The Singularity, or How I Learned to Stop Worrying and Love AI

J. Mark Bishop

Abstract Professor Stephen Hawking recently warned about the growing power of Artificial Intelligence (AI) to imbue robots with the ability to both replicate themselves and to increase the rate at which they get smarter - leading to a tipping point or 'technological singularity' when they can outsmart humans. In this chapter I will argue that Hawking is essentially correct to flag up an existential danger surrounding widespread deployment of 'autonomous machines', but wrong to be so concerned about the singularity, wherein advances in AI effectively makes the human race redundant; in my world AI - with humans in the loop - may yet be a force for good.

1 Background: the 'technological' singularity

It is not often that you are obliged to proclaim a much-loved international genius wrong, but in his alarming prediction regarding Artificial Intelligence and the future of humankind, I believe Professor Stephen Hawking is. Well, to be precise, being a theoretical physicist - in an echo of Schrdinger's cat, famously both *dead and alive* at the same time - I believe the eminent Professor is both *wrong and right* at the same time[1].

Wrong because there are strong grounds for believing that computers will never be able to replicate all human cognitive faculties and *right* because even such emasculated machines may still pose a threat to mankind's future existence; an existential threat, so to speak.

In a television interview on December 2nd 2014 Rory Cellan-Jones asked how far engineers had come along the path towards creating artificial intelligence to which, slightly alarmingly, Professor Hawking replied "*Once humans develop artificial intelligence it would take off on its own and redesign itself at an ever increasing rate.*

J. Mark Bishop
Goldsmiths, University of London, UK, e-mail: m.bishop@gold.ac.uk

[1] This chapter extends a brief essay first published in *Scientia Salon* March 2015.

Humans, who are limited by slow biological evolution, couldn't compete, and would be superseded".

Although warranting headlines that week, such predictions are not new in the world of science and science fiction; indeed my ex-colleague at the University of Reading, Professor Kevin Warwick, made a very similar prediction back in 1997 in his book "March of the Machines". In the book Kevin observed that, even in 1997 there were already robots with the 'brain power of an insect'; soon, he predicted, there would be robots with the brain power of a cat, and soon after that there would be machines as intelligent as humans. When this happens, Warwick predicted, the science fiction nightmare of a 'Terminator' machine could quickly become reality, because these robots will rapidly become more intelligent and superior in their practical skills than the humans that designed and constructed them.

The notion of the singularity (with the accompanying vision of a future mankind subjugated by evil machines) is based on the ideology that *all* aspects of human mentality will eventually be instantiated by an artificial intelligence program running on a suitable computer; a so-called 'Strong AI'[2]. Of course *if* this is possible, accelerating progress in AI technologies - caused both by the use of AI systems to design ever more sophisticated AIs and the continued doubling of raw computational power every two years as predicted by Moore's law - will eventually cause a runaway effect wherein the artificial intelligence will inexorably come to exceed human performance *on all tasks*; the so-called point of [technological] 'singularity' popularised by the Google futurologist Ray Kurzweil [14].

And at the point this 'singularity' occurs, so Warwick, Kurzweil and Hawking suggest, humanity will have effectively been "superseded" on the evolutionary ladder and will be obliged to eek out its autumn days listening to Pink Floyd and gardening; or in some of Hollywood's more dystopian visions, cruelly subjugated or exterminated by 'terminator' machines.

I did not endorse these concerns in 1997 and do not do so now; although I do share - for very different and mundane reasons that I will outline later - the worry that artificial intelligence potentially poses a serious risk to humanity.

2 The humanity gap

There are many reasons why I am sceptical of grand claims made for future computational artificial intelligence, not least empirical. This history of the subject is littered with researchers who have claimed a breakthrough in AI as a result of their research, only for it later to be judged harshly against the weight of society's expectations. All too often these provide examples of what Hubert Dreyfus calls 'the first step fallacy' [10] - undoubtedly climbing a tree takes a monkey a little nearer

[2] Strong AI takes seriously the idea that one day machines will be built that can think, be conscious, have genuine understanding and other cognitive states in virtue of their execution of a particular program; in contrast weak AI does not aim beyond engineering the mere simulation of [human] intelligent behaviour.

the moon, but tree climbing will never deliver a would-be simian astronaut onto its lunar surface.

In previous work I have identified at least three classical *philosophico-technical* problems that illustrate why computational AI has historically failed, and will continue to fail, to deliver on its 'Grand Challenge' of replicating human mentality in all its raw and electro-chemical glory [5] and I will briefly summarise these below.

2.1 Computers lack [phenomenal] consciousness.

In Science and Science Fiction the hope is periodically reignited that a computer system will one day be conscious in virtue of its execution of an appropriate program.; in moves towards this goal: World Scientific Publishing produce the 'International Journal of Machine Consciousness'; the UK funding body EPSRC awarded an Adventure Fund grant of around £500,000 to a team of 'Roboteers and Psychologists' at Essex and Bristol universities, with a goal of instantiating machine consciousness in a 'humanoid-like' robot called Cronos through appropriate computational 'internal modelling' and already a group or researchers at the University of Reading, led by Kevin Warwick, have claimed that robots they have developed are "as conscious as a slug".

Conversely, in an argument entitled *Dancing with Pixies* I demonstrated that if a computer-controlled robot experiences a conscious sensation as it interacts with the world, then an infinitude of 'conscious sensation' must be realised in all objects throughout the universe: in this cup of tea that I am drinking as I write; in the seat that I am sitting as I type, etc etc. If we reject such 'panpsychism', we must reject 'machine consciousness'.

The underlying thread of the 'Dancing with Pixies' reductio [2], [3], [4] &[5] derives from positions originally espoused by Hilary Putnam [30], Tim Maudlin [18] and John Searle [33], with subsequent criticism from David Chalmers [8], Colin Klein [13] and Ron Chrisley [9] amongst others [19].

In the DwP reductio, instead of seeking to secure Putnam's claim that "*every open system implements every Finite State Automaton*" (FSA) and hence that "*psychological states of the brain cannot be functional states of a computer*", I establish the weaker result that, over a finite time window, every open physical system implements the particular execution trace of a Finite State Automaton Q on a specific input vector (I).

That this result leads to panpsychism is clear as, equating FSA Q(I) to a finite computational system that is claimed to instantiate phenomenal states as it executes, and employing Putnam's *state-mapping* procedure to map a series of computational states to any arbitrary non-cyclic sequence of states, we discover identical computational (and *ex hypothesis* phenomenal) states lurking in any open physical system (e.g. a rock); then an infinitude of 'disembodied experience of conscious sensation' [dancing little pixies] are realised in everything ..

Baldly speaking DwP is a simple *reductio ad absurdum* argument to demonstrate that if the assumed claim is true (that an appropriately programmed computer really does instantiate genuine phenomenal states) then panpsychism is true. However if, against the backdrop of our immense scientific knowledge of the closed physical world and the corresponding widespread desire to explain everything ultimately in physical terms, we are led to reject panpsychism, then the DwP reductio leads us to reject the claim that any formal computational processes could instantiate phenomenal consciousness.

2.2 Computers lack genuine understanding.

On the 25th June 2012, Google's 'Deep Learning' technology was reported by the New York Times to have been deployed to categorise unlabelled images. To do this it was "*turned loose on the Internet to learn on its own .. Presented with 10 million digital images found in YouTube videos, what did Googles brain do? What millions of humans do with YouTube: looked for cats*". Le et al conjecture [15] , "The focus of this work is to build high-level, class-specific feature detectors from unlabelled images. For instance, we would like to understand if it is possible to build a face detector from only unlabelled images. This approach is inspired by the neuro-scientific conjecture that there exist highly class-specific neurons in the human brain, generally and informally known as 'grandmother neurons'."

At first sight, if such unsupervised 'Deep Learning' algorithms can learn to classify images of 'faces, 'cats and 'human bodies' from unlabelled images on the internet, then, it would seem that the work must go some way towards demonstrating a genuine form of machine 'understanding' (in addition to potentially arbitrating on the age old philosophical question of 'natural kinds'[3]).

However a thought experiment from the American philosopher John Searle suggests a note of caution. In the now (in)famous *Chinese room argument* Searle demonstrated how it could be possible to program a computer to *appear* to understand, without it the machine *actually* understanding - in his thought experiment Searle famously described how it might be possible to program a computer to communicate perfectly with human interlocutor in a language such as Chinese, without the computer actually understanding anything of the interaction (cf. a small child laughing at a joke she doesn't understand)

Searle illustrates the point by demonstrating how he could follow the instructions of the program - *in computing parlance, we would say Searle is 'dry running' the program* - and carefully manipulating the squiggles and squiggles of the [to him] meaningless Chinese ideographs as instructed by the program, without ever under-

[3] In philosophy the term 'natural kind' is used to refer to a 'natural' grouping contra an artificial one; an objective contra subjective set. There is considerable debate in analytic philosophy about whether there are any natural kinds at all, since even plausible definitions of very familiar species (such as the cat and dog) leave the classification of some exemplars ambiguous.

standing a word of the Chinese ideographic responses the process is so methodically cranking-out.

The essence of the Chines room argument is that **syntax** - *the mere mechanical manipulation [as if by computer] of uninterpreted symbols* - is not sufficient for **semantics** (meaning) to emerge. In this way Searle asserts that no mere computational process can ever bring forth genuine understanding and hence that computation must ultimately fail to fully instantiate mind[4].

It is clear that Searle's argument could just as easily target the claim that a Deep Learning network understands the images it so adroitly processes[5].

2.3 Computers lack [mathematical/creative] insight.

In his book *The Emperor's New Mind*, the Oxford mathematical physicist Sir Roger Penrose deployed Gödel's first incompleteness theorem to argue that, in general, the way mathematicians provide their 'unassailable demonstrations' of the truth of certain mathematical assertions is fundamentally non-algorithmic and non-computational. Gödel's first incompleteness theorem states that *".. any effectively generated theory capable of expressing elementary arithmetic cannot be both consistent and complete. In particular, for any consistent, effectively generated formal theory F that proves certain basic arithmetic truths, there is an arithmetical statement that is true, but not provable in the theory."* The resulting true but unprovable statement $G(\check{g})$ is often referred to as 'the Gödel sentence' for the theory (albeit there are infinitely many other statements in the theory that share with the Gödel sentence the property of being true but not provable from the theory).

Arguments based on Gödel's first incompleteness theorem - *initially from John Lucas [16] [17] - were criticised by Paul Benacerraf [1] then subsequently extended, developed and widely popularised by Roger Penrose [23] [24] [25] [26]* - typically endeavour to show that for any such formal system F, humans can find the Gödel sentence $G(\check{g})$, whilst the computation/machine (being itself bound by F) cannot.

---

[4] See [28] for extended discussion of the Chinese room argument by twenty well known cognitive scientists and philosophers.

[5] This philosophical position recently given additional empirical weight in a critical follow up paper from Szegedy et al [34] in which the researchers demonstrated that "*we can cause the network to misclassify an image by applying a certain imperceptible perturbation, which is found by maximizing the networks prediction error. In addition, the specific nature of these perturbations is not a random artefact of learning: the same perturbation can cause a different network, that was trained on a different subset of the dataset, to misclassify the same input*"; clearly whatever a Deep Learning network is doing when it has learnt to classify unlabelled data, it has not demonstrated that the "specificity of the 'grandmother neuron' could possibly be learned from unlabeled data" or, expressed more colloquially, that such a network could shed light on the human ability to categorise 'cat' images.

In [24] Penrose develops a subtle reformulation of the vanilla argument that purports to show that "the human mathematician can 'see' that the Gödel Sentence is true for consistent F even though the consistent F cannot prove $G(\check{g})$".

NB. A detailed discussion of Penrose's formulation of the Gödelian argument is outside the scope of this chapter - *for a critical introduction see [7] and for Penrose's response see [25]* - here it is simply important to note that although Gödelian-style arguments purporting to show 'computations are not necessary for cognition' have been extensively and vociferously critiqued in the literature (see [29] for a review), interest in them - both positive and negative - still regularly continues to surface (e.g. [6] [35]), with Penrose and Hammeroof asserting that recent developments in physics have gone a long way to proving their case [27].

3 Artificial Intelligence and Artificial Stupidity

Taken together, these above arguments undermine the notion that the human mind can be completely instantiated by mere computations; if correct, although computers will undoubtedly get better and better at many particular tasks - say playing chess, driving a car, predicting the weather etc - there will always remain broader aspects of human mentality that future AI systems will not match. Under this conception there is a 'humanity-gap' between the human mind and mere 'digital computations'; although raw computer power - and concomitant AI software - will continue to improve, the combination of a human mind working alongside a future AI will continue to be more powerful than that future AI system operating on its own; the singularity will never be televised ..

Furthermore it seems to me that without understanding and consciousness of the world and lacking genuine creative [mathematical] insight, any apparently goal directed behaviour in a computer controlled robot is, at best, merely the reflection of a deep rooted longing in its designer. Furthermore, lacking an ability to formulate its own goals, on what basis would a robot set out to subjugate mankind unless, of course, it was explicitly programmed to do so by its [human] engineer? But in that case our underlying apprehension regarding future AI might better reflect the all too real concerns surrounding current Autonomous Weapons Systems, than casually re-indulging Hollywood's vision of the post-human 'Terminator' machine.

Indeed, in my role as one of the AI experts co-opted onto the 'International Committee for Robot Arms Control' (ICRAC), I am particularly concerned by the potential military deployment of robotic weapons systems - *systems that can take decisions to militarily engage without human intervention* - precisely because current AI is still very lacking and because of the underlying potential of poorly designed interacting autonomous systems to rapidly escalate situations to catastrophic conclusions; in my view such systems all too easily exhibit genuine 'Artificial Stupidity'.

I am particularly sceptical that current and foreseeable AI technology can enable autonomous weapons systems to *reliably* comply with extant obligations under International Humanitarian Law; specifically three core obligations: (i) to identify

combatants from non-combatants; (ii) to make nuanced decisions regarding proportionate responses to a complex military situation and (iii) to arbitrate on military or moral necessity regarding when to apply force.

The extreme difficulty in lawfully identifying combatants from non-combatants is powerfully highlighted in the following example from the Human Rights Watch report "*Losing humanity: the case against killer robots*" [12]:

> .. According to philosopher Marcello Guarini and computer scientist Paul Bello,"[i]n a context where we cannot assume that everyone present is a combatant, then we have to figure out who is a combatant and who is not. This frequently requires the attribution of intention." One way to determine intention is to understand an individuals emotional state, something that can only be done if the soldier has emotions. Guarini and Bello continue, "A system without emotion could not predict the emotions or action of others based on its own states because it has no emotional states." Roboticist Noel Sharkey echoes this argument: "Humans understand one another in a way that machines cannot. Cues can be very subtle, and there are an infinite number of circumstances where lethal force is inappropriate."
>
> "For example, a frightened mother may run after her two children and yell at them to stop playing with toy guns near a soldier. A human soldier could identify with the mothers fear and the childrens game and thus recognize their intentions as harmless, while a fully autonomous weapon might see only a person running toward it and two armed individuals. The former would hold fire, and the latter might launch an attack. Technological fixes could not give fully autonomous weapons the ability to relate to and understand humans that is needed to pick up on such cues."

In addition to the technical challenges of meeting obligations under International Humanitarian Law, whenever autonomous systems interact without human supervision there is also a very real danger of catastrophic unintended escalation as underlying problems of 'Artificial Stupidity' forcefully come to bear ..

A light-hearted example demonstrating just how easily autonomous systems can rapidly escalate situations out of control occurred in April 2011, when Peter Lawrence's book 'The making of a fly' was auto-priced upwards by two 'traderbots' competing against each other in the Amazon reseller market-place. The result of this process is that Lawrence can now comfortably boast that his modest scholarly tract - first published in 1992 and currently out of print - was once valued by one of the biggest and most respected companies on Earth at $23,698,655.93 (plus $3.99 shipping).

As stark contrast, in "*Machine gun-toting robots deployed on DMZ*" a report in 'Stars and Stripes' magazine (July 12th, 2010), Jon Rabiroff outlines the following terrifying scenario:

> DEMILITARIZED ZONE, Korea Security along the DMZ has gone high-tech, as South Korea has quietly installed a number of machine gun-armed robots to serve as the first line of defense against the potential advance of North Korean soldiers.
>
> The stationary robots which look like a cross between a traffic signal and a tourist-trap telescope are more drone than Terminator in concept, operated remotely just outside the southern boundary of the DMZ by humans in a nearby command center.
>
> Officials refuse to say how many or where the robots have been deployed along the heavily fortified border between the two Koreas, but did say they were installed late last month and will be operated on an experimental basis through the end of the year.

> South Korean military officials will then decide how many, if any, robots they want complementing the soldiers who man the area adjacent to the 2.5-mile-wide DMZ, which stretches 160 miles across the peninsula.
>
> "The robots are not being deployed to replace or free up human soldiers," said Huh Kwang-hak, a spokesman for Samsung Techwin, the manufacturer of the SGR-1 robot. "Rather, they will become part of the defense team with our human soldiers. Human soldiers can easily fall asleep or allow for the depreciation of their concentration over time," he said. "But these robots have automatic surveillance, which doesnt leave room for anything resembling human laziness. They also wont have any fear (of) enemy attackers on the front lines."
>
> South Korea Ministry of National Defense spokesman Kwon Ki-hyeon said his agency is overseeing the project so he could not comment on theDMZrobot experiment. He referred questions to Samsung Techwin.
>
> Huh said no government officials would talk about the robots: "This experimental project is highly classified."

With armed robot border guards patrolling one side of the DMZ and the potential for North Korea to respond in kind, the darkly dystopian 'Science Fiction' vision of two quasi-autonomous robot armies squaring-up to each other begins to look all too possible; furthermore, given that one of the protagonists is an unstable nuclear armed state, the unintended dangers from a relatively minor military transgression, say a minor border incursion, escalating into a very serious, potentially nuclear, confrontation, begin to look alarmingly possible.

4 The body in question

I believe that the 'Dancing with Pixies' reductio, John Searle's 'Chinese Room Argument' and Roger Penrose' reflections on the non-computable nature of mathematical insight suggest that we need to move away from purely computational explanations of cognitive processes and instead reflect on how meaning, teleology and human creative processes are fundamentally grounded in the human body, society and the world; obliging us, in turn, to take issues of embodiment - the body and our social embedding - much more seriously. And such a *strong* notion of embodiment most certainly cannot be realised by simply co-opting a putative computational creative system into a conventional *tin can robot*[6] ..

As Slawomir Nasuto and I set out in our recent discussion of *Biologically controlled animats*[7] and the so-called *Zombie animals*[8] (two examples carefully chosen to lie at polar ends of the spectrum of possible engineered robotic/cyborg systems), because the induced behavioural couplings therein are not the effect of the intrinsic 'nervous' system's constraints (metabolic or otherwise) at any level, a fortiori,

[6] Whereby an appropriate AI is simply bolted onto a classical robot body and the particular material of that 'embodiment' is effectively unimportant.

[7] Robots controlled by a cultured-array of real biological neurons.

[8] E.g. An animal whose behaviour is 'remotely-controlled', by an external experimenter, say by optogenetics; see also Gradinaru et al [11], who used optogenetic techniques to stimulate neurons selectively, inducing motor behaviour without requiring conditioning.

merely instantiating appropriate sensorimotor coupling is not sufficient to instantiate any meaningful intentional states [21].

On the contrary, in both *Zombie animals* and *Biologically controlled animats* the sensorimotor couplings are actually the *cause* of extrinsic metabolic demands (made via the experimenter's externally directed manipulations). But since the experimenter drives the sensorimotor couplings in a completely arbitrary way (from the perspective of the intrinsic metabolic needs of animal or its cellular constituents), the actual causal relationship between the bodily milieu and the motor actions and sensory readings can never be genuinely and appropriately coupled. Hence our conclusion (ibid) that *only the right type and directionality of sensorimotor couplings* can ultimately lead to genuine understanding and intentionality.

In the light of such concerns, and until the challenges of the CRA, DwP and the mystery of mathematical insight have been *fully* met and the role of embodiment more strongly engaged (such that neurons, brain and body fully interact with other bodies, world and society), I suggest a note of caution in labelling any artificial system as 'strongly intelligent' - *a computational mind* - in its own right; any 'cognition' displayed therein being merely a projection of its engineer's intellect, aesthetic judgement and desire.

5 Conclusion

Without having to fantasise that it has now, (or will ever), reach the level of superhuman intelligence that Professors Warwick and Hawking have graphically warned us of, the all too real-world example of armed robots (as described above) precisely illustrate why it is easy to concur that already current AI systems pose a real 'existential threat' to humanity; the threat of *Artificial Stupidity*. For this reason, in May 2014, members of the International Committee for Robot Arms Control gathered in Geneva to participate in the first multilateral meeting ever held on Autonomous Weapons Systems (LAWS); a debate that continues to this day at the very highest levels of the UN; in a firm, but refracted, echo of Warwick and Hawking on AI - I think we should be very concerned.

Nonetheless, it is equally obvious that even current-state AI has a rich potential to transform society :- from the 'trivial' replacement of tedious human labour (e.g. by controlling robots to clean the floor and mow the lawn); to a more complex new role as an international social facilitator (by helping people communicate more easily by instantaneously offering an approximate translation from one language to another); to helping the State make substantially better use of scarce public resources (e.g. one project that I was personally closely involved with - the UKPLC 'SpendInsight" system - was recently evaluated by the UK National Audit Office and used to identify potential annual saving in the UK National Health Service purchasing budget in excess of £500million per annum [22] [31]; clearly if such savings were realised they would buy a significant number of additional frontline doctors, nurses and drugs).

Already, post 'Lighthill', post 'connectionist winter', post 'Terminator blues', the recent practical realisation of ambitious *real world, nouveau AI, machine learning, big-data systems* is tempting the engineered geek-in-me with too many lucrative, new and seductive headline images; any one of which could so easily prompt me to fall headstrong-in-love with AI again ..

Mark Bishop is Professor of Cognitive Computing at Goldsmiths, University of London and Director of The Goldsmiths Centre for Intelligent Data Analytics (TCIDA). He was Chair of the AISB, the UK Society for Artificial Intelligence and the Simulation of Behaviour, [2010-2014] and currently serves on the International Committee for Robot Arms Control.

References

1. Benacerraf P (1967) God, the Devil & Godel. Monist 51: 9-32
2. Bishop JM (2002) Dancing with Pixies: strong artificial intelligence and panpyschism. in: Preston J, Bishop JM (eds) (2002) Views into the Chinese Room: New Essays on Searle and Artificial Intelligence, Oxford University Press, Oxford
3. Bishop JM (2005) Can computers feel? The AISB Quarterly 199: 6
4. Bishop JM (2009) Why Computers Can't Feel Pain. Minds and Machines 19(4): 507-516
5. Bishop JM (2009) A Cognitive Computation fallacy? Cognition, computations and panpsychism. Cognitive Computation 1(3): 221-233
6. Bringsjord S, Xiao H (2000) A refutation of Penrose's Gödelian case against artificial intelligence. J. Exp. Theoret. AI 12: 307-329
7. Chalmers DJ (1995) Minds, Machines And Mathematics: a review of 'Shadows of the Mind' by Roger Penrose. PSYCHE 2(9).
8. Chalmers DJ (1996) The Conscious Mind: In Search of a Fundamental Theory, Oxford University Press, Oxford
9. Chrisley R (2006) Counterfactual computational vehicles of consciousness. Toward a Science of Consciousness April 4-8 2006, Tucson Convention Center, Tucson Arizona
10. Dreyfus, HL (2012) A history of first step fallacies, Minds and Machines, 22 (2 - special issue 'Philosophy of AI' ed. Vincent C. Mller): 87-99
11. Gradinaru, V., Thompson, K.R., Zhang, F., Mogri, M., Kay, K., Schneider, M.B., Deisseroth, K., (2007), Targeting and readout strategies for fast optical neural control in vitro and in vivo. J. Neurosci 26:27(52): 14,23114,238.
12. Human Rights Watch (2012) Losing humanity: the case against killer robots, Human Rights Watch Report (Nov.19 2012)
13. Klein C (2004) Maudlin on Computation (working paper)
14. Kurzweil R (2005) The singularity is near: When humans transcend biology, Viking, London.
15. Le Q, Ranzato MA, Monga R, Devin M, Chen K, Corrado G, Dean J, Ng A (2012) Building high-level features using large scale unsupervised learning. Proc. 29th International Conference in Machine Learning, Scotland.
16. Lucas JR (1962) Minds, Machines & Godel. Philosophy 36: 112-127
17. Lucas JR (1968) Satan Stultified: A Rejoinder to Paul Benacerraf. Monist 52: 145-158
18. Maudlin T (1989) Computation and Consciousness. Journal of Philosophy (86): 407-432
19. Minds and Machines (1994) Special Issue: What is Computation? Minds and Machines 4(4)
20. Mnih V, Kavukcuoglu, K, Silver, D, Rusu AA, Veness J, Bellemare MG, Graves A, Riedmiller M, Fidjeland AK, Ostrovski G, Petersen S, Beattie C, Sadik A, Antonoglou I, King H, Kumaran D, Wierstra D, Legg S, Hassabis D (2015). "Human-level control through deep reinforcement learning". Nature 518, 529533.

21. Nasuto, S.J. and Bishop, J.M., (2013), Of (zombie) mice and animats. In: Mller, V. C. (ed.) Philosophy and Theory of Artificial Intelligence. Studies in Applied Philosophy, Epistemology and Rational Ethics (5): 85-106, Springer Berlin Heidelberg.
22. National Audit Office (2011) The Procurement of Consumables by NHS Hospital Trusts [online], http://www.nao.org.uk/publications/1011/nhs_procurement.aspx (accessed 16 July 2012)
23. Penrose R (1989) The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics. Oxford University Press, Oxford
24. Penrose R (1994) Shadows of the Mind: A Search for the Missing Science of Consciousness. Oxford University Press, Oxford
25. Penrose R (1996) Beyond the Doubting of a Shadow: a reply to commentaries on 'Shadows of the Mind'. PSYCHE 2(23)
26. Penrose R (1997) On Understanding Understanding. International Studies in the Philosophy of Science 11(1): 7-20
27. Penrose R, Hammeroff S (2014) Consciousness in the universe: A review of the Orch OR theory, Physics of Life Reviews 11(1): 39-78
28. Preston J, Bishop JM (eds) (2002) Views into the Chinese Room: New Essays on Searle and Artificial Intelligence. Oxford University Press, Oxford
29. Psyche (1995) Symposium on Roger Penrose's Shadows of the Mind. PSYCHE 2 http://psyche.cs.monash.edu.au/psyche-index-v2.html
30. Putnam H (1988) Representation and Reality. Bradford Books, Cambridge MA
31. Roberts PJ, Mitchell RJ, Ruiz VF, Bishop JM (2014) Classification in e-procurement. Int. J. Applied Pattern Recognition, 1(3): 298314
32. Searle J (1980) Minds, Brains and Programs. Behavioral and Brain Sciences 3(3): 417-457
33. Searle J (1990) Is the Brain a Digital Computer? Proceedings of the American Philosophical Association (64): 21-37
34. Szegedy C, Zaremba W, Sutskever I, Bruna J, Erhan D, Goodfellow IJ, Fergus, R (2013) Intriguing properties of neural networks. Proc. Inc. Conf. International Conference on Learning Representations.
35. Tassinari RP, D'Ottaviano IML (2007) Cogito ergo sum non machina! About Gödel's first incompleteness theorem and Turing machines, CLE e-Prints 7(3)