Why current AI and neuroscience fail to replicate or explain ancient forms of spatial reasoning and mathematical consciousness

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

Very crudely, David Hume, depicted above, on the left, claimed that there are only two kinds of knowledge:

1. **empirical** knowledge that comes through sensory mechanisms, possibly aided by measuring devices of various kinds; which he called knowledge of "matters of fact and real existence". Deep learning mechanisms can achieve this.
2. 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 ancient knowledge of arithmetic and geometry, which Hume seemed to suggest was no more informative than the bachelor example.
"True by definition" applies to all truths that can be proved using only logic and definitions. An example is "No bachelor uncle is an only child", which can easily be proved from the definitions of "bachelor", "uncle" and "only child".

He famously claimed that if someone claims to know something that is neither of type 1 (empirical) nor of type 2 (mere definitional truths) 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.

Immanuel Kant’s response (1781)
In response to Hume, Immanuel Kant, depicted above, on the right, claimed that there are some important kinds of knowledge that don’t fit into either pair of 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**, e.g. not provable simply using definitions and logic, and
- they also are not based on experience in such a way that future experiences might prove them false, as could happen with most supposed knowledge about the world around us: i.e. they are not **empirical**.
- If they are true there are no possible circumstances in which they could be false, i.e. they are **necessarily** true, not **contingently** true.

Likewise if they are false, e.g. $3 + 5 = 9$, then there are no possible circumstances in which they could be true, i.e. they are **necessarily** false, not **contingently** false.

Example of an empirical, contingent, synthetic proposition: The fact that for millions of years sunset at any location on this planet has always been followed by sunrise does not prove that that will always be the case, or that it is necessarily 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.

In his argument against Hume, Kant drew attention to kinds of mathematical knowledge that 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 he thought was deeply mysterious, since he was unable to explain it), that "5+3=8" is a necessary truth, but not a mere matter of definition, nor derivable from definitions using only logic. (Unlike most philosophers I think such propositions are ambiguous and in one interpretation they conform to Kant’s theories, but not the interpretation Hume gave them.)

Kant 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**.

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, such as Carl G. Hempel who thought Kant had been proved wrong

- because work by Frege, Russell and others had shown that arithmetical knowledge was reducible to logic, and therefore analytic,

- because Kant's claim that Euclid's geometrical axioms and theorems were necessarily true of 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](https://en.wikipedia.org/wiki/Solar_eclipse_of_May_29,_1919)

It is widely, but erroneously, believed that Immanuel Kant's philosophy of mathematics in his Critique of Pure Reason (1781) was disproved by Einstein's theory of general relativity (confirmed by Eddington's observations of the solar eclipse in 1919, establishing that physical space is non-Euclidean).

When I encountered the claim that mathematical knowledge fell into Hume's second category ("relations of ideas", i.e. definitional truths), thereby refuting Kant, I knew from my own experience of finding mathematical proofs e.g. proofs in geometry, that this argument against Kant was fallacious.

My 1962 DPhil thesis (now online) defended a slightly modified version of Kant's claim that many important mathematical discoveries are non-empirical, non-contingent, and non-analytic (i.e. not just logical consequences of axioms and definitions), but did not explain how brains or machines could make such discoveries.

There were several different sorts of argument, but a key part was to generalise the notion taken for granted by many logicians and philosophers since the work of Frege and Russell that sentences in which there are predicates and relations can be construed as applications of functions to arguments, a notion familiar from mathematics.

On this view the sentence "London is a city" applies the predicate "is a city" to the object London, and because that function produces the value TRUE, the statement made by the sentence is true.

Likewise if Jack and Jill are two individuals the sentence "Jack is shorter than Jill" is analysed as applying the two-argument function "... is shorter than ..." which could also be written

```
Shorter(x,y)
```

to the individuals Jack and Jill. If the function produces the value TRUE then what is said is true. If the function produces the value FALSE, then what is said is false, but

```
not(Shorter(x,y))
```

would be true.

In the case of mathematical functions, used in statements like

```
six is greater than three
```

or in standard notation

```
6 > 3
```

nothing that happens to be the case in the physical world can affect whether it is true or false. But in general we do need to look beyond the functions, and the arguments to which they are applied, to discover the value of a function. To check whether there are more blocks than balls on a table
you need to know what blocks are, what balls are and how to compare numbers of objects. You
don’t need that capability in order to decide whether six is greater than three.

So the functions that are used as predicates and relation words in non-mathematical utterances
typically have an implicit additional argument, namely the state of the universe (or a relevant
portion of the universe).

Nevertheless there are many cases where we are able to tell that what is asserted in a proposition
is incapable of being made false by the universe, or incapable of being made true.

Some of those are the cases that Hume described as merely expressing relations of ideas, or
which we can regard as derivable from definitions using only logic.

But Kant’s point was that in other cases what is said is incapable of having a different truth value
(i.e. it is necessarily true or necessarily false) but not because of definitions and and their purely
logical consequences.

An example which is centrally relevant to our ability to use the natural numbers is that we can use
the relationship of two collections being in a one to one correspondence, e.g.

```
[apple banana elephant mouse] == [africa asia europe australia]
```

and

```
[africa asia europe australia] == [water salt wood smoke]
```

to infer that there is also a one-one correspondence if the items in one of the sets are reordered,
and the correspondence will necessarily be preserved if any item in one of the sets is replaced by
another not already in that set.

Moreover, every child learning these number concepts has to come to understand that the relation of
one to one correspondence, is both transitive and symmetric, in order to understand the natural
numbers. Moreover, neither property is merely an empirical property of the relation. (It is not merely
an empirical generalisation.)

Research by Piaget suggests that such understanding does not come until year five or six in most
young humans. So it is not innate, even if Kant is correct in saying that knowledge becomes
non-empirical as a learner’s understanding develops.

Neither is it just an empirical generalisation. Why not? See this draft interview (Sept 2019):

After encountering AI around 1969, I later encountered AI, learnt to program, and hoped to show
how to build a baby robot that could grow up to be a mathematician making discoveries like those
of Archimedes, Euclid, Zeno, etc., and many other deep discoveries made long before the
development of modern logic and formal proof procedures.
I think those mathematical abilities are closely related to the spatial intelligence of pre-verbal human toddlers, and other intelligent animals, e.g. squirrels, elephants, crows, apes, and perhaps octopuses[#] -- whose abilities are not yet replicated in AI/Robotics systems nor explained by current theories in neuroscience or psychology.

[#] https://www.bbc.co.uk/iplayer/episode/m0007snt/natural-world-20192020-5-the-octopus-in-my-house

Insofar as such mathematical discoveries involve necessity or impossibility they cannot be substantiated by mechanisms that collect statistical information and derive probabilities.

This version of Kant’s theory rules out natural and artificial neural nets and related forms of deep learning.

E.g. they cannot learn that something is impossible, such as a largest prime number, or a finite volume bounded by three plane surfaces. I have a large, and steadily growing, collection of examples to be explained by any adequate theory of mathematical consciousness.

I’ll give more examples later.

Many more examples can be found here, and in documents referenced herein: http://www.cs.bham.ac.uk/research/projects/cogaff/misc/impossible.html

Alan Turing’s comments in his PhD thesis on the difference between mathematical intuition and mathematical ingenuity seem to me to echo Kant’s insights, and I suspect (though the evidence is flimsy) that his 1952 paper on chemistry-based morphogenesis (nowadays his “most” cited paper) was at least partly motivated by a search for a new model of computation, combining continuous and discrete components. The most likely location for such a mechanism is sub-neural chemistry, for reasons related to Schrodinger’s analysis in What is life? (1944) of the role of chemistry in reproduction. A few neuroscientists are exploring related ideas (e.g. Seth Grant in Edinburgh).

I’ll present examples of spatial/mathematical reasoning illustrating Kant’s claims. E.g. what sorts of brain mechanisms enable a child to understand that it’s “impossible” to separate linked rings made of impermeable material? Why are you sure that no planar triangle can have one side whose length exceeds the combined lengths of the other two sides?) Current neurally inspired AI mechanisms cannot discover, or even represent, necessity or impossibility, or understand paragraphs like this. Logic-based mechanisms don’t explain what was going on in mathematical brains before the development of logic in the last few centuries, or squirrel brains, or human toddler brains, e.g. this one: http://www.cs.bham.ac.uk/research/projects/cogaff/movies/ijcai-17/small-pencil-vid.webm (Skip the introduction.)

The implications for the current wave of enthusiasm for deep learning are potentially devastating -- but invisible to people who have never studied Kant, or philosophy of mathematics. Which is not to deny that deep learning can be very useful, if used properly.

[xx] A disorganised collection of additional examples can be found here, with links to many more: http://www.cs.bham.ac.uk/research/projects/cogaff/misc/impossible.html (also pdf)
REFERENCES AND LINKS
(To be pruned later.)


Kenneth Craik, 1943, The Nature of Explanation, Cambridge University Press, London, New York Craik drew attention to previously unnoticed problems about biological information processing in intelligent animals. For a draft incomplete discussion of his contribution, see http://www.cs.bham.ac.uk/research/projects/cogaff/misc/kenneth-craik.html

David Deutsch, 1997 The Fabric of Reality, Allen Lane and Penguin Books


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


Immanuel Kant’s *Critique of Pure Reason* (1781) has relevant ideas and questions, but he lacked our present understanding of information processing (which is still too limited) http://archive.org/details/immanuelkantscri032379mbp

Imre Lakatos, *Proofs and Refutations*, Cambridge University Press, 1976,


Piaget, 1981, 1983 Jean Piaget’s last two (closely related) books written with collaborators are relevant, though I don’t think he had good explanatory theories.  
*Possibility and Necessity*  
Vol 1. The role of possibility in cognitive development (1981)  
Vol 2. The role of necessity in cognitive development (1983)  
University of Minnesota Press, Tr. by Helga Feider from French in 1987

(Like Kant, Piaget had deep observations but lacked an understanding of information processing mechanisms, required for explanatory theories.)


Erwin Schrödinger (1944) *What is life?* CUP, Cambridge,
I have an annotated version of part of this book here (also PDF):
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/schroedinger-life.html

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

Frege on the Foundation of Geometry in Intuition *Journal for the History of Analytical Philosophy*
Vol 3, No 6. pp 1-23,
https://jhaponline.org/jhap/issue/view/271


Sloman, A. (1962). *Knowing and Understanding: Relations between meaning and truth, meaning and necessary truth, meaning and synthetic necessary truth* (DPhil Thesis), Oxford University. (Transcribed version online.)

http://www.cs.bham.ac.uk/research/projects/cogaff/07.html#714

http://www.cs.bham.ac.uk/research/projects/cogaff/62-80.html#1965-02

http://www.cs.bham.ac.uk/research/cogaff/62-80.html#1971-02

A slightly expanded version was published as chapter 7 of Sloman 1978, available here.

http://www.cs.bham.ac.uk/research/cogaff/62-80.html#crp

http://www.cs.bham.ac.uk/research/cogaff/62-80.html#1978-02

Aaron Sloman (2012--...), The Meta-Morphogenesis (Self-Informing Universe) Project (begun 2012, with several progress reports, but still work in progress).
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/meta-morphogenesis.html
Aaron Sloman, 2015-18, Some (possibly) new considerations regarding impossible objects, (Their significance for mathematical cognition, current serious limitations of AI vision systems, and philosophy of mind, i.e. contents of consciousness), Online research presentation, 
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/impossible.html

Aaron Sloman, 2013--2018, Jane Austen’s concept of information (Not Claude Shannon’s) 
Online technical report, University of Birmingham, 
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/austen-info.html

http://ceur-ws.org/Vol-1651/

A. Sloman (with help from Jackie Chappell), 2017-8, The Meta-Configured Genome (unpublished) 
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/meta-configured-genome.html

A. Sloman, 2018a, A Super-Turing (Multi) Membrane Machine for Geometers Part 1 
(Also for toddlers, and other intelligent animals) 
PART 1: Philosophical and biological background 
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/super-turing-phil.html

A. Sloman, 2018b A Super-Turing (Multi) Membrane Machine for Geometers Part 2 
(Also for toddlers, and other intelligent animals) 
PART 2: Towards a specification for mechanisms 
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/super-turing-geom.html

Aaron Sloman, 2018c, 
Biologically Evolved Forms of Compositionality 
Structural relations and constraints vs Statistical correlations and probabilities 
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/compositionality.html (also PDF). 
Expanded version of paper accepted for First Symposium on Compositional Structures (SYCO 1) 
Sept 2018 School of Computer Science, University of Birmingham, UK 
http://events.cs.bham.ac.uk/syco/1/

Draft interview (Sept 2019) about how I got into AI and the consequences: 
Notes for an interview to be published in an AI journal.

https://en.wikipedia.org/wiki/Tarski%27s_axioms

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, (1952), 'The Chemical Basis Of Morphogenesis', in *Phil. Trans. R. Soc. London B* 237, 237, pp. 37--72. (Also reprinted(with commentaries) in S. B. Cooper and J. van Leeuwen, EDs (2013)).


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