Science*, Vol 3, No. 8, 160439.
DOI: 10.1098/rsos.160439
https://research-repository.st-andrews.ac.uk/handle/10023/9283

A. Sakharov (2003 onwards)
Foundations of Mathematics (Online References)
Alexander Sakharov, with contributions by Bhupinder Anand, Harvey Friedman, Haim Gaifman,

Vladik Kreinovich, Victor Makarov, Grigori Mints, Karlis Pdnieks, Panu Raatikainen, Stephen Simpson,
"This is an online resource center for materials that relate to foundations of mathematics (FOM). It is intended to be a textbook for studying the subject and a comprehensive reference. As a result of this encyclopedic focus, materials devoted to advanced research topics are not included. The author has made his best effort to select quality materials on www."
http://sakharov.net/foundation.html
NOTE: some of the links to other researchers' web pages are out of date, but in most cases a search engine should take you to the new location.

Dana Scott, 2014, Geometry without points. (Video lecture, 23 June 2014,University of Edinburgh)
https://www.youtube.com/watch?v=sDGnE8eja5o

Erwin Schrödinger, *What is life?,* CUP, Cambridge, 1944.
Commented extracts available here:
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/schrodinger-life.html

Claude Shannon, (1948), A mathematical theory of communication, in *Bell System Technical Journal,* July and October, vol 27, pp. 379--423 and 623--656,
https://archive.org/download/pdfy-nl-WZBa8gJFI8QNh/shannon1948.pdf

Stewart Shapiro, 2009 We hold these truths to be self-evident: But what do we mean by that? *The Review of Symbolic Logic*, Vol. 2, No. 1
https://doi.org/10.1017/S175502030909011X

A. Sloman, 1962, *Knowing and Understanding: Relations between meaning and truth, meaning and necessary truth, meaning and synthetic necessary truth* (DPhil Thesis), PhD. dissertation, Oxford University, (now online)
http://www.cs.bham.ac.uk/research/projects/cogaff/sloman-1962

A. Sloman, 1971, "Interactions between philosophy and AI: The role of intuition and non-logical reasoning in intelligence", in *Proc 2nd IJCAI,* pp. 209--226, London. William Kaufmann. Reprinted in *Artificial Intelligence,* vol 2, 3-4, pp 209-225, 1971.
http://www.cs.bham.ac.uk/research/cogaff/62-80.html#1971-02
An expanded version was published as chapter 7 of Sloman 1978, available here.

A. Sloman, 1978 *The Computer Revolution in Philosophy,* Harvester Press (and Humanities Press), Hassocks, Sussex.
http://www.cs.bham.ac.uk/research/cogaff/62-80.html#crp

A. Sloman, 1984, The structure of the space of possible minds, in *The Mind and the Machine: philosophical aspects of Artificial Intelligence,* Ed. S. Torrance, Ellis Horwood, Chichester,
http://www.cs.bham.ac.uk/research/projects/cogaff/81-95.html#49a

A. Sloman, 1996, Actual Possibilities, in *Principles of Knowledge Representation and Reasoning* (Proc. 5th Int. Conf on Knowledge Representation (KR '96)), Eds. L.C. Aiello and S.C. Shapiro, Morgan Kaufmann, Boston, MA, pp. 627--638,
http://www.cs.bham.ac.uk/research/cogaff/96-99.html#15

A. Sloman, (2000) "Interacting trajectories in design space and niche space: A philosopher speculates about evolution', in *Parallel Problem Solving from Nature (PPSN VI),* eds. M.Schoenauer, et al. Lecture Notes in Computer Science, No 1917, pp. 3-16, Berlin, (2000). Springer-Verlag.

A. Sloman, 2001, Evolvable biologically plausible visual architectures, in *Proceedings of British Machine Vision Conference,* Ed. T. Cootes and C. Taylor, BMVA, Manchester, pp. 313--322, http://www.cs.bham.ac.uk/research/projects/cogaff/00-02.html#76 http://dx.doi.org/10.5244/C.15.33},

A. Sloman, 2002, The irrelevance of Turing machines to AI, in *Computationalism: New Directions,* Ed. M. Scheutz, MIT Press, Cambridge, MA, pp. 87--127, http://www.cs.bham.ac.uk/research/cogaff/00-02.html#77

A. Sloman and R.L. Chrisley, (2003) "Virtual machines and consciousness', *Journal of Consciousness Studies,* 10(4-5), 113-172.

A. Sloman and R. L. Chrisley, 2005, More things than are dreamt of in your biology: Information-processing in biologically-inspired robots, *Cognitive Systems Research*, Vol 6, No 2, 145--174,, http://www.cs.bham.ac.uk/research/projects/cogaff/sloman-chrisley-cogsys.pdf

A. Sloman, 2008, The Well-Designed Young Mathematician, *Artificial Intelligence,* 172, 18, pp. 2015--2034, Elsevier. http://www.cs.bham.ac.uk/research/projects/cogaff/08.html#807

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

A. Sloman, 2009, Architecture-Based Motivation vs Reward-Based Motivation, *Newsletter on Philosophy and Computers,* American Philosophical Association, 09,1, pp. 10--13, Newark, DE, USA http://www.cs.bham.ac.uk/research/projects/cogaff/misc/architecture-based-motivation.html

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

A.Sloman (2010-2017), A (Possibly) new kind of (non?) Euclidean geometry, based on an idea by Mary Pardoe. http://www.cs.bham.ac.uk/research/projects/cogaff/misc/p-geometry.html

A. Sloman, 2013a "Virtual Machine Functionalism (The only form of functionalism worth taking seriously in Philosophy of Mind and theories of Consciousness)', Research note, School of Computer Science, The University of Birmingham. http://www.cs.bham.ac.uk/research/projects/cogaff/misc/vm-func.html

A. Sloman, 2013b "Virtual machinery and evolution of mind (part 3) Meta-morphogenesis: Evolution of information-processing machinery', in *Alan Turing - His Work and Impact*, eds., S. B. Cooper and J. van Leeuwen, 849-856, Elsevier, Amsterdam.
http://www.cs.bham.ac.uk/research/projects/cogaff/11.html#1106d
Project Web page: https://goo.gl/9eN8Ks

A. Sloman, (2013c), Meta-Morphogenesis and Toddler Theorems: Case Studies, Online discussion note, School of Computer Science, The University of Birmingham, http://goo.gl/QgZU1g

Aaron Sloman (2007-2014)
Unpublished discussion Paper: Predicting Affordance Changes: Steps towards knowledge-based visual servoing. (Including videos).
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/changing-affordances.html

A. Sloman (2015). What are the functions of vision? How did human language evolve? Online research presentation.
http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#talk111

A. Sloman 2017, "Construction kits for evolving life (Including evolving minds and mathematical abilities.)" Technical report (work in progress)
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/construction-kits.html

(An earlier version, frozen during 2016, was published in a Springer Collection in 2017:
https://link.springer.com/chapter/10.1007%2F978-3-319-43669-2_14
in *The Incomputable Journeys Beyond the Turing Barrier*
Eds: S. Barry Cooper and Mariya I. Soskova
https://link.springer.com/book/10.1007/978-3-319-43669-2)

Aaron Sloman, Jackie Chappell and the CoSy PlayMate team, 2006, Orthogonal recombinable competences acquired by altricial species (Blankets, string, and plywood) School of Computer Science, University of Birmingham, Research Note COSY-DP-0601,
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/orthogonal-competences.html

Aaron Sloman and David Vernon. A First Draft Analysis of some Meta-Requirements for Cognitive Systems in Robots, 2007. Contribution to euCognition wiki.
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/meta-requirements.html

P. F. Strawson, *Individuals: An essay in descriptive metaphysics,* Methuen, London, 1959.

Max Tegmark, 2014, *Our mathematical universe, my quest for the ultimate nature of reality,* Knopf (USA) Allen Lane (UK), (ISBN 978-0307599803/978-1846144769)

Arnold Trehub, 1991, *The Cognitive Brain,* MIT Press, Cambridge, MA,
http://people.umass.edu/trehub/

Trettenbrein, Patrick C., 2016, The Demise of the Synapse As the Locus of Memory: A Looming Paradigm Shift?, *Frontiers in Systems Neuroscience,* Vol 88,
http://doi.org/10.3389/fnsys.2016.00088

A. M. Turing, "Computing machinery and intelligence’, *Mind*, 59, 433-460, (1950). (reprinted in E.A. Feigenbaum and J. Feldman (eds) Computers and Thought McGraw-Hill, New York, 1963, 11-35).

A. M. Turing, (1952) "The Chemical Basis Of Morphogenesis", *Phil. Trans. Royal Soc. London B* 237, 237, 37-72.

Note: A presentation of Turing’s main ideas for non-mathematicians can be found in Philip Ball, 2015, "Forging patterns and making waves from biology to geology: a commentary on Turing (1952) ‘The chemical basis of morphogenesis’",
http://dx.doi.org/10.1098/rstb.2014.0218

Barbara Vetter (2011), Recent Work: Modality without Possible Worlds, *Analysis,* 71, 4, pp. 742--754,
https://doi.org/10.1093/analys/anr077

C. H. Waddington, 1957 *The Strategy of the Genes. A Discussion of Some Aspects of Theoretical Biology,* George Allen & Unwin, 1957.

R. A. Watson and E. Szathmary, "How can evolution learn?’, *Trends in Ecology and Evolution,* 31(2), 147-157, (2016).

Weir, A A S and Chappell, J and Kacelnik, A, (2002) Shaping of hooks in New Caledonian crows, *Science,* vol 297, p 981,

L. Wittgenstein (1956), *Remarks on the Foundations of Mathematics*, translated from German by G.E.M. Anscombe, edited by G.H. von Wright and Rush Rhees, first published in 1956, by Blackwell, Oxford. There are later editions. (1978: VII 33, p. 399)

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

---

Installed:15 Apr 2019
Last updated: 30 May 2019 (modified youtube links)
28 Apr 2019: Link to new shorter version of main video. Slight re-formatting.
22 Apr 2019: added abstract, table of contents, better formatting, and a little more detail.

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