

John McCarthy - Some Reminiscences

Aaron Sloman

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

School of Computer Science,
University of Birmingham.

Extended version of Memoir in
AISB Quarterly, No 132, Sept 2011 pp 7-10

<http://www.aisb.org.uk/>

<http://www.aisb.org.uk/membership/>

(Likely to be extended further as memories are revived.)¹

See also: tribute to Marvin Minsky, 2016:

<http://www.cs.bham.ac.uk/research/projects/cogaff/sloman-minsky-tribute.html>

NOTE ADDED 10 Nov 2019

Unfortunately, someone at Stanford University who apparently does not understand how the internet is used by researchers (and many others) was given the task of "prettifying" John's web site. So there is now a new web site here: <http://jmc.stanford.edu/>

However it does not include everything that used to be accessible on his old web site, e.g. the latex source files for his presentations, which are required to make some of the presentations work because the pdf files are generated from latex files that fail to specify "landscape" mode. So many of his presentations are truncated on the right. That would not matter so much if all the old links via <http://www-formal.stanford.edu/jmc/> were still available: someone could then download the latex files modify them and produce a collection of properly formatted PDF slide presentations. (I told John about this and he said he would try to get the presentations fixed, but he died soon after that.) I have left all the old links below, in the hope that someone at Stanford university will set up mechanisms to handle them, because there must be thousands of them on the internet. I would expect the relevant competence to exist in Stanford University Computer Science Department. However, it looks to me as if even Stanford has allowed its web site to be mangled into the style lampooned here: <http://xkcd.com/773/>

In contrast, MIT has had the good sense to leave everything on Marvin Minsky's web site working as it did before he died (though there may have been some additions since then, and various links to other sites are broken):

<http://web.media.mit.edu/~minsky/>

John McCarthy died aged 84 on 24th October 2011. Since then, much has been written about his life and work (e.g. search for his name and "homage", or "obituary"), and no doubt there will be much more. I shall not attempt to emulate or compete with any of the formal obituaries. Instead, I'll offer a few personal recollections and reflections.

There is also much to read on his web site², since he was one of the people who led the way in making everything he wrote freely available to all. It was from him that I learnt to cross out any part of a publisher's copyright agreement that restricted my right to post versions of my papers on my web site. Only one publisher has ever objected (so I withdrew the paper).

One of the most important events in my academic life occurred when Max Clowes [[Clowes Tribute](#)], then the leading AI researcher at Sussex university, introduced me to AI, allowed me to attend his programming tutorials, and gave me things to read, by Simon, Newell, Minsky, McCarthy and others. It quickly became clear that AI was very relevant to old philosophical problems, especially in the papers I read by Minsky and McCarthy. One day Max suggested that I should read the 1969 paper by McCarthy and Hayes, and lent me his copy. I found it very interesting, especially the distinction between metaphysical, epistemological and heuristic adequacy of forms of representation of the world (echoing, but different from, the three kinds of adequacy in [[Chomsky 1965](#)]).

However, I thought the main claim that a logical formalism would suffice for an intelligent machine was mistaken.

This (and much pushing by Max) provoked me into writing a dissenting paper presented at IJCAI 1971, subsequently reprinted in the AI Journal and elsewhere [[Sloman 1971](#)]. The McCarthy/Hayes formalism, first order predicate calculus enhanced with modal operators and fluents, was an example of a "Fregean" form of representation, i.e. one whose syntax used only function/argument structures, first identified as a core part of the structure of ordinary languages by Frege. I could see that logical and other Fregean formalisms (including algebraic formulae, and other mathematical and programming notations) are very useful in many contexts, but I thought it far from obvious that the *only* form of representation required by an intelligent machine is a form of logic. My objection was that we need, and robots will need, different forms of representation for different purposes, and it is sometimes useful (both on epistemological and on heuristic grounds) to employ non-Fregean "analogical" representations (often mistakenly assumed to be isomorphic with what they represent). Examples of the latter include maps, diagrams used in proving geometric theorems, pictures of mechanisms that could be used to reason about causal connections, and 2-D pictures of 3-D scenes. Programming languages that use syntactic ordering of commands to represent the temporal sequence of processing, or use syntactic ordering of items in a data-structure to represent ordering of items in some application domain, include "analogical" representations, in which properties of and relations between parts represent properties of and relations between things represented, though they need not be isomorphic with what they represent, since e.g. a 2-D picture can represent a 3-D object despite being far from isomorphic with it.

In many cases the information in an analogical representation can be re-formulated using a Fregean representation (e.g. specifying locations and orientations in fragments of terrain in a collection of logical assertions rather than a map) yet using the information in that form will often be dreadfully inefficient, because it loses the structural correspondences between representation and what is represented, which can lead to a loss of efficiency during searching, for example. The paper also showed how the notion of a "valid inference" could be extended to include inferences represented by manipulations of spatial representations, as in mathematical reasoning with diagrams - whether in the head, or on paper, in sand, etc.³

I remember that JMC attended my talk and that because we ran out of time we decided to continue the discussion after the final session that day. But I cannot now remember what he said in response! However, Pat Hayes later wrote a critical response [[Hayes 1984](#)].

As a result of writing that paper I was later able to spend a year (1972-3) in Edinburgh, in Bernard Meltzer's group, learning about AI and having my brain rewired, which substantially changed the subsequent direction of my teaching and research. So I owe a very great personal debt to

McCarthy and Hayes.

Thereafter I met JMC occasionally at conferences, e.g. a conference in Edinburgh on "Expert Systems in the Microelectronic Age" organised by Donald Michie in 1979. In the discussion of the ethics of using AI in development of weapons (e.g. Cruise missiles), I remember JMC arguing that a good (and ethical) use of AI would be to enable a missile to fly down the chimney of a munitions factory and destroy it, instead of missing the target and destroying a civilian accommodation block.

In 1980, apparently as a result of reading [Sloman 1978], he invited me to visit Palo Alto, where he had a collection of researchers in AI and philosophy (including Dan Dennett, Pat Hayes, John Haugeland, and possibly one or two others) funded by the Sloan Foundation, meeting and talking about philosophy and AI at the Centre for Advanced Studies in the Behavioral Sciences (CASBS). The other participants came for a year, but family and other commitments meant I could visit for only a month, a very interesting and enjoyable month. Alas, I don't have any detailed recollections of our discussions (though I recall writing comments on a draft version of "Beyond Belief" by Dennett). I also recall sitting at my desk in CASBS with screen and keyboard connected to a computer in the Stanford AI lab, via a modem that made a buzzing noise while transmitting (at about 9k bits/s). I think we used a text editor implemented by Art Samuel. We had neither mouse nor graphics in those days.

I think our next meeting was at IJCAI 1981 in Vancouver, where I presented a paper on emotions in robots, jointly authored with Monica Croucher [Sloman & Croucher 1981]. JMC, like many others since then, had misread the paper as claiming that we should try to give robots emotions. Unlike most others, he objected that that would be a bad idea. I agree with him that if we want our robots to be useful we should try to minimise their emotionality.

However, our paper did not claim that robots *should* have emotions because they are desirable in intelligent systems, a claim that is often made, usually based on fallacious arguments⁴. Instead, we argued that there are resource constraints and knowledge limitations which require mechanisms that sometimes have to react quickly on the basis of partial knowledge, including sometimes overriding other, more intelligent, mechanisms, and that emotional states could result from the operation of such mechanisms. Similar points had been made earlier, by Herbert Simon, in response to Ulric Neisser's claim that only cold cognition, not hot cognition, could be explained or modelled computationally [Simon 1967]. For machines with more knowledge and much greater computational power, such mechanisms might not be necessary, and avoiding such emotional episodes would be preferable. If I ever need brain surgery, I hope I'll have a completely unemotional but highly competent surgeon.

Thereafter, I met JMC from time to time at conferences and during visits to Palo Alto, always finding our conversations interesting and rewarding. On one occasion we discussed limitations of symbol-grounding theory, the latest incarnation of the old philosophical doctrine of concept empiricism, much discussed by past philosophers, including Hume and Berkeley who regarded it as obviously true, and Kant who refuted the theory [Kant 1781], and later philosophers of science who showed that many of the deep concepts of science (e.g. "neutrino", "gene", "magnetic field") could not be derived by abstraction from experience of instances. An alternative, summarised in [Sloman 1985, Sloman 1987], is that such concepts are implicitly defined by their role in an explanatory and predictive theory. But it may be difficult to make the theory rich enough to exclude all unwanted models, since in general any axiomatic system with undefined symbols can have multiple models in different parts of the universe. When I said, in one of our conversations, that

there might be no alternative to using what [Carnap 1947] called "meaning postulates, which link theoretical statements with observable evidence and measurements to help at least partially restrict the possible interpretations (or "tether the theory"⁵), JMC responded that he thought it would always be possible to avoid the need for that by enriching the axioms in the theory. For any intended portion of the world it may be in fact possible to produce a unique identifying description (not including any references to particulars) even if we can never prove that the referent is unique. I don't think he provided an argument that it was always possible: he merely thought it was true. If he is right, then philosophical discussions about the "Twin earth" problem are ill-informed.⁶ This may be one of several philosophical debates in which philosophers wrongly conclude from the fact that they *think* that they can imagine something that they really *can* imagine it, or that it could possibly exist.

On one of my visits I noticed that his car bumper had a sticker saying "More people died at Chappaquiddic⁷ than Three Mile Island"⁸. He was in favour of developing use of nuclear energy. People who have not encountered his commentaries on contemporary debates may enjoy this: <http://www-formal.stanford.edu/jmc/commentary.html>

I think it was during a visit in 1985 that he insisted on taking me for a spin in a two-seater plane that he liked to fly, from San Francisco airport. I was concerned about insurance, but could not find a way to refuse the invitation. The flight was certainly very enjoyable - until he could not make contact when he tried requesting permission to land. Nothing he tried made the radio work, so he decided to head for the airport hoping the controllers would understand what was happening and take charge of the situation. I asked whether the problem could be that the map we had been looking at had altered a switch centrally located above the windscreen. He was sure that switch was irrelevant. I pleaded with him to try it, and he did, and it worked. I guess he had never previously had to use it because the radio was always on. We were both using common sense reasoning, but with different premisses!

We have both always had a strong emphasis on the importance of trying to unpack common sense (which includes a great deal of implicit knowledge and know-how) in order to identify both what needs to be explained by theories of how minds work, and what needs to be implemented in intelligent machines. But there were differences of emphasis: JMC was mostly interested in how we represent, reason about and make use of relatively abstract logically representable, information about the environment or the constraints on some collection of actions [McCarthy 1958], whereas much of my interest was on continuous variation in structures, e.g. surfaces with changing curvature, such as a tea-cup, processes such as rotation of nut on a thread, or straightening a string, or getting a finger through the handle of a mug. That included wanting to understand how humans or machines can discover or prove theorems in Euclidean geometry by manipulating real or imagined spatial configurations, as human mathematicians often do. A lot of progress has been made on JMC's problems. One of the reasons for the limitations of current robots is the lack of progress on the problems concerned with spatial structures and processes, including continuous variation. It's clear that there is considerable development regarding the latter in the first few years of a human's life, but what exactly that development amounts to is far from clear. I don't think the recent emphasis on embodied AI really addresses the problems [Sloman 2009]. I suspect JMC would agree.

I wish I had kept records of our interactions, which were neither frequent nor extended. I enjoyed our conversations and I think he did, though in retrospect I also wish I had pressed him harder on our points of disagreement. I have heard others say they found him difficult to converse with. I did not notice that, possibly because we had enough disagreements to discuss and enough shared assumptions to make the disagreements fruitful. I suspect we both were incompetent at small talk and social chat. Moreover, he had always been interested in and fairly well-read in philosophy but probably did not often meet philosophers who could actually program and were doing AI research. He immediately accepted when I invited him to join me in a two hour special session entitled "A philosophical encounter" at IJCAI 1995, in Montreal. As requested, he submitted a two page summary of his position [McCarthy 1995 2]. Marvin Minsky also accepted the invitation to take part, after some uncertainty as to whether he could attend (which is why there isn't a paper by him in the proceedings). During the discussion session I was amazed when Herbert Simon, who had made important contributions to philosophy, and who was the recipient of the IJCAI research achievement award that year, stood up and objected strongly to the inclusion of a philosophy session at an AI conference, as did Pat Hayes. Herbert Feigenbaum noted that it was the first occasion since the Dartmouth conference that so many of the founders of AI had been in the same room at the same time.

The following year we met at KR96⁹. John gave an invited talk "From Here to Human-level AI" (<http://www-formal.stanford.edu/jmc/human.html> later published as [McCarthy 2007]). Unfortunately I have neither detailed notes nor recollections of his talk, except that I was then, and still remain, uncomfortable with the notion "Human level" since humans are so varied in what they can do, especially if infants, toddlers, people with various brain-abnormalities, quantum physicists, mechanical engineers, trapeze artists, programmers, poets (e.g. Shakespeare) and composers (e.g. Bach), are regarded as humans. He might have rejected this quibble by specifying that what he meant by 'human level AI' was defined by the examples he provided rather than by some notion of normal humanity. More importantly, I have often argued that a deep science of intelligence cannot restrict itself to one case, but needs to investigate the space of possible designs - including those produced by biological evolution - ("design space"), and the space of possible requirements for intelligent systems of various sorts ("niche space") just as a science of chemistry needs to study the generative basis for all molecules and their capabilities.

My paper at the conference [Sloman 1996], was entitled "Actual Possibilities". Looking back at both papers I see that there was some unplanned (and as far as I recall, unnoticed) overlap between our presentations, since, for example, when attempting to characterise "human level" intelligence, JMC considered humanly possible answers to a physics exam question about how to find the height of a building using a barometer, pointing out that, in addition to the "intended" answers, there are many other answers connected with different possible actions that can be performed with a barometer when on or near the building to be measured. That example was very closely related to the main point of my paper, namely that besides being able to perceive, think about, and reason about *actual* entities and situations, normal humans, and some animals, can also perceive and reason about possibilities and constraints. J.J. Gibson's "affordances" [Gibson 1979] are a special subset. I also claimed that some intelligent systems use the ability to recognise and think about "possibility-transducers".

We both concluded that the question of how information about possibilities and constraints on possibilities should be represented in intelligent animals and machines was still open. JMC wrote: *"Since it seems clear that humans don't use logic as a basic internal representation formalism, maybe something else will work better for AI. Researchers have been trying to find this something*

else since the 1950s but still haven't succeeded in getting anything that is ready to be applied to the common sense informatic situation. Maybe they will eventually succeed. However, I think the problems listed in the later sections of this article will apply to any approach to human-level AI." In that paper, he identified important research problems and was careful to phrase them in a manner that could be accepted by researchers who did not share his hope that an intelligent machine could get by using only variants of mathematical logic.

On another occasion, I forget when, I told him I was trying to defend a view that all life involves *information processing*, which contrasts with the mere ability to respond to physical forces. Whereas a non-living object's movements will normally be fully explained by the resultant of all forces acting on the object, like a ball rolling down a helter-skelter (designed artefacts, like mouse-traps, excepted), a living object will typically have a store of chemical energy whose deployment can be turned on or off at least partly under the control of the organism - using not only sensors detecting external states, but also internal sensors detecting needs, etc. He immediately pointed out that that characterisation is not general enough since some animals can use external forces whose deployment they control, e.g. a bird using air-currents to control some of its flight, using only a small amount of its own energy. This required a reformulation of the distinction.

There was a period of at least 10 years, possibly more, when John McCarthy, Marvin Minsky, and other well known AI figures were regular contributors to discussions, including philosophical discussions, on usenet - before that medium was destroyed by the combination of universal access, allowing people with no relevant prior knowledge to pontificate at great length, and worse, the rise of spamming by advertisers. Before that, there was something very valuable about people all over the planet, who had never met, ignoring all distinctions of status, presenting questions, arguments and counter-arguments on both technical problems in AI and also philosophical problems. I presume there are online records of all those interactions. I hope someone will one day produce an edited version without the spam and without the wasteful duplication usually included by those who have not learnt email discussion etiquette. JMC's contributions (and Minsky's) will be a major feature of such an archive.

At a workshop in 2002, Marvin Minsky mentioned McCarthy's 1996 paper "The well-designed child", an early version of [McCarthy 2008]. So I looked it up soon after, liked it very much, and started recommending it to others. It was triggered by his reading [Spelke 1994]. The difference between psychologists who have no experience of the problems of designing working systems and thinkers like JMC, who do have that experience, is very striking. It should especially be read by all those AI researchers working on learning, who need to be reminded that

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

"Animal behavior, including human intelligence, evolved to survive and succeed in this complex, partially observable and very slightly controllable world. The main features of this world have existed for several billion years and should not have to be learned anew by each person or animal."

Let's hope the next 50 years of AI and cognitive science research will be more strongly influenced than the last 50 years by that viewpoint, and the implication that in order to design human-like robots we need a deep understanding of the structure of the world that shaped our evolution, including the evolution of our potential to use logic! A slightly modified version of that paper was published as [McCarthy 2008]. (I was honoured that the journal accepted my "follow on" paper for

the same issue [[Sloman 2008](#)].)

There are many critics of so-called classical AI, or symbolic AI, whose criticisms are based on a very superficial (and usually biased) understanding of the breadth and depth of the problems addressed by AI. For instance, criticisms that early AI systems were mainly concerned with abstract problem solving and planning, as opposed to interacting with a dynamic environment ignore the fact that in the 1960s and early 1970s CPU speeds were measured in kilocycles per second, and memories of a quarter megabyte were rare. If it takes about 20 minutes for a computer vision system to find the rim of a mug in an image, dynamic interaction with the environment is not an option. The look, think, plan, act cycle was the only kind of design that could be used: concurrent visual servoing while using a hand to manipulate an object was out of the question. However I think it is fair to say that the founders, including JMC, seriously underestimated the difficulties of the tasks, and as a result made rash predictions that seriously harmed AI. I've never understood why they did not see the complexities. When Margaret Boden and I wrote about AI we found it obvious that the problems were very deep and would take many years to address [[Boden 1978](#),[Sloman 1978](#)].

There's far more to McCarthy's work than I have touched on. A taste of the breadth of his influence can be found in the recent special issue of the AI journal on his legacy [[Morgenstern McIlraith 2011](#)]. I recently stumbled across an interview by William Aspray [[Aspray McCarthy 1989](#)] that may be of interest to those who would like to know more about the early days of AI at Stanford. I don't believe his goal of basing all of AI on logic can be achieved, and I suspect he also realised that there are problems with that approach. What's important, however, is taking something as powerful as logic and pushing it as far as possible. That will help to identify the problems that need to be solved by combining logic based AI with alternatives. We need an AI educational system that is much less factional and produces graduates with a broad and deep knowledge of the full range of approaches, their strengths, their weaknesses, the problems solved so far, and some of the hard unsolved problems. Alas we have instead a fragmented field with factions that pontificate on the basis of incomplete knowledge of both the problems and the achievements of the various strands. I don't think I heard JMC pontificate in that way, though he did show impatience with discussions that lacked mathematical or logical rigour. Fortunately for me, that did not stop him listening to my half-baked ideas, and commenting on them.

JMC will be remembered with approval by many different researchers, including both engineers trying to solve practical problems, and scientists and philosophers, trying to understand the world and what's possible. I would say he made one huge mistake, whose consequences will go on being harmful for a long time, namely naming the new field "Artificial Intelligence", rather than, for example, "Computational Intelligence", or the more cumbersome "Natural and Artificial Intelligence". The mistake is puzzling insofar as it is clear that from the start his interests went far beyond just trying to make useful machines. He was trying to understand human intelligence as one example of a space of possible forms of intelligence, and he hoped that eventually we'll be able to produce better forms than human intelligence - e.g. intelligent machines unencumbered by emotions. Moreover he understood very well that being that sort of scientist involved also being a philosopher, as shown by the title of the 1969 paper.

However, it was sometimes hard for philosophers to take him seriously, for example when he claimed that a thermostat has desires and beliefs [[McCarthy 1979](#)]. I think that what he was trying to say was right, namely that even in a thermostat we can distinguish what I prefer to call "belief-like" and "desire-like" states, distinguished by what some philosophers have called "direction

of fit"¹⁰. So I was delighted to read this blog entry by David Krane a couple of days ago: "Got the Nest learning thermostat installed today. Neat! Pretty easy install. I had one issue where a wire was pressing down on one of the wire mounts, and that made the Nest think there was a wire plugged in there"¹¹.

Many of his slide presentations are on his web site¹² but don't work because the latex source does not include [landscape] on the top line. So anyone wanting to read the slides will have to fetch the latex files, edit and run. I've reported the problem to a member of his department.

The last few times I met him it was clear that his health was deteriorating, at the AAAI Spring symposium in 2004, and AAAI 2006 in Boston, and most recently at the AAAI conference in August 2011 in San Francisco, when he was in a wheel chair. Alas it was not possible in the circumstances to follow up any of our loose ends.

APPENDICES: AFTERTHOUGHTS

After this had been given to the AISB Newsletter I remembered things I should have included. One of them was mentioned above in connection with the KR96 conference. Future additions will be included as notes here:

NOTE 2 - 25 Sep 2015:

I've found an excellent online video interview: John McCarthy talking to Jeffrey Mishlove:

<https://www.youtube.com/watch?v=Ozipf13jRr4>

NOTE 1 - 6 Jun 2012:

I recently came across a reference to John's work on Elaboration tolerance.¹³ He summarised it thus: "*A formalism is elaboration tolerant to the extent that it is convenient to modify a set of facts expressed in the formalism to take into account new phenomena or changed circumstances. Representations of information in natural language have good elaboration tolerance when used with human background knowledge. Human-level AI will require representations with much more elaboration tolerance than those used by present AI programs, because human-level AI needs to be able to take new phenomena into account.*"

I think this is closely related to but different from a requirement I have been emphasising in discussions of how to evaluate AI models, namely the requirement to *scale out*, as opposed to *scaling up*.

A mechanism can be said to *scale up* if its performance degrades in a reasonable way (e.g. not exponentially) as the size of problem or input increases.

A mechanism has the ability to *scale out* insofar as it can be combined with new mechanisms to perform a variety of tasks. E.g. a vision system that can only be used to label items in images, and cannot be combined with a natural language mechanism to produce descriptions of scenes, and cannot be used in conjunction with a motor control mechanism to control movements by a robot, and cannot provide information for use as the input of an action planner (e.g. information about possibilities and constraints) and cannot be trained to read text, or understand flow-charts or check geometric proofs, might be said to lack the capability to scale out, even if it is extremely efficient at what it can do and scales up well.

McCarthy's notion of "elaboration tolerance" does not apply to a working mechanism but to a formalism. This requires a form of compositional semantics in the language that is not restricted to a narrow semantic domain. I think the ability to scale out requires something like McCarthy's elaboration tolerance in the formalisms used to specify tasks or designs. But it's not yet clear to me whether elaboration tolerance entails the ability to scale out except for a restricted set of linguistic tasks.

References

[Aspray McCarthy 1989]

Aspray, W., & McCarthy, J. (1989, March). An Interview with John McCarthy Conducted by William Aspray. Available from <http://purl.umn.edu/107476>

[Boden 1978]

Boden, M A. Boden, M. A. (1978). *Artificial intelligence and natural man*. Hassocks, Sussex: Harvester Press. (Second edition 1986. MIT Press)

[Cabalar&Santos 2011]

Cabalar, P., & Santos, P. E. (2011). Formalising the fisherman's folly puzzle. *Artificial Intelligence*, 175 (1), 346 - 377. <http://www.sciencedirect.com/science/article/pii/S0004370210000408> (John McCarthy's Legacy)

[Carnap 1947]

Carnap, R. Carnap, R. (1947). *Meaning and necessity: a study in semantics and modal logic*. Chicago: Chicago University Press.

[Chomsky 1965]

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

[Clowes Tribute]

Obituary and biography of Max Clowes

<http://www.cs.bham.ac.uk/research/projects/cogaff/sloman-clowestribute.html>

<http://www.cs.bham.ac.uk/research/projects/cogaff/sloman-clowestribute.pdf>

[Minsky Tribute]

Tribute to Marvin Minsky (2016)

<http://www.cs.bham.ac.uk/research/projects/cogaff/sloman-minsky-tribute.html>

[Gibson 1979]

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin.

[Hayes 1984]

Hayes, P. (1984). Some problems and non-problems in representation theory. In R. Brachman & H. Levesque (Eds.), *Readings in knowledge representation* (pp. 3-22). Los Altos, California: Morgan Kaufmann.

[Kant 1781]

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

[McCarthy 1958]

McCarthy, J. McCarthy, J. (1958). Programs with Common Sense. In *Proceedings Conference on the Mechanization of Thought Processes*. Teddington. (Also in (Minsky, 1968) and online <http://www-formal.stanford.edu/jmc/mcc59.html>)

- [McCarthy 1979]
McCarthy, J. (1979). Ascribing mental qualities to machines. In M. Ringle (Ed.), *Philosophical perspectives in artificial intelligence* (pp. 161-195). Atlantic Highlands, NJ: Humanities Press.
<http://www-formal.stanford.edu/jmc/ascribing/ascribing.html>
- [McCarthy 1990]
McCarthy, J. (1990). *Formalising common sense*. Norwood, New Jersey: Ablex.
- [McCarthy 1995 1]
McCarthy, J. (1995a). Making robots conscious of their mental states. In *AAAI Spring Symposium on Representing Mental States and Mechanisms*. Palo Alto, CA: AAAI. (Revised version:
<http://www-formal.stanford.edu/jmc/consciousness.html>)
- [McCarthy 1995 2]
McCarthy, J. (1995b). What has artificial intelligence in common with philosophy? In *Proc. 14th International Joint Conference on Artificial Intelligence*. <http://www-formal.stanford.edu/jmc/aiphil/aiphil.html>
- [McCarthy 2004]
McCarthy, J. (2004). What is Artificial Intelligence? Stanford University.
<http://www-formal.stanford.edu/jmc/whatisai/whatisai.html>
- [McCarthy 2007]
McCarthy, J. (2007). From Here to Human-Level AI. In *Artificial Intelligence* (Vol. 171, p. 1174-1182). Elsevier. (doi:10.1016/j.artint.2007.10.009, originally in KR96)
- [McCarthy 2008]
McCarthy, J. (2008). The well-designed child. *Artificial Intelligence*, 172 (18), 2003-2014.
<http://www-formal.stanford.edu/jmc/child.html>
- [McCarthy Hayes 1969]
McCarthy, J., & Hayes, P. (1969). Some philosophical problems from the standpoint of AI. In B. Meltzer & D. Michie (Eds.), *Machine Intelligence 4* (pp. 463-502). Edinburgh, Scotland: Edinburgh University Press.
<http://www-formal.stanford.edu/jmc/mcchay69/mcchay69.html>
- [McCarthy 2002]
McCarthy, J., Minsky, M., Sloman, A., Gong, L., Lau, T., Morgenstern, L., et al. (2002). An architecture of diversity for commonsense reasoning. *IBM Systems Journal* 41 (3), 530-539.
<http://www.research.ibm.com/journal/sj41-3.html>
- [Minsky 1968]
Minsky, M. L. (Ed.). (1968). *Semantic Information Processing*. Cambridge, MA: MIT Press.
- [Morgenstern McIlraith 2011]
Morgenstern, L., & McIlraith, S. A. (2011). John McCarthy's legacy. *Artificial Intelligence*, 175 (1), 1-24.
- [Simon 1967]
Simon, H. A. (1967). Motivational and emotional controls of cognition. Reprinted in H. A. Simon (Ed.), *Models of thought* (pp. 29-38). Newhaven, CT: Yale University Press.
- [Sloman 1971]
Sloman, A. (1971). Interactions between philosophy and AI: The role of intuition and non-logical reasoning in intelligence. In *Proceedings 2nd IJCAI* (pp. 209-226). London: William Kaufmann.
(<http://www.cs.bham.ac.uk/research/cogaff/04.html#200407>)
- [Sloman 1978]
Sloman, A. (1978). *The computer revolution in philosophy*. Hassocks, Sussex: Harvester Press (and Humanities Press). Revised edition available from <http://www.cs.bham.ac.uk/research/cogaff/crp>

[Sloman 1985]

Sloman, A. (1985). What enables a machine to understand? In *Proc 9th IJCAI* (pp. 995-1001). Los Angeles: IJCAI. <http://www.cs.bham.ac.uk/research/projects/cogaff/81-95.html#4>

[Sloman 1987]

Sloman, A. (1987). Reference without causal links. In J. du Boulay, D.Hogg, & L.Steels (Eds.), *Advances in artificial intelligence - II* (pp. 369-381). Dordrecht: North Holland. <http://www.cs.bham.ac.uk/research/projects/cogaff/81-95.html#5>

[Sloman 1996]

Sloman, A. (1996). Actual possibilities. In L. Aiello & S. Shapiro (Eds.), *Principles of knowledge representation and reasoning: Proc. 5th Int. Conf. (KR '96)* (pp. 627-638). Boston, MA: Morgan Kaufmann Publishers. <http://www.cs.bham.ac.uk/research/cogaff/96-99.html#15>

[Sloman 2008]

Sloman, A. (2008). The Well-Designed Young Mathematician. *Artificial Intelligence*, 172 (18), 2015-2034. <http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0807>

[Sloman 2009]

Sloman, A. Sloman, A. (2009). Some Requirements for Human-like Robots: Why the recent over-emphasis on embodiment has held up progress. In B. Sendhoff, E. Koerner, O. Sporns, H. Ritter, & K. Doya (Eds.), *Creating Brain-like Intelligence* (pp. 248-277). Berlin: Springer-Verlag. Available from <http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0804>

[Sloman Croucher 1981]

Sloman, A. Croucher, M. Sloman, A., & Croucher, M. (1981). Why robots will have emotions. In *Proc 7th Int. Joint Conference on AI (IJCAI-07)*, pp. 197-202. Vancouver: IJCAI. <http://www.cs.bham.ac.uk/research/cogaff/81-95.html#36>

[Spelke 1994]

Spelke, E. Spelke, E. (1994). Initial knowledge: six suggestions. In *Cognition*, 50 , 431-445.

Footnotes:

¹This version is

<http://www.cs.bham.ac.uk/research/projects/cogaff/12.html#1202>

(Last updated 1 Aug 2016; 21 Nov 2016).

²<http://www-formal.stanford.edu/jmc/>

³Interestingly one of the papers in the recent special issue of AIJ on *John McCarthy's legacy* includes a paper that attempts to show how a type of 3-D spatial puzzle that humans would normally reason about spatially can also be treated in a Fregean formalism - provided that someone has worked out how to express the problem in the appropriate form, which the authors do [[Cabalar & Santos 2011](#)]. Whether and how a machine could do that re-formulation is a hard problem.

⁴As explained in

<http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#cafe04>

⁵"Tethering" was suggested later by Jackie Chappell.

⁶<http://en.wikipedia.org/wiki/Twin\ Earth\ thought\ experiment>

⁷http://en.wikipedia.org/wiki/Chappaquiddick_incident

⁸http://en.wikipedia.org/wiki/Three_Mile_Island_accident

⁹The Fifth International Conference on Principles of Knowledge Representation and Reasoning held in Cambridge Mass. November 1996. When I wrote the original version of this paper for AISB Quarterly I had forgotten about that meeting. This paragraph was added later.

¹⁰<http://en.wikipedia.org/wiki/Direction\ of\ fit>

¹¹Original post by David Krane no longer publicly available (1 Aug 2016). Instead see <https://store.nest.com/>

¹²<http://www-formal.stanford.edu/jmc/slides.html>

¹³<http://www-formal.stanford.edu/jmc/elaboration/elaboration.html>

File translated from T_EX by L^AT_EX, version 3.85.

On 28 Jul 2016, 12:58.