**Machine Consciousness**

Slavoj Zizek was asked recently in the Guardian - What is the future of philosophy? It will become more important than ever, even for the so-called “ordinary people”. Think about biogenic intervention, which may even change your character, how your psyche works. The 21st century will be the century of philosophy.

What is real.. what is reality? We have explanations of what works that we call science and explanations of why it works that we called philosophy. I want explore both of these strands. The past is a mixture of science and philosophy, the present is about science and the future philosophy.

Evolution is a good working theory of a process- natural selection / survival of the fittest. We started from an amoeba and ended where we are by way of monkeys. Up to now it has been based on biology but It seems quite reasonable to me that machines are just another step on this journey. In 1863, Samuel Butler wrote *Darwin Among the Machines* he pointed out that machines are evolving far faster than humans – very prescient.

‘Why’ is a big question that has driven humanity onwards throughout time. For some that is the purpose of religion. For others who are rationalists we need an explanation that we can believe in. The third group who not interested and find that question irrelevant.

In *Death and the Afterlife* by Samuel Scheffler says that earthly afterlife is a structural feature of being a death bound creature. He posit that if the end of humanity was by meteor strike which we knew was coming in the very near future it would happen fast or by global infertility which would happen more slowly. We might react very differently in these two scenarios though both implied the end of humanity in the near future. In this book there are lots of questions about what it means to be human, is it our bodies or our minds?

Computers are just at the birth of where they will be in our life time. They used to be a single fixed task machine. This is such an old fashioned view so we should see it as such. We are giving computers capabilities now and these are very primitive compared to what will see in our lifetimes. For example IBM has just introduced a chip called SyNAPSE. This is hopeless at traditional step­ by step computing but it is very good at pattern matching. It has a network of 1m ‘neurons’ connected by 256m ‘synapses’, roughly on the scale of a bee’s brain.

Do not confuse machine intelligence with consciousness. The singularity is forecast to happen at the end of our lives when the computer will become more ‘powerful’ than the brain. Wikipedia defines this “ technological singularity hypothesis is that accelerating progress in technologies will cause a runaway effect wherein artificial intelligence will exceed human intellectual capacity and control, thus radically changing or even ending civilization in an event called the singularity. Because the capabilities of such an intelligence may be impossible to comprehend, the singularity is an occurrence beyond which events are unpredictable or even unfathomable”. The father of the singularity is Ray Kurzweil (now heading up robotics for Google) says “there will be no distinction, post-Singularity, between human and machine”. Vernon Vigne uses the analogy between the breakdown in our ability to predict what would happen after the development of superintelligence and the breakdown of the predictive ability of modern physics at the space-time singularity beyond the event horizon of a black hole. If a superhuman intelligence were to be invented it would bring to bear greater problem-solving and inventive skills than current humans are capable of. It could then design an even more capable machine, or re-write its own software to become even more intelligent. This more capable machine could then go on to design a machine of yet greater capability. These iterations of recursive self-improvement

So we can expect machines to understand what we want more and more. The Turing test (Imitation Game) is about computers as they used to be. Pretending to be human is getting simpler and simpler for computers. A much stronger test is can they invent new things.

During our children’s lifetime computers will become companions and during our grandchildren’s it will dawn on some people that they may take over. It is a bit like climate change – it is clear that is happening but nobody will take serious responsibility until it is too late.

To illustrate this point . Weapons can now decide for themselves what to target using AI. |Humans launch them and ‘tell’ them what the objective is then leave it to them to decide how to behave. An example of us losing control. When Google bought Deepmind(AI company) founder Demis Hassabiswas was wise enough to force them to set up an ethics body. Maybe we need a technology philosopher in chief.

Experience is just the inputs we receive together with the memory system that is inside us so that is all very mechanistic. Thoughts (like emotion, altruism..) where there are no inputs must rely on memory solely. Memory is not fixed so trying to look for it may be self defeating as it may be a process.

The computer model of memory is storage and indexing. A recent paper in Topics in Cognitive Science suggests age related decline in cognitive function is not true. As you get older the database increases in size so you take longer to find things. It is possible that the indexing system needs to be exercised and as we get older we do less.

Google processes 24 petaB (peta/tera/giga) per day.!

Computational neuroscience is the trying to understand how humans work and refining models that aid part of that process.

**What’s happened**

Since I last talked about what is happening in the world of computers (a couple of years ago) I have been filing articles. As part of reorganising my filing system I got them out and went through them. A surprising number of them were cutting edge at the time but now seem curiously dated. A good example is IBM’s machine Watson (who won at Jeopardy). Its knowledge was based on the fact that it had (& remembered) the whole of Wikipedia.

An article in the Guardian of 15.03.11 was entitled “Reality Check” . It talked about web 3.0 where the boundary between ‘real life’ and ‘online’ disappears. The internet of things is already happening and with the advent of 3-D printing who knows.

I expect the internet to be connected directly to our brains at some stage in the future (mid term). This will not happen suddenly so we won’t get the “blinded by the headlights” experience. We will have local memory (ie our mind) and global memory which we will access when we want.

‘Potty for ipads’ was a Guardian article where the title had a double meaning. Tablets are becoming a part of our kids’ lives. In the near future we can expect every primary school to be using them in the same way that paper and pencils are seen as essential tools. Does that affect how the human mind works? How many parents will fight against that and deskill their children from the world that they will grow up in.

One film getting attention is “Her”, the story of a man who falls in love with his computer operating system. This may be just semantics but the separation between hardware and software seems to have gone. Operating systems (OS) used to be the basic level software that meant that the hardware could do the basic things that all programmers needed. To do anything specific required an application (now called app). In this SF film the computer has merged hardware, operating system and app. It has solved the problem of natural language processing and the singularity may have happened. It is set in a world where it is perfectly natural to have a relationship with a computer.

Leaving all the technical issues aside there are three main things that the film brings out. First physicality – the OS is just a mental construct so interacting physically with humans is clearly brought to light. Next is the OS behaving like humans irrationally (falling in love). Is it behaving this way because we expect human to be like that and the OS realises that it needs to do this to make its behaviour appear more human.

Finally computers aren’t individuals, they can have a relationship with lots of people at the same time. Dr Genevieve Bell, cultural anthropologist at Intel Labs , says that devices that will have relationships, rather than mere interactions, with people. So at end of the film one is left with the feeling that computers won’t need people.

So where does this lead us?

The Cambridge University Centre for Existential Risk is concerned about the end of humanity. At recent conference they assembled experts in lots of disciplines. A question they asked them was ‘would we still be here at the end of the century’, 19% answered no. So this is a very real and immediate issue. One of the threats is AI and this suggests that machines will eventually not need us and they might set about killing us. Not directly but by not providing the services that we need to keep ourselves alive. In the future we are going to be wholly dependent on them for things like health and power.

I’m sure you want to know a bit about quantum computing. It has the potential to be many million times more powerful than existing computers. What is it and how does it work? Digital computers (DC) work using the binary system, on or off, 1 or 0, quantum computers (QC) can be on and off at the same time. To evaluate a complex model we try millions of different combination one after another, with QC all that can happen at the same time. This biggest prize is examining a very large number to find two prime numbers than multiply to make it. Using DC it takes about as long as the time since the universe was created and is the basis for internet security, with QC ..... The problem is keeping the little blighters where you want them and getting them to do want you want. There is a least one QC but it is fearsomely expensive. They will probably happen one day but probably not till the middle term.

Facebook has just bought Oculus for $2bn. This is a virtual reality (VR) system which presents the world through a set of goggles. It potentially has lots of application as it feels like you are there, wherever that is. So sporting events, music and theatre allow a lot more “cooks to get round the stove”. But this just relies on sight, do we trust what we see on a screen? Images can be “airbrushed” in real time. In the VR world you can be just who you want to be so the separation of mind and body is starting to happen. To set against this is the power of live. Humans appear to love live events ie being there not experiencing it as it happens. Music groups make much more from live performances than they do from recordings, particularly in this increasingly digital world which makes almost compulsory to share things for free.

They say that the brain is the most complex object known to man so what will happen when we have a more complex object.

**Rich and Poor**

All these are example of what is possible today . Who wouldn’t want to correct biological deficiencies like a failing heart, poor hearing or weak eyesight. Who wouldn’t want more memory, a more powerful brain or to extend life. Very few, so it may be up to those who can afford it. Upgrading yourself is not an egalitarian project it is elitist. The latest human enhancements will only be accessible to the rich, leading to a society that is more and more unequal.

In *Sapiens* by Yuval Noah Harari says “In the 21st century, there is a real possibility of creating biological castes, with real biological differences between rich and poor... In the not too distant future we could have more than one human species on earth again.”

In *Inferno*  by Dan Brown calls this master race transhuman. At the centre of this book is a geneticist who thinks that the population of the earth is running out of control (exponential growth) and creates a virus to control this. The ending is unsatisfactory as it would have far better if the only people who could survive this virus were transhuman. This would have been a classic moral dilemma but that doesn’t sell blockbusters.

Anders Sandberg , at the Future of Humanity Institute at Oxford Martin School disagrees. He agrees that speciation might happen but thinks it will be driven by culture. You might get a country that wants to make people healthier by subsidising the healthcare budget. So Singaporeans might become their own species.

**Consciousness**

All this suggests that technology will become an increasing part of our lives. Could machines become conscious?

What is consciousness? It could be wrapped up in the word “I”. This implies a single entity that is aware of itself. It also implies the word “other” (not-me), separate sentient beings that are like me. I use the word sentient being because this talk is all about machines (constructed originally by humans).

“I” may also signify hierarchy. This certainly is so with humans and animals. We need a social grouping for this to be true. Another thought is “I” is an historical construct – I am my history – and it is this construct that interprets the present.

Is there more to the brain than just the physical component? Descartes was aware that one couldn’t rule out the possibility that matter may be conscious..( Article on Real Naturalism by Galen Strawson in LRB 26/9/13).

Is it just a matter of complexity? Is this ‘I’ness just happened through that.

Once the singularity has happened then unless humans have a non-physical element sometimes called a “soul” then machines will certainly takeover.

The “soul” is very evocative word. All religions believe that there is something that is non physical. This is one of the answers to the question ‘why’. If we discount religious beliefs then do we believe that there is something special about humanity that cannot be reproduced by a machine.

I believe that consciousness is a just one aspect of complexity. When computers get more complex than brains (particularly in a smaller space) then why can’t they become conscious?

A couple of quotes from Joseph Priestley (1770)

“the faculty of thinking is the result of a certain arrangement of the parts of matter”

“sensation and thought do necessarily result from the organisation of the brain.”

This clearly agreed with what Hobbes(1641) & Tolland(1704) thought.

Roger Jennings in his talk “Mind & Brain” states - Factually, all the evidence that ‘stuff’ in the form of silicon-based digitally-operating machines is incapable of generating and experiencing thoughts and feeling or possessing a notion of ‘self’.

This seems to me to be a very backward looking view. In our lifetime technology is going to become so sophisticated that it will become part of our lives. The current forecast for the singularity is 30 years from now. So we need start thinking about what world that implies.

One big mistake is to believe that computer will never think/behave like us. Why should they?

A good example comes from David Mitchell in the Observer (23/11/14) in his article about not pricking the Christmas bubble as it may keep capitalism alive. “At the core of our midwinter festivities is something deeply irrational, an urge that robot would never understand: a need to make merry, to paint the town glittery, to lavish gifts...”. This seems to me totally rational from a robot’s point of view as it shows how humans behave. The last statement implies that robots will understand humans and possibly human thought processes will be seen by them as primitive.

As complexity grows we will have less and less knowledge of what is going on. Machines may be a new chain of evolution, i can imagine a world where machines look at humans as we look at monkeys. They will probably keep a few remaining humans in a zoo so they can be seen by their children (anthropomorphism).

This raises the “I” question. Will machines become individuals? We think in terms of aging so having children can be a part of this process. Machines don’t have to ‘wear out’ so don’t need to have children for this reason. Creating a child machine may make the parent redundant so that would be a moral dilemma. This presupposes that machines behave like individuals.

We are creating a world that relies more and more on technology. A cross-over point could come when we ask machines what to do. We live in a world that is becoming more and more divided. How do we decide between personal and species ethics? How individual’s self-interest can be reconciled with the larger impact of human survival? We might need machines to help us with this dilemma and then they may decide that the main risk is humans.