A Philosopher-Scientist's View of AI
(First draft)

Aaron Sloman

Editors:

Abstract

This is a **draft** reaction to a question about whether and how to define AI or AGI. Since the "official" launch of AI in 1956 (preceded by earlier work on mechanical and later electronic calculators and controllers of various sorts), work in AI has included a wide range of activities, by scientists, engineers, and others with very different short term and long term aims, including sub-groups with conflicting interests, aims, methods and theories – mostly dominated by engineering aims, but including research with scientific and philosophical aims. I was originally introduced to AI, in 1969, by Max Clowes (pronounced "Clues") as an activity with scientific goals (extending the resources of psychology and neuroscience) rather than engineering goals. As I learnt more about programming and AI, I used it (with home-grown tools) to investigate architectural ideas about how *cognitive* functions could 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 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,

1. 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 (Minsky, 1963) with over 100 bibliography entries (and still downloadable from his web site[1]) included many such ideas. Another important publication in the first decade of AI, discussing its implications for psychology, was Miller, Galanter, and Pribram (1960).

In 1969, an important, but more methodologically focused, paper on the scope and methods of AI was McCarthy and Hayes (1969), advocating the use of *logical* forms of expression as a *metaphysically*, *epistemologically*, and *heuristically* adequate form 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. As explained below I disagreed with their claim that logic met the requirement of heuristic adequacy.

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

Like many branches of pure and applied science, including physics, chemistry, biology, geology, astronomy, psychology, and the social and economic sciences, AI research builds on earlier research, including designs for calculators and controllers, as well as logic, philosophy, psychology, linguistics and social sciences. It includes research with practical (engineering), scientific and philosophical aims, although a fairly narrow subset of the engineering aims now dominates news about AI.

2. My introduction to AI

I was introduced to AI by Max Clowes around 1969 when I was a young lecturer in philosophy at Sussex University. He was an inspiring researcher in vision using computer models to present new theories about how some aspects of monocular 3D human vision worked, at least in a subset of interesting cases. He kindly lent me books and papers, let me attend his classes and discussed AI and its links to philosophy and psychology with me, and others. He was then one of the leading AI vision researchers, but not primarily trying to solve engineering problems. He was trying to use AI techniques to express explanatory theories about functions and mechanisms of human vision, partly inspired by 19th Century ideas of perception as "controlled hallucination".[2]

The rest of this paper is a case study: reporting aspects of my personal journey triggered by encountering AI under Max's influence. It illustrates aspects of AI that are not always acknowledged: in particular, as hinted above, it is far more than an engineering discipline concerned with making smart machines. I'll also draw attention to some deep gaps in current AI that generally go unnoticed, and which may require development of new forms of computation.

3. AI as science and philosophy

Research fields, including AI, can change their goals, methods, tools, theories and applications as they develop (Sloman, 2002). So, I see no reason to attempt to *define* AI in terms of its current tools, as opposed to *contributing* to it, *using* it and in some cases extending it. E.g. John McCarthy was disconcerted by the suggestion in Sloman and Croucher (1981) 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. McCarthy's scientific and philosophical goals were to some extent blunted by his engineering goals.

Debates about what should be included in AI risk being as pointless as some debates about the scope of mathematics: e.g. does it include parts of theoretical computer science? Some boundary debates concern what topics should be included in an educational system for young learners, which is not usually a pointless question, for example because restricting diversity in education can have bad effects. Instead of stipulating boundaries it may be better to require AI researchers and teachers (like all other researchers and teachers) to be clear about their explanatory or practical goals, how they relate to preceding ideas and possible future developments, and if possible also what the limitations are.

2. I did not learn until later that my colleague in Philosophy at Sussex, Margaret Boden, had already been writing about AI and reviewing AI publications, including some of Max's work.

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. However, historical surveys should be allowed to limit their scope, provided they acknowledge incompleteness, as in Margaret Boden's books on AI and cognitive science (1977; 2006).

4. Pattern recognition *vs* AI scene analysis

In contrast with the *pattern recognition* approach, which attempted to segment images into 2D portions with learnt labels attached, the *scene analysis* approach adopted by Clowes and others attempted to use 2D 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. 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 (Clowes, 1971, 1973).[3] 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. My own later work attempted to show how visual perception could use multiple layers of structure in different "domains", an already obvious feature of language understanding.

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. Max 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. Gombrich (1960) also influenced AI vision researchers.[4][5]

Talking to Max Clowes inspired me to try to use AI to specify some of the ancient mechanisms of mathematical discovery, especially in geometry and topology: I hoped to use AI to provide a new defence of Immanuel Kant's philosophy of mathematics, which I had previously defended at length without any knowledge of AI, in Sloman (1962)[6]. Sloman (1965) gives a very short summary of Kant's claims, namely that there are kinds of mathematical knowledge that are (a) *non-empirical*, (b) *synthetic/non-analytic* i.e. not based merely on logic and definitions and (c) include *necessary* (= non-contingent) truths and falsehoods.

From my own experience of doing mathematics, especially finding constructions and proofs in Euclidean geometry, I thought Kant was broadly correct, contrary to popular opinion among philosophers and mathematicians, at that time, who thought that Kant had been refuted by Einstein's theory of General Relativity, and Eddington's observation of the 1919 solar eclipse, as argued in Hempel (1945)[7], among others. I hoped to use AI to show how Kant's theory might work (which turned out to be a far more difficult task than I had imagined). I was aware that most other AI researchers had more engineering oriented goals – attempting to produce useful, more or less

3. 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`
4. For more on the work of Max Clowes see the obituary notice and bibliography (Sloman, 1984 to 2018).
5. `http://www.cs.bham.ac.uk/research/projects/cogaff/81-95.html#61`
6. `http://www.cs.bham.ac.uk/research/projects/cogaff/62-80.html#1962`
7. Also at `http://www.ditext.com/hempel/geo.html`

intelligent, artefacts, rather than using AI to defend Kant's claims about the nature of mathematical knowledge, which I hoped could one day be justified by an AI program making mathematical discoveries as described by Kant: e.g. a "baby" robot that could grow up to be a mathematician, as a baby human might.

5. Challenging representational constraints in AI

Although I was impressed by much of the content of the paper by McCarthy and Hayes on claimed sufficiency of logic-based forms of representation for AI, I felt that if I had been required to use only logical representation as a young student searching for constructions and proofs in Euclidean geometry, I definitely would not have found the use of logic *heuristically* adequate: it would have made my mathematical tasks much harder than using diagrams and diagrammatic constructions (including *imagined* diagrams and constructions) rather than logical proofs alone. 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 tried designing 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 reasoning.

Strongly encouraged by Max Clowes, I presented some of these ideas, in a paper for IJCAI-71 (Sloman, 1971), disagreeing with McCarthy and Hayes, claiming that for heuristic purposes intelligent (human-like) machines would need what I called *analogical* forms of representation using parts and relations to represent parts and relations—but not necessarily requiring *isomorphism* between representations and things represented. Isomorphism is typically impossible in 2D pictures of 3D scenes: a point that many researchers failed to understand.

As a result of that IJCAI paper, I met and became friendly with McCarthy and was also invited to spend a year (1972-3) in Edinburgh, one of the five or so world-leading AI centres – where I talked to local and visiting AI researchers in several different subfields of AI, including the developers of the Freddy II robot: `https://en.wikipedia.org/wiki/Freddy_II`. In that year I had my brain rewired, and completely changed my way of doing philosophy, as explained in Sloman (1978), now freely available online including afterthoughts.[8]

I hoped to show in a working AI model how ancient mathematical reasoning about spatial structures involves recognition of *impossibilities* and *necessary* connections, which Immanuel Kant (1781) had identified as an essential component of ancient mathematical intelligence. Against McCarthy and Hayes, I also argued that logic-based AI systems could not explain, or replicate, abilities to make ancient mathematical discoveries involving spatial necessity and impossibility, and related competences in young children and other animals with spatial intelligence, such as crows, squirrels and apes.

But the challenge proved much harder than I had expected. I now suspect that replicating ancient mathematical discovery processes, and also everyday processes of spatial reasoning, cannot be done

8. `http://www.cs.bham.ac.uk/research/projects/cogaff/62-80.html#1962`

on digital computers and may, in brains, make use of sub-neural chemical processes with a mixture of continuous and discrete changes.[9]

Neural net models using statistical evidence to derive probabilities, including "deep learning" systems, also fail, but for different reasons: they cannot even *represent* impossibility or necessity. Neither can neural nets in brains, for the same reason, which suggests to me that understanding ancient mathematical discovery processes may require us to understand how brains use sub-neural chemical mechanisms, with their mixture of continuous and discrete processing. Some neuroscientists are now investigating sub-neural computations for other reasons, e.g. Trettenbrein (2016); Grant (2018). Perhaps 22nd Century (or later) AI system will use mechanisms that are now unimaginable: one of the themes of the Turing-inspired "Meta-morphogenesis" project.[10]

6. Symbolic, logic-based AI

One of the less-visible major strands in current AI inspired by the early work of McCarthy and others is based on the use of logic, algebra and arithmetic for reasoning and discovery. Although I am not an expert in the field I am aware of the existence of powerful theorem provers many of which are used in practical applications such as proving that a certain class of programs will always terminate, and that others will never terminate (unless limited by physical space or number of memory locations). Such conclusions cannot be reached by statistics-based learning systems or any mechanism whose results always have attached probabilities. (Perhaps other contributors to this issue have given examples.)

Although some theorem provers can prove theorems in Euclidean geometry e.g. Gelernter (1964) and the far more sophisticated Chou, Gao, and Zhang (1994), they work only because their designers have provided a logicised version of Euclid's axioms and postulates e.g. Hilbert (1899), which the original ancient geometers did not have and did not need: they used other, still unknown, mechanisms for studying spatial structures and processes, a process I tried to defend in Sloman (1962), before I learnt about AI.

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. That possibility, however, requires further research. This would render out of date many 20th and 21st century specifications of what AI is.[11]

References

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

9. 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`
10. `http://www.cs.bham.ac.uk/research/projects/cogaff/misc/meta-morphogenesis.html`
11. My discussion presupposes notions of information and information processing. But I am not referring to Shannon information Shannon (1948), which is basically a syntactic information. Instead I use the much older semantic concept of information, used, for example, in Jane Austen's novels a century before Shannon. `http://www.cs.bham.ac.uk/research/projects/cogaff/misc/austen-info.html`

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: Oxford University Press.

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

Chou, S.-C.; Gao, X.-S.; and Zhang, J.-Z. 1994. *Machine Proofs In Geometry: Automated Production of Readable Proofs for Geometry Theorems*. Singapore: World Scientific.

Clowes, M. 1971. On seeing things. *Artificial Intelligence* 2(1):79–116.

Clowes, M. 1973. Man the creative machine: A perspective from Artificial Intelligence research. In Benthall, J., ed., *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 Feigenbaum, E. A., and Feldman, J., eds., *Computers &Amp; Thought*. New York: McGraw-Hill. 134–152. Re-published 1995 (ISBN 0-262-56092-5).

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

Grant, S. G. N. 2018. Synapse molecular complexity and the plasticity behaviour problem. *Brain and Neuroscience Advances* 2:1–7.

Hempel, C. G. 1945. Geometry and Empirical Science. *American Mathematical Monthly* 52. Repr in Readings in Philosophical Analysis, ed. 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 edition.

Kant, I. 1781. *Critique of Pure Reason*. London: Macmillan. Translated (1929) by Norman Kemp Smith.

McCarthy, J., and Hayes, P. 1969. Some philosophical problems from the standpoint of AI. In Meltzer, B., and Michie, D., eds., *Machine Intelligence 4*. Edinburgh, Scotland: Edinburgh University Press. 463–502.

Miller, G.; Galanter, E.; and Pribram, K. 1960. *Plans and the Structure of Behaviour*. New York: Holt.

Minsky, M. L. 1963. Steps toward Artificial Intelligence. In Feigenbaum, E., and Feldman, J., eds., *Computers and Thought*. New York: McGraw-Hill. 406–450.

Shannon, C. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27:379–423 and 623–656.

Sloman, A., and Croucher, M. 1981. Why robots will have emotions. In *Proc 7th Int. Joint Conference on AI*, 197–202. Vancouver: IJCAI.

Sloman, A. 1962. *Knowing and Understanding: Relations between meaning and truth, meaning and necessary truth, meaning and synthetic necessary truth (DPhil Thesis)*. Ph.D. Dissertation, Oxford University.

Sloman, A. 1965. ‘Necessary’, ‘A Priori’ and ‘Analytic’. *Analysis* 26(1):12–16.

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. Reprinted in *Artificial Intelligence*, vol 2, 3-4, pp 209-225, 1971.

Sloman, A. 1978. *The Computer Revolution in Philosophy*. Hassocks, Sussex: Harvester Press (and Humanities Press). (Now freely available online, with additional notes.).

Sloman, A. 1984 to 2018. Experiencing Computation: A Tribute to Max Clowes. In Yazdani, M., ed., *New horizons in educational computing*. Chichester: Ellis Horwood Series In Artificial Intelligence. 207–219. (Online version with expanded obituary and biography.).

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

Trettenbrein, P. C. 2016. The Demise of the Synapse As the Locus of Memory: A Looming Paradigm Shift? *Frontiers in Systems Neuroscience* 10(88).