[Preprint]
A Philosopher-Scientist's View of AI

Aaron Sloman

http://www.cs.bham.ac.uk/~axs

School of Computer Science, University of Birmingham, UK

January 2020
This is available at

Abstract:

Since the "official" launch of AI in 1956, preceded by earlier mathematical and philosophical work by Turing and even earlier practical uses of automated calculators and controllers of various sorts, including mechanical looms and "player pianos", AI has included a wide range of activities, by scientists, engineers, and others with widely varying aims, now mostly dominated by practical, engineering aims. Some of the early work had scientific and philosophical, rather than engineering goals. My own work is of the former type, including use of AI to investigate architectural ideas about how *cognitive* functions interact with motivation, emotions and other varieties of *affect*, addressing old problems in philosophy and the sciences of mind. Some of the difficulties encountered suggest that modelling/replicating ancient mathematical and spatial reasoning abilities of humans and other intelligent species may require digital computers to be enhanced with mechanisms that combine discrete and continuous forms of computation, in ways that nobody understands at present, although sub-neural chemistry-based mechanisms with such a combination are attracting increasing attention. Regarding the recent use of the label "AGI" (Artificial General Intelligence) I have always assumed that AI should accommodate any mechanisms that work, including specialised subsystems common in robotics, so adding a "G" for "general" seems to me to be a misleading publicity gimmick.

Keywords:

AI, Biology, Education, Engineering, Evolution, Kant, Mathematics, Minds, Philosophy, Science

Introduction: Surveys by pioneers

Anyone wishing to understand the scope and methods of AI can still benefit from the vision of some of the pioneers, not because they had a right to limit future developments, but because their work often included useful/powerful ideas that are still important. Minsky's remarkable survey originally written around 1960 [Minsky1963] with over 100 bibliography entries (and still downloadable from his web site[1]) included many such ideas. An important early publication recognizing implications of AI for psychology, was [Miller, Galanter, Pribram 1960]. In 1969, an important, but more methodologically focused, paper on the scope and methods of AI from a philosophical standpoint was [McCarthy Hayes 1969], arguing that *logical* forms of expression are *metaphysically*, *epistemologically*, and *heuristically* adequate forms of representation for intelligent machines. Those ideas are still used by many AI researchers employing logic-based representations, sometimes in hybrid systems, e.g. combined with diagrammatic or probabilistic reasoning, challenging the heuristic adequacy of pure logic-based AI, as in [Sloman1971].

Like many branches of pure and applied science, AI builds on earlier achievements, including designs for calculators and controllers (e.g. automated looms), as well as research in logic, philosophy, psychology, neuroscience, linguistics and social sciences. AI has always included research with both scientific and philosophical aims, although engineering aims and achievements now dominate news about AI. Research fields can also include participants focusing on very different aims, e.g. some more interested in solving old practical problems, some seeking new explanations for old phenomena, and some seeking new practical applications.

This paper focuses on relationships between AI and natural intelligence that are not always acknowledged or widely understood. As indicated above, AI has always been far more than an engineering discipline concerned with making smart machines. For example, Alan Turing, John von Neumann, John McCarthy, Herbert Simon, and Marvin Minsky were as interested in explaining natural intelligence, and, in some cases, answering philosophical questions, as in making smart new machines. I'll also try to show that there are deep explanatory gaps in current AI that generally go unnoticed, and which may require development of new forms of computation. Any attempt to *define* "Artificial Intelligence" should at least allow for the possibility that over time it can change its aims and methods and mechanisms, at least as much as physics has done since ancient attempts to understand such things as levers and planetary motion. Some of this evolution was documented in great detail in Boden(2006).

So, attempting to *define* AI in terms of its current tools and aims at any time is seriously misguided. Despite his breadth of vision, McCarthy was disconcerted by the suggestion in [Sloman Croucher1981] that some intelligent machines will unavoidably have emotions, as a side-effect of design requirements for intelligence with limited knowledge and resources. He thought AI systems should be prevented from having emotions, since that could reduce their reliability. In part this reflects a difference between AI as engineering and AI as science. On that occasion, McCarthy's scientific and philosophical goals were to some extent blunted by his engineering goals. Contrast the broad aims of [Minsky2006].

Debates about what should be included in AI risk being pointless, like some debates about the scope of mathematics: e.g. does mathematics (or AI!) include parts of theoretical computer science? Debates about what should be included in education for young learners are not pointless, however, because restricting diversity in education can have bad effects. Instead of stipulating boundaries it is more important that AI researchers and teachers (like all other researchers and teachers) are clear about their explanatory or practical goals, how they relate both to preceding ideas and possible future developments, and when disagreements about goals are not disagreements about facts. Although individual teachers or schools cannot cover everything relevant, national educational systems should allow, and even encourage, diversity, in order not to hobble future research.

People offering services, products, courses, degrees and certificates should, of course, be clear about the scope of what they are offering, but stipulating *definitions*, especially for research fields, can restrict freedom to explore new directions and may block scientific and engineering advances, as well as constraining educational opportunities for young minds. Historical surveys may limit their scope provided they acknowledge incompleteness, as Boden does Boden(1977) Boden(2006).

Pattern recognition *vs* AI scene analysis

Sometimes disagreements about the scope of AI, or branches of AI, are based on different assumptions about natural intelligence. For example, a strand in AI since its earliest days was *pattern recognition*, designing self-extending programs trained to segment recorded speech into words, phrases, sentences, etc., or 2D visual images into 2D portions with learnt labels attached, e.g. "head", "arm" "finger", "eye", in contrast with the *scene analysis* approach adopted by Clowes and others in the late 1960s, attempting to use 2D input image structures (e.g. lines, line-junctions, and 2D regions) to *derive* descriptions of 3D structures with parts and relationships, on the basis of general principles of projection, or

attempting to derive semantic structures from written or spoken language input using syntactic and semantic theories, sometimes augmented with prior world knowledge. For example, a junction in a 2D image where several lines meet might be interpreted as representing a 3D vertex where several edges meet, some interpreted as convex and some concave, even if that particular configuration of lines and junctions had never previously been encountered in a "training" session [Clowes1971,Clowes1973].[2] A crucial feature of such work was use of context to resolve local ambiguities--important in both language understanding and visual perception. Later research extended the ontologies used by such scene analysis systems.

The 1960s AI work in vision was partly inspired by work in linguistics, e.g. Chomsky (1965), on the relationships between syntactic structures in sentences and semantic descriptions of portions of the world. Clowes was also influenced by ideas in Abercrombie(1960), concerning visual learning in trainee medical researchers learning to derive descriptions of minute physiological structures from images perceived using microscopes. [Gombrich1960] also influenced AI vision researchers.[3]

Proceedings of the 2nd IJCAI https://www.ijcai.org/Proceedings/1971 indicate the breadth AI had achieved by 1971. Alan Turing, Herbert Simon, John McCarthy, and Marvin Minsky had previously recognized its deep relevance for philosophy, including philosophy of mind. Arguing for the heuristic inadequacy of pure logic-based AI, [Sloman1971] offered a new defence of Immanuel Kant's philosophy of mathematics, summarised in [Sloman1965], claiming that some kinds of mathematical knowledge are (a) *non-empirical*, (b) *synthetic/non-analytic* i.e. not based merely on logic and definitions and (c) include *necessary* (i.e. non-contingent) truths.

An important potential (future!) use of AI is explaining why Kant's philosophy of mathematics was broadly correct, especially about discoveries concerning constructions and proofs in Euclidean geometry--contrary to popular opinion among philosophers and mathematicians who think Kant was refuted by Einstein's theory of General Relativity, and Eddington's observation of the 1919 solar eclipse, as argued in [Hempel1945].[4] A future AI system making mathematical discoveries with the features described by Kant, might replicate in a "baby robot" the ability of some baby humans to grow up to be mathematicians. This will require deep advances in biology, neuroscience, and philosophy, as well as AI. [Sloman1962] offered a purely philosophical defence of Kant that could be considerably strengthened by advances in AI replicating human and non-human spatial reasoning competences.

Challenging representational constraints in AI

Despite McCarthy's and Hayes's claims for adequacy of logic-based forms of representation for AI, it is arguable that if ancient mathematicians had been restricted to exploring what can be done using logic they would not have discovered the constructions and proofs in Euclidean geometry that are still in use world wide. Rather than logic being *heuristically* adequate, being restricted to using and thinking with logical forms of representation would have made ancient discoveries much harder than the use of diagrams and diagrammatic constructions (including *imagined* diagrams and constructions). Although some theorem provers can prove theorems in Euclidean geometry e.g. [Gelernter 1964] and the far more sophisticated [Chou, Gao, Zhang .1994], they work only because their designers provided logicised versions of Euclid's axioms and postulates e.g. [Hilbert1899], which the original ancient geometers did not have and did not need: they used other, still unknown, mechanisms for studying spatial structures and processes. [Sloman1962] attempted to defend the validity of ancient diagrammatic forms of reasoning, without reference to AI. Future AI and neuroscience, explaining the roles of sub-neural chemistry in spatial reasoning in brains, may produce a much better defence of Kant.

Similar remarks can be made about mechanical engineers designing or debugging complex machines with 3D interacting parts, such as gears (including worm and pinion gears), pulleys, levers, cables, pistons, etc. Has any engineer ever tried to design a functioning crane or other complex piece of machinery, using only predicate calculus (plus modal logic if needed) to describe the structures, their relationships, their functions, and the processes that can occur during their operation? A computer might be programmed to do it using only logic and arithmetic, but it would not be an accurate model of human design processes, if it replaced all spatial reasoning by numerical and logical reasoning. Moreover, it is very unlikely that replacing all the spatial toys used by pre-verbal children and trying to teach them logic, and formal versions of Euclid's axioms instead, will increase their spatial understanding and future powers as scientists, engineers, architects, or carpenters. Neither would replacing their chemistry-based brains with statistics-based neural nets, if that were possible.

Likewise, I suspect that replicating ancient mathematical discovery processes, and also everyday processes of spatial reasoning, cannot be done on digital computers, whether they use logical theorem provers or artificial neural nets, if brains make essential use of sub-neural chemical processes with a mixture of continuous and discrete changes.[5] Neural net models using statistical evidence to derive probabilities, cannot even *represent* impossibility or necessity, let alone find

proofs of impossibility or necessity. Neither can neural nets in brains, for the same reason, which suggests that understanding ancient mathematical discovery processes will require an understanding of how brains use sub-neural chemical mechanisms, with a mixture of continuous and discrete processing, which I suspect motivated the research reported in [Turing1952], very different from Turing's earlier work on Turing machines [Sloman2002-3].

Some neuroscientists are now investigating sub-neural computations for other reasons, e.g. [Trettenbrein2016, Grant 2018]. Perhaps 22nd Century (or later) AI systems will use mechanisms that are now unimaginable: one of the themes of the Turing-inspired "Meta-morphogenesis" project.[6]

Symbolic, logic-based AI

One of the less-visible, less-fashionable, major strands in current AI inspired by the early work of McCarthy and others is the use of logic, algebra and arithmetic for reasoning and discovery. There are powerful theorem provers used in practical applications such as proving termination of programs, or satisfaction of formal requirements (subject to adequate physical memory and time limits), e.g. https://www.embedded.com/you-think-your-software-works-prove-it/. Such definite conclusions cannot be reached by statistics-based learning systems or any mechanism whose results always have attached probabilities.

When we fully understand human spatial reasoning mechanisms and their roles in ancient mathematical discoveries, we may not be able to replicate them in current computer-based systems, in which case AI will have to be expanded to include the study of biologically evolved computational mechanisms, perhaps including sub-neural chemical computations, a possibility requiring further research. This would render out of date many 20th and 21st century specifications of what AI is.

Finally, this discussion presupposes notions of information and information processing. But I am not referring to Shannon information introduced in (shannon48), which is basically a *syntactic* property. Instead I have been using the much older *semantic* concept of information, used, for example, in Jane Austen's novels a century before Shannon, as explained in http://www.cs.bham.ac.uk/research/projects/cogaff/misc/austen-info.html, a far more important concept for organisms or machines perceiving, interacting, and learning in a complex, richly structured, constantly evolving environment.

POSTSCRIPT
Some ideas about varieties of consciousness as products of biological evolution especially the kinds of ancient mathematical consciousness involved in the deep discoveries reported in Euclid's

Elements, can be found in a draft, incomplete, paper on evolution of very many different forms of consciousness http://www.cs.bham.ac.uk/research/projects/cogaff/sloman-evo-consc-preprint.pdf

---

NOTES

[1] https://courses.csail.mit.edu/6.803/pdf/steps.pdf

[2] A very brief, incomplete, introduction to the ideas can be found in
http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/OWENS/LECT8/node2.html

[3] For more on the work of Max Clowes see the obituary notice and bibliography [Sloman1984 to 2018].

[4] Also at http://www.ditext.com/hempel/geo.html. (Comment added post-publication: Hempel, like many others, failed to realize that the deep content of Kant's philosophy of mathematics was no more refuted by discovery of a large scale non-Euclidean feature of the physical universe than it would have been by the discovery that the surface of a sphere is non-Euclidean. There still remains a problem of explaining what great ancient mathematicians were doing when they thought about Euclidean space and made great discoveries that were *proved*, albeit proved only for Euclidean space, not merely postulated, and were not proved by starting from logical specifications and performing logical derivations, nor by generalising from empirical observations. If Kant ever claimed that any space in which objects can exist, move and interact must be Euclidean then that claim was disproved in the 20th Century, as well as by features of the surface of a ball, on which 2D shapes can move and interact. I am not a Kant expert, but I don't recall him making such an extreme claim.)

[5] I have several partially analysed online examples, and would welcome help with making further progress, e.g.
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/deform-triangle.html,
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/super-turing-geom.html

[6] Also at
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/meta-morphogenesis.html

---

Bibliography
(Formatting unfinished)

Abercrombie, M. (1960). *The Anatomy of Judgement* New York Basic Books.

Boden, M.A. (1977) *Artificial Intelligence and Natural Man* Hassocks, Sussex, Harvester Press. (Second edition 1986. MIT Press)

Boden, M.A. (2006) *Mind As Machine: A history of Cognitive Science* (Vols 1-2) Oxford University Press.

Chomsky, N. (1965)
Aspects of the theory of syntax. Cambridge, MAMIT Press.

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

Clowes, M. (1971), On seeing things. *Artificial Intelligence* 2,1, pp79-116. http://dx.doi.org/10.1016/0004-3702(71)90005-1

Clowes, M. (1973), Man the creative machine: A perspective from Artificial Intelligence research, in J. Benthall, *The Limits of Human Nature* London Allen Lane.

Craik, K. (1943). *The nature of explanation*
London, New York Cambridge University Press.

Gelernter, H. 1964. *Realization of a geometry-theorem proving machine*, in EA. Feigenbaum J. Feldman (Eds) *Computers &Thought* (134-152). New York, McGraw-Hill. (Re-published 1995)

Gombrich, E. H. 1960. *Art and Illusion: A Study in the Psychology of Pictorial Representation*. New York, Pantheon.

Grant, S.G.N, 2010, Synapse molecular complexity and the plasticity behaviour problem *Brain and Neuroscience Advances* 21-7. https://doi.org/10.1177/2398212818810685

Hempel, C. G., 1945 Geometry and Empirical Science, *American Mathematical Monthly52.* http://www.ditext.com/hempel/geo.html Repr., in *Readings in Philosophical Analysis, eds.*, Eds H. Feigl and W. Sellars, New York: Appleton-Century-Crofts, 1949

Hilbert, D, 1899. The Foundations of Geometry (Salt Lake City, Project Gutenberg). Translated 1902 by E.J. Townsend, from 1899 German ed. http://www.gutenberg.org/ebooks/17384

J. McCarthy and P.J. Hayes, (1969). Some philosophical problems from the standpoint of AI. B. Meltzer D. Michie, Eds. **Machine Intelligence 4** Machine Intelligence 4 (463-502). Edinburgh University Press. http://jmc.stanford.edu/articles/mcchay69.html

Miller, Galanter &Pribram .1960

Miller, G., Galanter, E., Pribram, K. (1960). Plans and the Structure of Behaviour Plans and the structure of behaviour.
New York, Holt.

Minsky, L. 1963. Steps toward Artificial Intelligence, in E. Feigenbaum J. Feldman, Eds., *Computers and thought* pp (406-450). New York, McGraw-Hill. https://courses.csail.mit.edu/6.803/pdf/steps.pdf (Originally in *Proceedings of the IRE* 1961)

Minsky, M. L. 2006. *The Emotion Machine.*, New York, Pantheon.

Shannon, C. 1948, A mathematical theory of communication, in *Bell System Technical Journal*, vol 27, pages 379-423 and 623-656.

Sloman, A. 1962. *Knowing and Understanding: Relations between meaning and truth, meaning and necessary truth, meaning and synthetic necessary truth* (DPhil Thesis) Oxford University.
http://www.cs.bham.ac.uk/research/projects/cogaff/62-80.html#1962

Sloman, A. 1965. 'Necessary', 'A Priori' and 'Analytic'. *Analysis* 26, pp 112-16.
http://www.cs.bham.ac.uk/research/projects/cogaff/62-80.html#1965-02

Sloman, A. 1971. Interactions between philosophy and AI: The role of intuition and non-logical reasoning in intelligence, in *Proc 2nd IJCAI* (209-226). London, William Kaufmann.
http://www.cs.bham.ac.uk/research/cogaff/62-80.html#1971-02
Reprinted in *Artificial Intelligence*, vol 2, 3-4, pp 209-225, 1971

Sloman, A. and Croucher, M. 1981. Why robots will have emotions. *Proc 7th Int. Joint Conference on AI* (pp 197-202). Vancouver. http://www.cs.bham.ac.uk/research/cogaff/81-95.html#36

Sloman, A., 1984 to 2018. Experiencing Computation: A Tribute to Max Clowes, Part of this originally in: *New horizons in educational computing* (207-219). Chichester, Ellis Horwood, Series In Artificial Intelligence. Ed. M. Yazdani,
http://www.cs.bham.ac.uk/research/projects/cogaff/81-95.html#61
(Online version with expanded obituary and biography.)

Sloman, A. 2002-1. Architecture-based conceptions of mind, in P. Gärdenfors, K. Kijania-Placek J. Woleński, Eds., *In the Scope of Logic, Methodology, and Philosophy of Science (Vol II)(pp 403-427). Dordrecht, Kluwer. http://www.cs.bham.ac.uk/research/projects/cogaff/00-02.html#57*

Sloman, A. 2002-2. Diagrams in the mind. in M. Anderson, B. Meyer P. Olivier Diagrammatic Representation and Reasoning (pp 7-28). Berlin Springer-Verlag.
http://www.cs.bham.ac.uk/research/projects/cogaff/00-02.html#58

Sloman, A. 2002-3. The irrelevance of Turing machines to AI, in M. Scheutz (Ed), Computationalism: New Directions (p 87-127). Cambridge, MA, MIT Press.
http://www.cs.bham.ac.uk/research/cogaff/00-02.html#77

Sloman, A. 2005. Perception of structure: Anyone Interested? (Research Note COSY-PR-0507). Birmingham, UK. http://www.cs.bham.ac.uk/research/projects/cosy/papers/#pr0507

Sloman, A. 2007-1. What evolved first and develops first in children: Languages for communicating? or Languages for thinking (Generalised Languages: GLs)? Birmingham, UK. Presentation given to Birmingham Psychology department.
http://www.cs.bham.ac.uk/research/projects/cosy/papers/#pr0702

Sloman, A. 2007-2.
Why symbol-grounding is both impossible and unnecessary, and why theory-tethering is more powerful anyway (Research Note COSY-PR-0705). Birmingham, UK.
http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#models

Sloman, A. 2008. Architectural and representational requirements for seeing processes, proto-affordances and affordances., in AG. Cohn, DC. Hogg, R. Möller B. Neumann, {Dagstuhl Seminar Proceedings}, Logic and probability for scene interpretation.
Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany.

http://drops.dagstuhl.de/opus/volltexte/2008/1656 Trettenbrein, P. C., 2016, *The Demise of the Synapse As the Locus of Memory: A Looming Paradigm Shift?* in *Frontiers in Systems Neuroscience* 1088 *http://doi.org/10.3389/fnsys.2016.00088*

Turing, A. M.% 1952. *The Chemical Basis Of Morphogenesis,* in *Phil. Trans. R. Soc. London B*, Vol 237, pp 37-72. *http://dx.doi.org/10.1098/rstb.1952.0012*

About this document ...

This document was generated using the LaTeX2HTML translator Version 2018.3
The translation was initiated on 2020-02-07, and post-edited by hand.