

How to develop a conscious sentient AI

Are we there yet?

David M W Powers

Professor of Computer Science and Cognitive Science
Flinders University, South Australia

powe0093@flinders.edu.au

Abstract

The history of robotics is older than the invention and exploitation of robots. The term ‘robot’ came from the Czech and was first used in a play a century ago. The term ‘robotics’ and the ethical considerations captured by ‘The Three Laws of Robotics’ come from a SciFi author born a century ago. SF leads the way! Similarly, the idea of Artificial Intelligence as a thinking machine goes back to the earliest days of computing, and in this paper we follow some of the key ideas through the work of the pioneers in the field.

We’ve come a long way since then, but are we there yet? Could we now build a conscious sentient thinking computer? What would it be like? Will it take over the world?

1 Introduction

Are robots going to take over the world? Or are they going to destroy it? Are going computers going to become more intelligent than us, then exterminate us? Or will they stop us destroying the world?

Will robots ever reach the point that we can think of mechanical machines as forms of life?

The negative image of artificial beings arguably started with Mary Shelley’s (1880) ‘Frankenstein’ while the word ‘Robots’ was coined for mechanical men in Karel Čapek’s (1920) ‘R.U.R.’ (Rossum’s Universal Robots), which initiated the world domination trope.

Isaac Asimov as a teenager despaired of seeing only negative robot stories so tried to develop positive stories, leading to the formulation of “The Three Laws of Robotics” (1940-1942) that were hardwired into a robot’s

positronic brain – and meant to stop the robot doing any harm. His stories in general revolve around ways in which these laws produce unexpected consequences or can be circumvented. Asimov is also credited with invention of the word ‘robotics’, which partly explains why the ‘three laws’ are so well enshrined in the lore of the field.

Asimov’s (1954 & 1957) detective stories (involving a human-robot detective team) look in particular at how you can use a robot to commit murder despite the laws that are meant to prevent this. The authorized continuations (Allen, 1993-6; Tiedemann, 2000-3) continue in this robot crime vein. Just don’t assume the robot is the criminal!

Asimov went further eventually (1982) with his 0th law – we don’t just have to worry about harming or saving individual humans, but harming or saving humanity... and showing humanity...

Arthur C. Clarke’s (1968) *HAL in 2001: A Space Odyssey*, and the Stanley Kubrik film of the same name, replace a robot with a computer-controlled spaceship and addresses similar issues, with *HAL* becoming iconic in his own right – the problems in *2001* being of a similar character to those Asimov illustrated in his stories about individual robots.

In the mean time, people like Alan Turing were involved in developing the first computers in the 1930s and 40s, and Turing (1936/37) was already thinking about the limits of computation – indeed his characterization of what a computer can do, his tape-based Turing Machine remains the basis for our understanding of what a computer can’t do, or can’t do efficiently (although for

Turing a computer was still a person, and he saw the rewritable tape as like a stack of paper). This extended (Turing, 1950) to considering whether a machine could be considered to think (and now the machine in one room and person in another, originally communicating on sheets of paper, has become the Turing Test, in all its many forms).

Turing made the prediction that by 1950, a computer would be capable of fooling people around 30% of the time in a five minute conversation. In fact, 30% of judges have been fooled in various forms of the Loebner Prize competition, as early as 1997-1998, but an expert has no trouble seeing that they haven't really advanced much beyond Weizenbaum's Eliza program from the 1960s. Indeed they largely rely on the same tricks and techniques (Warwick & Shah, 2016; Weizenbaum, 1966; Shah *et al.*, 2016).

2 Science Fiction and Technology

2.1 Learning

Clarke's (1968 & 1982) *2001* and *2010* SciFi sequence is interesting for the technical effort that went into being accurate about the technology of space travel and artificial intelligence, in particular how to build a sentient artificial intelligence like *HAL*. However, the artificial stupidity *HAL* represents is somewhat more controversial, even "unforgivable" (Lenat, 2001).

Just as Turing (1950) recommended, *HAL* was not programmed, but learned as a child – note that although not normally thought of as a robot, the *HAL* was embodied in a ship with sensors and actuators, providing the essential grounding.

Another more early, obvious and well-known example is Osamu Tezuka's (1952-1968) *Astro Boy*. This *manga* comic book series morphed into the *anime* cartoon series, and its author into a film director. This is the series that defined the genre and captured 40% of Japan's TV audience, perhaps inspiring the modern robotic focus of Japan, and indeed a generation of AI and robotic researchers

around the world. *Astro Boy* is a robot child, who has to learn about the world in every sense.

Powers and Turk (1989) argue that this is how an AI must develop if it is to have a human-like understanding of language and the universe we live in, and cites both *HAL* and *Astro Boy* in this PhD thesis alongside founder of psycholinguistics Jean Piaget – who wrote over twenty books exploring different facets of the question of how children learn to talk about the world, the 'sticky mirror' concept being Powers' formulation of Piaget's ideas about 'reflection' (Piaget, 1923 & 1928).

The importance of Piaget's contribution was developing this area of child development as a scientific field, open to falsification (Popper, 1935) – he also was a student of the philosophy of science. Some of the predictions made by his early theories were indeed overturned by later experiments.

While Turing (1950) mentions some primitive experiments with teaching a computer, Block et al. (1975) tried to provide a more convincing model to explain how a robot could learn English with a more detailed model of language involving syntax and semantics, and using a language learning game to explore the process with humans playing the role of computer – a different twist on the Turing Test.

2.2 Grammar

Powers (1983-84;1989) used both statistical and neural network methods to learn basic grammar, arguing for an unsupervised approach – babies don't have teachers who give them grammar lessons and mark their work. Surprisingly, to some computer scientists, this led to functional words like 'the', 'and' and 'for' being learned before content words like 'cat', 'chased', 'bit' and 'dog'.

But this was not so surprising to psycholinguists who were aware that children's understanding capability led their capability in imitation and production (Brown, 1970). And even older research on reading had

shown how a good reader’s eyes jump from functional word to functional word skipping over the content words (Huey, 1908). The grammatical words, like the prosody, are picked up early, so that a child can recognize not only their mother’s voice at birth (Mehler et al. 1988) but their language (versus another language with a similar sound) and indeed are learning about their linguistic and social environment in the womb (even responding to external stimuli) from at least the start of the second trimester.

2.3 Semantics and Pragmatics

One problem AI faced from the beginning was what to do about meaning, how to represent it, and graphical and network representations became the norm. Conceptual semantic networks (Quillian 1967; Schank, 1975) and in particular Cyc (Lenat et al. 1984-1990) and WordNet (Miller, 1990) are examples of this kind of semantic representation, which can enable matching of concepts due to word similarity.

Cyc was designed to provide the basic commonsense default everyday knowledge that a child could be expected to have, at encyclopaedic scope but lower level. Combined with a natural language capability the idea was that this would allow learning about arbitrary topics. Its aim was specifically to address the limitations of AI systems and provide this bootstrap capability for learning.

Wordnet was designed for linguistic purposes, and being free has become a very popular tool, with versions appearing for other languages (although they are not in general free). This network connects parts to wholes and more general concepts to more specific instantiations (like animals to humans, who then have heads, hands and fingers). A similar hierarchy for verbs provides analogous connections (like ‘go’ having specific specializations for walk, run, fly, etc.). But already this part-of-speech structure makes assumptions that are not the same across languages – for example some languages don’t distinguish nouns and verbs, and even in

English nouns can be verbified, and routinely are when there is no existing or more specialized word for the job.

This allows the idea of similarity to extend beyond noun-noun relations to verb-verb relations, noun-verb relations, and even less direct connections due to shared lexical forms (Yang and Powers, 2005-6). Interestingly systems built on this are more accurate than the average humans (but of course human experts/linguistics can do better).

But it is also possible to induce such information statistically, building a thesaurus or network based on words occurring in similar contexts – the overlap between such a self-organized thesaurus, WordNet and Roget’s Thesaurus is about one third at the 1000 word grouping level (Yang and Powers, 2008).

2.4 Ontology

Learning a bit of syntax or some semantic relations is not the same as learning language – and modern deep neural networks for speech recognition and machine translation are little different. We need to focus on understanding the world as much as understanding language.

Winograd’s (1973) early “SHRDLU” robot-arm natural language system demonstrated the power of going beyond network representations to a world representation, but the traditional representation, and the lack of the ability to learn, still gave rise to grave limitations and a system that was not easy to extend (Winograd & Flores, 1986).

Analogous to this in, in robotics, Brooks (1986-1990) argued that the world is our best representation, that our knowledge is transformed sensory-motor information. Brooks (1986-1990) focused on the “subsumption” idea that you want to have simple sensory motor learn simple behaviours, add those to your arsenal of behaviours, and then learn more at higher and higher levels of abstraction.

Powers (1983,1984) borrowed the word “ontology” from philosophy to distinguish this kind of real world grounding from the idea of

“semantics” that was growing in computer science – connecting words with similar meanings into a network is pretty much like chasing words round a dictionary. Think about looking up Chinese characters in a Chinese dictionary (if you even can) – all you have is more characters.

Powers (1984; Hume, 1984) also defined a software simulated world that allowed humans and robots to interact in a world we would could get complete information about what was happening. This involved simulating some basic (naïve) physics (Hayes, 1979). Concepts could be learned and grounded in this world much more cheaply and safely than using real robots (Sammut, 1985).

Basically, the world and the language were “parsed” using the same bottom up approach to understanding the environment (physical, social and linguistic).

This need for grounding was further popularized by Hanard (1990-1991) as “The Symbol Grounding Problem” although he argued that this could be achieved in a simulation, arguing that simulated thinking is no more thinking than simulated flying is really flying.

But we are now in a world of virtual and augmented reality and sophisticated computer graphics and first person games and interactive fiction. Serious games for educational purposes show that humans are learning real world capabilities through simulations, so why shouldn’t computers. In fact, sophisticated computer games can have millions of non-player characters each with their own intelligence and personality. Moreover, learning language and social skills have been some of the first such educational applications (Milne et al. 2010-18; Powers et al. 2008-16, Stevens et al. 2016).

2.5 Robots vs Animals and Babies

Looking at animals, even insects, we see highly intelligent behaviour (compared to what our computer AIs do). Basically, there is a very tight connection between sensors and actuators that allows simple reflex like

activity, and simple learning can learn environmentally relevant behaviours.

It is not just about understanding the environment but exploring and exploiting the environment, and surviving.

In AI, concept learning is often about understanding nouns – a concept like ‘arch’, or ‘chair’. Originally people approached this descriptively, but in fact it needs to be approached functionally. What is the purpose of an ‘arch’ or a ‘chair’? This has to do with what you do with them – arches support something so something else can go under them, chairs are for sitting on and are designed to support someone while they are sitting.

But this brings us to verbs, like ‘support’ or ‘sit’ or ‘go’ – and many specific kinds of ‘go’ like ‘walk’, ‘run’, ‘hop’, ‘skip’, ‘jump’. And these distinctions have to do with understanding 4D spatiotemporal relationships – so for example this 1984 system learned that ‘jump’ involved passing through a point that was higher than the landing point. Block et al. (1975) famously enunciated the mantra “the verbs are the parts of the robot”.

One part of the human that is of especial importance is the hand, with our opposable thumb. To understand all the many verbs that to do with the hand requires some understanding of a hand – and current robotic manipulators don’t share many of these properties. Thus the structure of the animal or robotic system provides a bias in terms of the kind of concepts it can learn

Also of particular importance is the way human languages handle prepositions like ‘at’, ‘in’, ‘on’ and ‘under’. The basic meanings related to the 3D world, but we can also understand them in 1D (point on a line) or 2D (fly on a wall) or 4D (spatiotemporal context: in progress, on the way to work, at 9 o’clock, in/under/over a minute).

This also relates to an idea of convexity, that leads us to recognize smooth bumpless and holeless shapes ant one level (e.g. person, then at a lower level hand, finger, and knuckle). And of course seeing these shapes in other

contexts gives us a hand of bananas, a finger of land, as well as verb form uses of the words, and so on.

The recognition that metaphor was the basis for our understanding and talking about the world, this ability to recognize things in one context and apply them in an entirely different dimension, is fundamental to the way language works – and the way it worked. Lakoff and Johnson (1980) explored this in extraordinary detail – and this recognition became the basis for the founding of the field of Cognitive Linguistics.

But a lot of progress into understanding how language actually works has focused on prepositions – a concept like ‘in’

Many linguists and psycholinguists have closely monitored their own children’s learning of language (e.g. Brown, 1970), while Deb Roy (2009) goes a step further by bugging the whole house and capturing everything in video to provide a comprehensive corpus of what a child experiences, that can then be used to train a computer or a robot like a child.

By contrast, Luc Steels (1995-2015) initially allowed his community of robots to learn from scratch, inventing their own language – as indeed children do when without parental input.

This suggests that one reason why we haven’t seen intelligent computers/robots is because we haven’t trusted them with robotic sensors and bodies and allowed them to play in the real world with us.

3 Conscience, Consciousness and Emotion

3.1 Emotions, Instincts and Drives

Sloman and Croucher (1981) famously argue that robots must have emotions. In fact, this even goes back to some of the points made by Turing (1950). To survive in the world it needs to be aware of danger, it needs instincts and drives, it needs to know when it is low on energy and needs ‘food’.

3.2 Sentience and Autism

Powers and Turk (1989) argue that babies learn to understand not just language, and the world, but family, culture and society – and multimodal actuators and sensors (internal and external) give us sentience. Learning to see your situation reflected in others’ is also essential for survival and leads to conscience; sequential focus of attention in a vast and vastly parallel array of sensorimotor data is necessary for planning, so you have the essence of consciousness. There’s a saying ‘an intelligent person learns from their mistake, but the wise ones learn from the mistakes of others.

A lack of this understanding of others from their point of view is at the heart of autism, but interestingly something where we can provide training, and AI’s are particular well received and get good results (Milne et al. 2010-2018).

3.3 Neuroscience

Powers (1984) discovered that his computer models learned the close class of functional words first. These words are actually recognized in an area of the brain, Broca’s area, that is adjacent to the part of the motor-cortex that controls the mouth. We have a tendency to focus on recognizing words and structure, and understanding language, but in fact we also have to produce it. Of course, it only takes a print statement to get a computer to print whatever you want – and that is what the current Eliza-like conversational agents tend to do. AI concentrates on the recognition side of the problem.

The ability to look at the brain in action through techniques like fMRI and EEG gives us additional insight into how our cognitive processes work, including helping us to understand what we mean by attention and consciousness. Incorporate an idea of attention into deep artificial neural networks is one of the key innovations that have led to the leap ahead in AI in recent years – but at a cost: it has also spawned the field of Adversarial Machine Learning because an AI system that doesn’t have the same biases and drives as a

human, is not going to learn the same concepts because there is always a bias (Powers, 2016).

Something Winograd and Flory (1986) emphasize is that we should be thinking so much about AI as IA, not Artificial Intelligence, but Intelligence Augmentation. This is closely related to the difference between VR and AR, Virtual Reality versus Augmented Reality. In relation to Human Computer Interface (HCI) we can now with varying success control computers with speech and gesture, including using a Brain Computer Interface (BCI).

There are consumer/games level interfaces that on the whole work well when the electrodes are in appropriate places (Grummett et al. 2015), but often have electrodes placed at easily accessible places (no hair) where they pick up thousands of times more muscle signal than EEG signal, but EMG and EOG and even ECG are signals that convey information in their own right, and an important step is filtering or separating these components (Fitzgibbon et al., 2007-16).

When we are looking at brain signals (EEG not EMG or EOG) what we pick up is often unconscious, and we can learn a lot about the difference between different states of awareness using EEG, and indeed exploit differences in attention to drive a prosthetic device. A very interesting development is Unconscious Computer Interface, where the computer is able to pick up information or intentions even before they become conscious to us, let alone we formulate a conscious thought or command. In particular, we can pick up the brain detecting errors or mismatch, which are important both from a BCI point of view and in understanding the learning process and developing educational and mental health interventions (Blankertz et al. 2002; Groen et al., 2008; Iscan and Nikulin, 2018).

Understanding the role of attention, and trial and error, in learning at the brain level is helpful for us to understand how to build autodidactic AIs that can direct their own attention and learning – a critical element of what we think of as human intelligence.

3.4 Psycholinguistics

Psycholinguistics looks at the whole question of how language is learned (Brown, 1970), and in particular identifies three phrases: recognition/understanding imitation and generation/production.

Generally, a new word or construction will be understood by an infant long before it is used, and when first used it will be in the imitation mode, using or repeating a word or phrase that was just used. In fact, this is how Eliza works. You say “I’ve been having bad dreams” and it responds “Tell me about your dreams”.

Except that this approach doesn’t need to know anything about dreams. Introduce a new word or name X, and it is likely to say “Tell me about X.”

So what about basic sounds, and the simple short morphemes (words, suffixes, prefixes)? When baby tries to say something in its voice, how does that relate to its mother’s and father’s very different voices?

What about a word like ‘mama’? Babies like to experiment, and open and close their mouths while vocalizing. An open vocalization obstructed at the lips will produce the ‘mama’ word (note the vowel and consonant harmony). In pretty well all languages and cultures this is associated with one of the primary caregivers (not always the mother). Of course, other places of obstruction produces different sounds like ‘dada’, ‘nana’ and ‘tata’ (all obstructing with the tongue behind the teeth with different timing relating to the back of the tongue and the vocal chords).

Decent mirrors are a pretty recent invention, but the baby needs to see the sounds it makes mirrored back – it turns out that parents (and siblings and other caregivers) imitate the infant *more* than the other way. This is the original Imitation Game.

Powers and Turk (1989) discuss this and the way closing this feedback loop meant that their model had neurons that reacted both to hearing/seeing the sound (made by someone else) and the motor activity and bone

conduction hearing of their own speech (which thus sounds different to recorded speech).

It was long before such neuron were found, and (not surprisingly) dubbed mirror neurons (Di Pellegrino, et al. 1992; Rizzolatti et al., 1996) and this role in language learning explored (Rizzolatti and Arbib, 1998; Arbib, 2009) – and of course the mirroring is important for more than what you do with your mouth.

This suggests that play is important – a child has to be able to both directly imitate their family members, but also explore and see what results from various muscles – including those in the mouth, and those that make speech sounds.

4 Science Fiction

In 2001 and 2010 Clarke (1968; 1982) gives some details about how HAL learned language. But there is a whole lot that is missing. Recently we have come to put more focus on attention, which is closely related to consciousness. Conscious thought is a succession of different evoked brain patterns that we attend to as our ‘free will’ determines which direction we are going to take them.

Without getting too much into mind-body and free will questions, let’s consider the models in which computers, the AIs of SF, make probabilistic decisions. Something tips the balance one way, and the system follows a path to a particular conclusion, a particular association is dominant, a particular brain rhythm or resonance or synchrony. This combination of external input and internal response is essentially what we mean by conscious thought. For many people thinking is done in their mother tongue, or in a well-learned second language – the percepts our perception gives us associate up through multiple levels to evoke concepts and words, and ideas and plans.

The Sawyer (2009-11) WWW series is interesting in that it assumes the idea of grounding as a necessary condition for emergence of intelligence: in this case

grounding into the real world through a blind girl’s experimental prosthetic.

But we can’t expect an AI to learn language and ontology just by watching, any more than we expect a child to learn what it needs just by being plonked down in front of a television. There needs to be a motor actuation element, stimulus-response. In Sawyer’s story this comes from the AI piggybacking on the girl’s responses and reactions, and eventually their interactions.

Marti Ward (2019, 2020) puts these AI/IA/AR theories to work in an intergalactic SciFi context, and has his AI AI expound on three levels of consciousness. He calls them ‘awake’, ‘aware’ and ‘await’. The highest level involves actively influencing your environment through language.

Awareness is a basic conscious state that we wake up to, but even when asleep we are (subconsciously) aware of things that are happening around us, alert to danger or a baby’s cry. The tight reflex connection between sensing and motor response doesn’t even have to involve the brain, and processes like breathing we are seldom aware of (and control unconsciously during speech for the most part) – and our heartbeat is not at all easy for us to control consciously.

Once patterns start to recur and gain semantic and pragmatic import (sign of food, danger, etc.) these associations gain a life of their own – and are well studied in cats, mice and monkeys as well as humans. We can also associate random (or fortuitous, onomatopoeic, evoked) sounds with these and recognize them in others. These associations can also be chained – and this is the beginning of both language and consciousness.

Of course, the language that works is what is most natural for our sensor system to process, that has the same part-whole spatiotemporal structure as the rest of our sensory-motor world. What starts as an evoked cry as you fall into danger can become a warning cry before you actually fall – or a cry from someone else that your recognize that saves you from falling.

It is important for our AI to have the right biases. If an AI doesn't have the same biases as us, then it is going to tend to learn different things and make decisions in different ways. This is why current AI and Deep Neural Networks are so easy to fool – spawning the new field of Adversarial Machine Learning (trying to find the bias and focus of a learner to trick it into the wrong decision, and then trying to find a way of training it to avoid that pitfall). But this endeavour is doomed to failure if it doesn't get the biases right.

Discussion, Questions and Conclusions

Artificial Intelligence still has a long way to go before it exhibits either the intelligence or the rebelliousness of AIs in Science Fiction stories and films. The best of these stories exhibit important ideas that underlie current approaches to developing real artificial general intelligence, allowing them to learn and be educated and make mistakes.

Some explore important issues relate to the questions of trust, freedom and free will.

Are we willing to take the risk of giving our computational children the same freedoms we give our literal children?

Or are we going to make a race of slaves with no rights or freedoms? And will depriving AIs and robots of these freedoms not actually cause the revolution we are trying to prevent?

Current laws in some countries require that a computer's memories of a person be wiped on request, or after a predetermined time.

Is that something you would do to your children? Are we fundamentally different from that intelligent robot whose development we so cheerfully abort?

Are your prepared to give a robotic entity the opportunity to develop this kind of sentience and consciousness

If we claim to be moral entities on the basis of *cogito ergo sum*, how can we deprive other entities of their rights under this same principle?

Do you believe you are more than a machine and thus have more rights than an intelligent robot? Do you have more rights than a baby

because you are more intelligent and more powerful, or because it is undeveloped and helpless? What about when the robots become more powerful and intelligent than you, when the accelerating pace of AI and robotic development overtakes an increasingly lazy, regressive and self-destructive human society?

Robot Intelligence takes us into a minefield that has social implications way beyond the most obvious ones. We are already seeing AI weaponized in autonomous systems. It is humans that are selfish and immoral, if not overtly irrational, in making war on each other, killing each other, stealing from each other...

Maybe we aren't moral entities but our AIs and robots will be...

Maybe AI is the best chance for humanity to survive...

References

- Roger McBride Allen. 1993. *Caliban*. Ace Books.
- Roger McBride Allen. 1994. *Inferno*. Ace Books.
- Roger McBride Allen. 1996. *Utopia*. Ace Books.
- Isaac Asimov. 1940. Robbie. in *Super Science Stories*, September. Reprinted in 1950 in *I Robot* and in 1982 in *The Complete Robot*.
- Isaac Asimov. 1942. Runaround. In *Astounding Science Fiction*. Reprinted in 1950 in *I Robot*.
- Isaac Asimov. 1950. *I Robot*. Doubleday
- Isaac Asimov. 1954. *The Caves of Steel*. Doubleday.
- Isaac Asimov. 1957. *The Naked Sun*. Doubleday.
- Isaac Asimov. 1982. *The Complete Robot*. Doubleday.
- D. V. M. Bishop and M. J. Hardiman , 2010. "Measurement of mismatch negativity in individuals: A study using single-trial analysis" **Psychophysiology** 47: 697–705.
- Benjamin Blankertz, Christin Schäfer, Guido Dornhege and Gabriel Curio, 2002. Single Trial Detection of EEG Error Potentials: "A Tool for Increasing BCI transmission rates", **International Conference on Neural Networks (ICANN)** pp 1137-1143
- H. D. Block, J. Moulton and G. M. Robinson. 1975. Natural Language Acquisition by a Robot, **Int.Jnl of Man-Machine Studies** 7 pp.571-608
- Rodney Brooks, 1986. "A robust layered control system for a mobile robot". **IEEE Journal of**

- Robotics and Automation.** 2 (1): 14–23. doi:10.1109/jra.1986.1087032. hdl:1721.1/6432.
- Rodney Brooks, 1989. "A Robot that Walks; Emergent Behaviors from a Carefully Evolved Network" **Neural Computation** 1 (2): 253–262, doi:10.1162/neco.1989.1.2.253
- Rodney Brooks. 1990. "Elephants Don't Play Chess", *Robotics and Autonomous Systems*, 6 (1–2): 3–15, doi:10.1016/S0921-8890(05)80025-9.
- Roger Brown. 1970. *Psycholinguistics: Selected Papers*. New York: Free Press.
- Karel Čapek. 1920. *R.U.R.*
- Arthur C. Clarke. 1968. *2001: A Space Odyssey*. Hutchison.
- Arthur C. Clarke. 1982. *2010: Odyssey Two*. Ballantine.
- L Davila, DMW Powers, D Meagher and D Menzies. 1987. Further Experiments in Computer Learning of Natural Language, **Australian Joint Artificial Intelligence Conference**, 458-468
- G Di Pellegrino, Fadiga, L.; Fogassi, L.; Gallese, V.; Rizzolatti, G (1992). "Understanding motor events: a neurophysiological study". **Experimental Brain Research.** 91(1): 176–180. doi:10.1007/bf00230027. PMID 1301372.
- EM Whitham, KJ Pope, SP Fitzgibbon, T Lewis, CR Clark, S Loveless 2007. "Scalp electrical recording during paralysis: Quantitative evidence that EEG frequencies above 20Hz are contaminated by EMG", **Clinical neurophysiology** 118 (8), 1877-1888 389
- SP Fitzgibbon, TW Lewis, DMW Powers, EM Whitham, JO Willoughby, 2013. Surface Laplacian of Central Scalp Electrical Signals is Insensitive to Muscle Contamination, **IEEE Transactions on Biomedical Engineering** 60 (1), 4-9
- SP Fitzgibbon, D DeLosAngeles, TW Lewis, DMW Powers, EM Whitham, 2015. "Surface Laplacian of scalp electrical signals and independent component analysis resolve EMG contamination of electroencephalogram", *International Journal of Psychophysiology* 97 (3), 277-284
- SP Fitzgibbon, D DeLosAngeles, TW Lewis, DMW Powers, TS Grummett, 2016. "Automatic determination of EMG-contaminated components and validation of independent component analysis using EEG during pharmacologic paralysis", *Clinical Neurophysiology* 127 (3), 1781-1793
- Yvonne Groen, Wijers, Albertus A.; Mulder, Lambertus J. M.; Waggeveld, Brenda; Minderaa, Ruud B.; Althaus, Monika 2008. "Error and feedback processing in children with ADHD and children with Autistic Spectrum Disorder", **Clinical Neurophysiology** 119:11 2476-2493 DOI: 10.1016/j.clinph.2008.08.004
- Edmund Huey. 1908. *The Psychology and Pedagogy of Reading*. Reprinted MIT Press 1968.
- Stevan Harnad. 1990. Harnad, Stevan, Hanson, S.J. and Lubin, J. (1991) *Categorical Perception and the Evolution of Supervised Learning in Neural Nets*. Powers, D. M. W. and Reeker, L. (eds.) *Proceedings of the AAAI Spring Symposium on Machine Learning of Natural Language and Ontology*.
- Stevan Harnad. 1991. The symbol grounding problem. *Physica D: Nonlinear Phenomena* 42 (1-3), 335-346
- Patrick I. Hayes. 1979. "The Naive Physics Manifesto", in D. Michie, ed., *Expert Systems in the Micro-electronics Age*, Edinburgh University Press, Edinburgh, Scotland, pages 245ff.
- Zafer İşcan and Vadim V. Nikulin, 2018. Steady state visual evoked potential (SSVEP) based brain-computer interface (BCI) performance under different perturbations, **PLOS ONE**, <https://doi.org/10.1371/journal.pone.0191673>
- Stan A Kuczaj. 1983. *Crib Speech and Language Play* Springer-Verlag, New York, NY.
- Douglas B. Lenat, Borning, Alan; McDonald, David; Taylor, Craig; Weyer, Steven, 1983. "Knoosphere: Building Expert Systems with Encyclopedic Knowledge". *Proceedings of the Eighth International Joint Conference on Artificial Intelligence - Volume 1. IJCAI'83*: 167–169.
- Douglas B. Lenat and John S. Brown, 1984. "Why AM and Eurisko appear to work". *Artificial Intelligence.* 23(3): 269–294. doi:10.1016/0004-3702(84)90016-X.
- Douglas B. Lenat, Prakash, Mayank; Shepherd, Mary. 1986. "CYC: Using Common Sense Knowledge to Overcome Brittleness and Knowledge Acquisition Bottlenecks". **AI Magazine** 6 (4): 65–85. ISSN 0738-4602.
- Douglas B. Lenat and Guha, R. V. 1989. *Building Large Knowledge-Based Systems; Representation and Inference in the Cyc Project* (1st ed.). Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc. ISBN 978-0201517521.
- Douglas B. Lenat and Guha, R. V.; Pittman, Karen; Pratt, Dexter; Shepherd, Mary. 1990. "Cyc: Toward Programs with Common Sense". *Communications of the ACM.* 33 (8): 30–49. doi:10.1145/79173.79176. ISSN 0001-0782.
- Douglas B. Lenat and Guha, R. V. 1991. "The Evolution of CycL, the Cyc Representation

- Language". SIGART Bull. 2 (3): 84–87. doi:10.1145/122296.122308. ISSN 0163-5719.
- Douglas B. Lenat, 2001. "Hal's Legacy: 2001's Computer as Dream and Reality. From 2001 to 2001: Common Sense and the Mind of HAL". Cycorp, Inc.
- Jacques Mehler, Peter Jusczyk, Ghislaine Lambertz, Nilofar Halsted, Josiane Bertoncini, Claudine Amiel-Tison, 1988. *Cognition* 29:2 143-178, Elsevier
- Miller, G. A. (1956). "The magical number seven, plus or minus two: Some limits on our capacity for processing information". *Psychological Review* 6, 81–97. doi:10.1037/h0043158.
- G. A. Miller, R. Beckwith, C. D. Fellbaum, D. Gross, K. Miller. 1990. WordNet: An online lexical database. *Int. J. Lexicograph.* 3:4, pp. 235–244.
- M Milne, MH Luerssen, TW Lewis, RE Leibbrandt, DMW Powers, 2010. "Development of a virtual agent based social tutor for children with autism spectrum disorders", *International joint conference on neural networks (IJCNN)*, 1-9
- M Milne, P Raghavendra, R Leibbrandt, DMW Powers, 2018. "Personalisation and automation in a virtual conversation skills tutor for children with autism", *Journal on Multimodal User Interfaces* 12: 257-269G.
- Jean Piaget. 1923. *The Language and Thought of the Child*, Routledge & Kegan Paul, 1926, translated from *Le Langage et la pensée chez l'enfant*, 1923.
- Jean Piaget. 1925. *The Child's Conception of the World*, Routledge and Kegan Paul, 1928, translated from *La Représentation du monde chez l'enfant*, 1926.
- Karl Popper 1935. *The Logic of Scientific Discovery*. Originally published in 1934 in German as *Logik der Forschung*. Revised and republished 2005 by Routledge.
- David M. W. Powers. 1983. Robot Intelligence. *Electronics Today International*, November pp.15-18
- David M. W. Powers, 1984. Experiments in Computer Learning of Natural Language, *Proc. Aust. Comp. Conf.*, Sydney NSW, 489-500.
- David M. W Powers, 1984. Natural language the natural way. *Computer Compacts* 2 (3-4), 100-109.
- David M. W. Powers and Christopher C. R. Turk. 1989. *Machine Learning of Natural Language*. Springer.
- David MW Powers, R Leibbrandt, D Pfitzner, M Luerssen, T Lewis. 2008. "Language teaching in a mixed reality games environment", Proceedings of the 1st international conference on Pervasive Technologies.
- David M. W. Powers, 2016. Computational Natural Language Learning: ±20years ±Data ±Features ±Multimodal ±Bioplausible. SIGNLL Conference on **Computational Natural Language Learning** pp.1-9 (Founder's invited talk for 20th anniversary)
- Quillian, J.R., "Word Concepts: A Theory and Simulation of some Basic Semantic Capabilities", *Behavioral Science* 12, pp.410-430.
- Giacomo Rizzolatti, Fadiga, Luciano; Gallese, Vittorio; Fogassi, Leonardo. 1996.. "Premotor cortex and the recognition of motor actions". *Cognitive Brain Research*. 3(2): 131–141. doi:10.1016/0926-6410(95)00038-0. PMID 8713554.
- Giacomo Rizzolatti and M.A. Arbib, 1998. "Language within our grasp." *Trends in Neuroscience* 21: 105-167
- Deb Roy. 2009. New horizons in the study of child language acquisition. In *Proceedings of Interspeech*.
- Claude Sammut. 1985. Concept Development for Expert System Knowledge Bases, *The Australian Computer Journal*, Vol. 17(1)49-55.
- Robert J. Sawyer. 2009. *Wake*, Ace
- Robert J. Sawyer. 2010. *Watch*, Ace
- Robert J. Sawyer. 2011. *Wonder*, Ace
- Schank, Roger. 1975. *Conceptual Information Processing*, North Holland Publishing.
- Shah, Huma; Warwick, Kevin; Vallverdú, Jordi; Wu, Defeng. 2016. "[Can machines talk? Comparison of Eliza with modern dialogue systems](#)". *Computers in Human Behavior*. 58: 278–95.
- Mary Shelley, 1880. *Frankenstein*.
- Aaron Sloman and Monica Croucher. 1981. Why Robots will have Emotions. *IJCAI*.
- Luc Steels. 1995. A self-organizing spatial vocabulary. *Artificial life*, 2(3):319–332.
- Luc Steels. 1997. The synthetic modeling of language origins. *Evolution of communication*, 1(1):1–34.
- Luc Steels. 2003. Evolving grounded communication for robots. *Trends in cognitive sciences*, 7(7):308–312.
- Luc Steels. 2015. *The talking heads experiment*. Language Science Press.
- Catherine J Stevens, Bronwyn Pinchbeck, Trent Lewis, Martin Luerssen, Darius Pfitzner, David MW Powers, Arman Abrahamyan, Yvonne Leung, Guillaume Gibert. 2016. [Mimicry and expressiveness of an ECA in human-agent](#)

interaction: familiarity breeds content!
Computational Cognitive Science 2:1

Mark W Tiedemann 2000. *Mirage*. iBooks.

Mark W Tiedemann 2001. *Chimera*. iBooks.

Mark W Tiedemann 2002. *Aurora*. iBooks.

Alan Turing (1937). "On Computable Numbers, with an Application to the Entscheidungsproblem". *Proceedings of the London Mathematical Society*. 2. **42** (1): 230–265. doi://10.1112/plms/s2-42.1.230

Alan Turing. 1950. Computing Machinery and Intelligence, **MIND VOL. LIX NO. 236**.

Marti Ward. 2019. *Casindra Lost*. Amazon. <http://tiny.cc/AmazonCL>

Marti Ward. 2020. *Moraturi Lost*. Amazon. <http://tiny.cc/AmazonML>

Kevin Warwick & Huma Shah (2016) Can machines think? A report on Turing test experiments at the Royal Society, **Journal of Experimental & Theoretical Artificial Intelligence**, **28:6**, 989-1007, doi://10.1080/0952813X.2015.1055826

Weizenbaum, Joseph (1966). "*ELIZA: A Computer Program for the Study of Natural Language Communication between Man and Machine*". **Communications of the ACM**. **9**: 36–35

Terry Winograd, 1972. *Understanding Natural Language*, Academic Press

Terry Winograd and Fernando Flores, 1986. *Understanding Computers and Cognition: A New Foundation for Design*. Intellect Books

Dongqiang Yang and David MW Powers, 2005. *Measuring semantic similarity in the taxonomy of WordNet*. Australian Computer Society.

Dongqiang Yang and David MW Powers, 2006. "Verb similarity in the taxonomy of WordNet." Third International WordNet Congress.

Dongqiang Yang and David MW Powers, 2006.

Dongqiang Yang and David MW Powers, 2008. "Automatic thesaurus construction", Proceedings of the Australasian Conference on Computer Science