Some notes on super-intelligence

David Bell

Abstract. Most people do not know much about the details of artificial intelligence (AI). However, if AI is potentially an imminent danger it would be prudent to know something about it, or simply to consider, and even if possible try to influence the directions AI should focus upon. Some 'prophets' say that if a machine's intelligence exceeds that of humans, we cannot possibly predict whether humans will be tolerated as pets by it, assisted by it in amazing ways, or wiped out by it. Some suggest that technology holds the promise of things like salvation and well-being, and might be the means whereby our race can persist for ever. This raises issues for traditional Christian views.

Despite disappointments in the past, AI has steadily made impressive progress. It now contributes very significantly to modern life, and there is realistic promise of more benefits to come. It is often hidden away in products and used in various important activities and professions. However, many respected technologists and scientists see AI as presenting a more serious problem for mankind than even global warming and Islamic State.

This talk presents an informal, personal look at the relevant scientific and technical state of the art and at hurdles to be overcome, and assesses briefly some of the theological problems that could be raised.

1. Introduction

The terms *Super-intelligence* or *Artilect* can be used to refer to artificial agents that surpass humans in our mental capabilities. Stephen Hawking and others have expressed concern recently about threats they might bring to humans: 'Once humans develop artificial intelligence, it would take off on its own and re-design itself at an ever increasing rate. The development of full artificial intelligence could spell the end of the human race.' The *Singularity* in this context is the point at which super-intelligence emerges, if it does emerge.

So – could or will there be thinking machines that take over?

Let us get a working definition: 'By a superintelligence (SI) we mean an intellect that is much smarter than the best human brains in practically every field, including scientific creativity, general wisdom and social skills' [1].

Taking an SI to be an agent and ignoring the word 'practically' for the purposes of our discussion, we take it that such an SI would be intellectually ahead of all human geniuses, such as Albert Einstein, as there are probably fields with which each genius wasn't all that well-acquainted. So we are setting the bar high. However this definition will do to go on with. The SI could be implemented on a variety of platforms. For example, it could be implemented on a network of electronic computers, and it could have some organic components. We will assume here that the platform is a digital computer.

An important engineering consideration arises immediately: what will any SI be able to do that an intelligent human cannot? Will it be able to tell us answers to hard, open mathematical questions such as whether any odd perfect numbers exist? Will it be able to sort out famous paradoxes? It is hard to say what it will do. But if we do not know what we are looking for, deciding whether it is possible is going to be difficult! So that is one of the main issues we will need to look at with regard to prospects for the achievement of SI.

2. Where do SIs fit in with visions of the future?

Are SIs going to be needed for our survival as a species? For many people considering where mankind is heading, 'survival in the long run' or some rather vague concept of 'making progress' - of the species and/or individuals - are seen as drivers, and this seems to swamp their cosmos views. Sometimes they preach a rather limited gospel of 'survival via scientific inquiry'. They might therefore conclude that anything on top of a survival-support framework, and maybe even the idea of having any sort of SIs in the picture, would be a mere luxury. This raises the wider question: is survival alone a fit objective for a human life?

To fix our thinking, let us look for illustration at some very heavily cut down examples of visions of the future of mankind. SIs could figure prominently in these, but we will see that it is sometimes claimed that they do not have to, and that itself is an interesting point for our present discussion. On the other hand, some futurologists consider the level of human intelligence to be rather low, and that we are moving forward to increased consciousness, complexity and personality, and even to some vague combination of the best capabilities man can offer with some key attributes of deity in machines. This would imply an advent of super-intelligence - at the Singularity, the point at which computational tools first outstrip human capacities.

A professor at Tulane University, the mathematical physicist and cosmologist Frank Tipler is well known for predicting that machines will build ever more powerful machines. He predicts that the minds of inhabitants of the Universe will migrate by some means to those machines, and they will be implemented in the end as computer programs that run on them. So there would be artefacts that displayed human-level intelligence. However, I have not seen anywhere that he claims explicitly that the computational tools themselves will be super-intelligent in our sense – at least not until very late in the whole sequence of events he sets out. In passing, at this point, we note that he tries to tie this all in with traditional concepts in Christianity, such as incarnation, resurrection of the dead and eternal life.

On a different but related tack, MIT professor Seth Lloyd says that the physical universe is in fact a computer, but not a conventional one, and he does not say that it is a super-intelligent one. He says that the observations we make are all consistent with seeing the universe as the ultimate big quantum computer.

The Harvard professor of psychology Steven Pinker is among those who have objected to the idea that the Universe is an information processor. He has commented that everything contains information – but that to process information, the information processed would have to stand for something, and the processing system would have to attain some goal. For example, a computer program might have the goal of sorting numbers which stand for populations of cities. This kind of thinking raises interesting questions for some 'futurologists'. For example, should we seek to progress, even if our goal is only to survive, by programming this existing big system in some way? If it has already been programmed, what does the universe compute ultimately?

Using roughly the same science as Lloyd, I presume, including models of particles and the universe as a whole, Tipler seems to focus on the possibility of us building the amazingly powerful computing systems he predicts. According to Tipler's calculations, the universe itself may provide us with the resources needed for this system building - unlimited memory capacity, with memory access times cut down arbitrarily, or unlimitedly, small. He writes about adventures such as harnessing the power of objects large and small in the universe, such as stars and molecules, to get energy, and developing computer entities/people that become collectively the 'Omega Point', that (or who) is already, if I understand Tipler correctly, drawing history forward!

Another physicist, Oxford University's David Deutsch, has given some support to Tipler and he spells out what he sees as the physical requirements for the knowledge creation that's essential for this sort of scenario - *Energy*, *Matter* and *Evidence*. Energy is needed for manufacturing processes and for conducting experiments, Matter to build storage for the expected steady flow of data and knowledge that is to be forth-coming, and that Evidence itself, originating from experiments in labs and the observation of astronomical entities, is clearly needed. Now, the total energy output of the Sun is estimated to be billions of times the amount of energy that reaches the Earth, and this is several thousands of times the amount of all human energy usage. So there is plenty of spare energy in this universe of about 10^{22} stars.

And according to Tipler, the universe's transformation/evolution would in fact be such that that that an infinite amount of processing would be feasible in a finite time before The End. The cosmological picture he had to start with asserts that there is an end-point of gravitational collapse - the Big Crunch. From up-to-date data, the universe appears to be expanding, and there is much uncertainty about its ultimate future. If, for example, the universe were to continue to expand for ever, this offers the potential of an unlimited source of material for the job, but it would cause various new physical problems- eg for getting material from far away - but Tipler has plenty of ideas!

In contrast to Tipler's theology-linked stance, some atheists, not being interested in any 'higher' meanings, say 'great, this can help us do without God in our cosmos-views'. Tipler argues for merging the physical picture with traditional Christian beliefs, but he says a lot of very controversial things, such as that precise simulations of an entity are identical to the entity. His vision is nonetheless often seen as being 'hopeful' for us or for our 'descendants'. Even for those who do not share Tipler's take on resurrection, etc, ensuring survival of the species might be the incentive needed to encourage humans and our descendants to devote attention to working towards his vision of the Omega Point.

But would a computer emulation of me really be me? And what reason could there be for future agents to produce emulations of people who lived long ago – complete with the horrors many of them caused? Moreover, many people would say that if they're to survive, they want to be as themselves, complete with their own original, if possibly somewhat altered, body. The Bible for example, teaches consistently that a body is on offer. My guess is that most people would want to exist physically, and not just to be, say, bits or qubits or circuits in a computer emulation. I for one would say: 'a representation of me is not what I want'. I prefer to anticipate and hope for my persistence and fulfilment, body and all.

The conceptions just described look at what is 'possible' within the laws of science – ie. what is not excluded by those laws. Some of those who write about this say that super-intelligence is not essential, although many do not make it completely clear what their views are on this. Its arrival would presumably speed the whole process up, but Deutsch for one argues unambiguously against the very possibility of super-intelligence, as we will see below.

There is another interesting slant on the view of the universe as a great computing system that is of relevance here, but it is largely beyond our present scope. There is a widespread belief around that mind plays an important basic role in the universe. Nobel laureates like George Wald in Biology, Werner

Heisenberg and Max Planck in Physics, and Christian Anfindsen in Chemistry, along with other prominent scientists, have made some, perhaps surprising, statements on this. The following is a selection of quotations from scientists:

'It is Mind that has composed a physical universe that breeds life, and so eventually evolves creatures that know and create.'

'…there exists an incomprehensible power or force with limitless foresight and knowledge that started the whole universe going.'

'…. mind and intelligence are woven into the fabric of our universe in a way that altogether surpasses our comprehension.'

It is well-accepted by scientists that the universe is orderly and makes sense, and many would say that we humans with our special mental capabilities must be involved in any purpose it has. Christians believe that we improve our understanding of the universe through observation and reflection, but primarily through humble involvement with the self-revealed personal Creator, whose mind is beyond any level of intelligence we can imagine. After all '….a man's reach should exceed his grasp, or what's a heaven for?' (Browning).

3. The state of the art

Watson is an IBM computing system. A few years ago it played and won against two of the best human players of the high-level TV quiz game *Jeopardy*! So it displayed remarkable capability in a task that is taken to require high intelligence when tackled by a human. Incidentally, that phrase is often taken as informally defining what is meant by 'AI'. Watson can address very obscure problems, expressed somewhat subtly and cryptically. It is able to access and process huge amounts of information expressed in natural language very speedily.

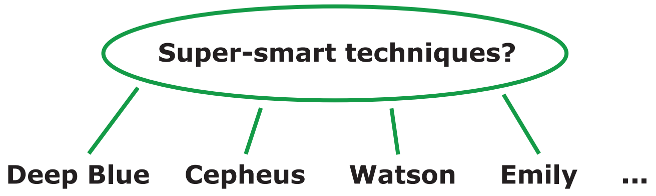
More mundane examples of current uses of AI than Watson include recommender systems on Amazon, Airbus, guided missiles, and driverless cars. All of these systems incorporate AI techniques in applications of computers to improve effectiveness and efficiency. But all these systems are playback - they accomplish limited, well-specified tasks effectively and efficiently. Other impressive AI systems include Deep Blue for chess, Cepheus for poker playing, Cyrano for mathematical pursuits and Emily for musical composition. The acronym 'HLI', is used below for human-level machine intelligence to distance it from the narrower level of AI they exemplify.

For a designer aspiring to include innovation and 'shaking things up' in computing machines, there are some things that can be contenders for supplements to, say, relatively familiar mathematical search methods, such as those in genetic adaptation algorithms. Examples of such 'tricks' include Bisociation, linking previously un-related entities or concepts, and Construction, adding in some missing link that completes a chain from the givens to the result desired. And we can try to look at things in new ways, expand or contract, generalise or particularise - perhaps to extremes. Or we can suspend assumptions that are latent in the givens, at least temporarily.

Top-quality HLI and certainly an SI, would have Newton-Leonardo-Bach-Einstein levels of capability. Such human geniuses stand out because of their demonstrations of originality, analysis and/or synthesis - sometimes in abstract cognitive space. One of the questions they seem to be outstanding at answering well is: What problem will I address next? Persistence and a 'frontiersman' or 'intrepid explorer' spirit would also be useful for this of course.

In 'Speakable and Unspeakable in Quantum Mechanics', Belfast-born John Bell, of Bell's Theorem fame, referred to Arthur Koestler's book 'The Sleepwalkers' about Copernicus, Kepler, and Galilei, heroes of the Copernican revolution. Koestler acknowledged their achievements, but he 'saw them as motivated by irrational prejudice, obstinately adhered to, making mistakes which they did not discover, which somehow cancelled at the important points, and unable to recognize what was important in their results, among the mass of details'.

Computers can also be very persistent! However, what seems to make most difference in break-throughs is when people appear to have and to use much advanced versions of the 'shaking things up' examples listed above. We'll call these 'genius tricks' for reference here, and they are very hard to identify, but if any were to be picked out clearly enough, it is conceivable that they could be incorporated in the 'super-smart techniques' super-structure of the reference schematic shown in our diagram below.



So, where might we get these techniques? That is a big question. While ideas ranging from physiological inputs to the operation of divine revelation have been expressed, there seem to be only hints in the literature of the sources of inspiration and usable details of the manner in which the smartest people arrived at novel ideas. In the world of mathematics, for example, the famous 19th century mathematician Henri Poincaré left us some details of his experiences as he made mathematical discoveries, and he revealed that he found it to be a somewhat mysterious process.

More recently another mathematician, Cedric Villani, who, incidentally, was appointed as director of Institut Henri Poincaré in Paris in 2009, also described the sort of experience he had when making an advance that led to him being awarded one of the highest honours there is for mathematicians. In his book 'Birth of a Theorem: A Mathematical Adventure' in 2015 he writes:

'*Princeton, Morning of April 9, 2009.* Uhhhhh ... man it is hard to wake up! Finally with the greatest of difficulty I managed to sit up in bed. Huh? I hear a voice in my head. *You've got to bring over the second term from the other side, take the Fourier Transform and invert in* L^2.'

He also found the key 'aha' aspects of the process mysterious, and commented on the lack of accommodation of this in publications in the mathematics research literature. If we could identify and acquire creative 'super-smart techniques' we could think of hiding them away in the super-structure that makes use of component AI systems. This is a 'big If' where 'voices in heads' are concerned. Then, could the appearance of a 'first SI seed' from, say, a genius trick, something like a 'Poincaré /Villani method', and/or some of the 'shaking up' manipulations we looked at, and/or some random fluctuations or mistakes, and/or even some trial-and-error moves, give a small advance beyond human level capability? And if we do not want that to happen, can we just make sure we reserve that 'super-structure role' for humans? These are good questions!

4. Possible road blocks?

The journey to better and better AI and in the rough direction of HLI promises to bring enormous payoffs such as improvements in medicine, education, and retailing and applications in commerce, defence, transportation and engineering. Most of the people working on the engineering side of the disciplines involved in the journey are motivated by the pull of applications. Up to HLI, the driving-force coming from promised improvements might be enough to overcome resistance. If greater-than-human-equivalence is contemplated seriously, though, there might be opposition to further development. Now, much pertinent history to date, especially concerning previous promises of AI advances and the 'AI winters' that were experienced in practice, and over-hyped outputs, does not instil confidence that SIs will ever be developed.

Ethical matters also have to be considered. There are, of course, ethical issues already present in, say hydraulic fracking, and the design of cars or power plants, but some agents with HLI could be expected to operate in unusually unpredictable contexts and to fill responsible, perhaps caring, roles in society. These agents, and any with more advanced mental capabilities than this, might have to have appropriate legal status. And a single profile can be copied onto many machines - so, do they have more than one vote? Do we control our new 'equals' by tightly constraining or even manacling any agents that have HLI, or the SIs?

Producing an SI by first imitating the functioning of the human brain would require appropriately powerful hardware and software, and lots of sensory inputs, possibly requiring unfettered mobility, which could be dangerous! Are there any serious Computing Technology barriers to progress?

On the hardware side, the fastest supercomputer today is Tianhe, working at 33.86 petaflops (ie. 33.86 x 10^{15} floating-point operations per second). When people have studied what we'd need to emulate all human history, various estimates have been obtained, but let us just chose one that has been used by Oxford University philosopher Nick Bostrum for illustration, namely around 10^{36} operations in total. So Tianhe won't do. However, a rough approximation of the computational power of a planetary-mass computer is 10^{42} FLOPS, so, on these figures, a single computer of that power could accommodate the simulating of the mental history of humankind many times over, at least conceivably. Tipler's models and calculations come into the picture again.

But we can't yet even control the weather on Earth or earthquakes, so any action plans for eg. Tipler's pattern of progress could still be premature! And what are the prospects for the required software being available? Current software development focuses on 'play-back' systems. For example, systems like Watson have access to very large factual knowledge-bases along with search, reasoning and learning capability. Enough might be discovered about our own learning algorithms and other cognitive capabilities to make it possible to copy some of these in computers.

But as we noted above, for SIs' software the goal is unclear. Hardware innovation is impressive and it would be necessary - but it will not be sufficient if we do not know where we are going. Current systems may perform particular well-specified tasks better than the best humans, but what type of cognitive task would an SI be able to do? Some people argue that the designers would have to be super-intelligent themselves to be capable of detailing a first working SI. Even HLI is still well beyond our grasp.

Another question is: How will we ever be able to arrange for the 'first SI seed' to be planted? Could we hope that we could generate HLI and SIs as a side-effect of a process where we start by linking some programs like Watson or Deep Blue, and trying to simply speed them up? Facts that are already 'available' but still hidden within computer stores, perhaps coming from e-Science projects and other Big Data sources, could be uncovered. A future Watson, for example, could give information that had never been sought or used before.

A very mundane illustration of this is where a timely 'transitive closure' result could be obtained that showed that there is a possible, previously hidden, chain of commercial aircraft routes between Belfast and some remote and obscure airport across the world. That sort of discovery is through search and the link-up of previously-known facts and relationships, in this case flight connection information. But SI-level discoveries would also require genius tricks, so we are back at our question: Where do they come from?

As we noted above, Deutsch for one says that we can't ever get an SI. He extolls the unique capabilities of humans, although he does not attribute their uniqueness as due to their being in the image of God, as basic Christian theology asserts. He says: 'There can be no such thing as a superhuman mind … No concepts or arguments that humans are inherently incapable of understanding'. When he talks like this about humans he seems mean mankind collectively. I have not seen anywhere where he explains the evolutionary leap from the smartest non-human animal to the human, or the chasm between the smartest genius and the pedestrian man-in-the-street. My guess is that he does not rule out the possibility of the equivalent of a genius trick, or similar, having produced a small advance to get to human level capability, analogous to any leap there would be from HLI upwards. Yet somehow he rules out the possibility of something similar happening again, at the higher levels.

A standard belief of Christians is that the first of those two leaps came when God '… breathed the breath of life; and man became a living soul'. Theological influences on all of this speculation are interesting, but they need careful handling by the layman. In particular, Tipler was very influenced by two theologians, Teilhard de Chardin and Wolfhart Pannenberg whose speculations raise deep questions. For example, could we 'universal predictors, knowers, explainers, deciders …' rely for our persistence on mankind's own provision, and could a move towards SIs lead to worship and service of 'the creature more than the Creator'?

The imago dei - 'the image of God' – identifies us as a special part of his 'very good' creation. We are a 'little lower than the angels', though, so are the minds of angels above or below SI level? Christian researchers in technology would need to think carefully before becoming involved deeply in any programme like those outlined above. In the case of AI, my own view is that we should restrict our efforts to producing those beneficial practical outputs that we listed earlier.

We can look to scripture passages such as 1 Cor. 15, 1 Thess. 4 and John 20, for example, to get contrasts to some of the more bizarre visions of future cosmic events above. However we are left with many uncertainties about the details of all of this. Another theologian, Henri Blocher, wrote helpfully, on progress in a different aspect of scientific knowledge and its link-up with Christian beliefs, 'We should not be embarrassed to conclude with uncertainty: it is a mark of mature faith, properly based on adequate evidence and … progress by faith, not sight.'

Proverbs 1:7: The fear of the LORD is the beginning of knowledge.

Proverbs 9:10: The fear of the LORD is the beginning of wisdom.

References / further reading

[1] Bostrum N (2014) Superintelligence: Paths, Dangers, Strategies, (OUP).
[2] Tipler FJ (1995) The Physics of Immortality: Modern Cosmology, God and the Resurrection of the Dead, London:Macmillan.
[3] Deutsch D (2011) The Beginning of Infinity, Penguin Books.
[4] Lloyd S (2013) The Universe as Quantum Computer , in A Computable Universe: Understanding and exploring Nature as computation, H. Zenil ed., World Scientific, Singapore, 2012.

About the Author

David Bell graduated in 1969 in Pure Mathematics, and has three research degrees in Computing topics. He became a full professor 1986, and has been at Queen's University Belfast since 2002, where he was the Director of Research in Knowledge and Data Engineering, a group focussing on research in Soft Computing until 2011. He is now Professor Emeritus.
He has produced several hundred publications, and supervised more than 35 PhDs to completion. He was prime investigator on many national projects and on about 18 EU-funded projects (eg MAP, ESPRIT, DELTA, COST, AIM) in IT in the eighties and nineties. His research interests are centred on data and knowledge management, involving the linking of reasoning under uncertainty, machine learning, and other artificial intelligence techniques with database research. This involves the exploration of other topics in computing, including, aspects of agent awareness and innateness.

The God and Science Papers

The God and Science Papers are published by Christians in Science Ireland, and are based on lectures held in the God and Science Lecture Series. Christians in Science Ireland is the Irish local group of Christians in Science, an international network of those concerned with the relationship between science and Christian faith, and incorporated in England and Wales. The God and Science Papers address a broad range of topics related to the interactions between science and religion. The opinions expressed in the God and Science Papers are those of authors and do not necessarily represent the views of Christians in Science Ireland. The full list of God and Science Papers can be found at http://www.cis.org.uk/ireland/godandscience/, from where free copies can be down-loaded in pdf format.

This paper is based on a lecture given by Professor David Bell on 24 March 2015 in Belfast.
Publication date: September 2015. © David Bell.