Maria Curtis

Summer School 2014 – Theme Talk – The Authentic Self -Tuesday 19 Aug

Hymn: 227 (Purple) Gathered here (4-part round)

Chalice lighting

Today may our chalice flame represent our humanity, our human-ness, our human nature, our authentic selves.

Let us celebrate what makes us human.
Let us gratefully appreciate the privileges of being human
And may we accept, too, the responsibilities:
To our own species, to other living things, and to our beautiful, long suffering planet earth.
May we use our amazing faculties and sensitivities
To bring compassion, peace and harmony to the world. Amen.

Introduction

I want to think today about what it is that makes us human. And I've chosen to look at this topic through the lens of *artificial* intelligence, AI, attempts to create machines that are like human beings. What can we learn about ourselves from their successes and failures? For some, it is only a matter of time before we create a machine that behaves, thinks, speaks and even feels like a human being.

Our young people have been working on a drama, loosely based on Stephen Spielberg's film, AI, which looks into some of these issues.

Children's drama

5 minute drama.

The film assumes the robot child has feelings. We can't help feeling empathy towards the abandoned child. But what I want to think about today is whether in theory a machine *could* have feelings programmed into it. My answer is a resounding No.

I recently watched the film Under the Skin in which an alien species (represented by Scarlett Johanssen) comes to earth to get some spare parts(?) She entraps men who are then "melted down" so that all their tissue and bones can be recycled(?) – I'm guessing here. She is distinctly odd at the beginning of the film - (it's set in Glasgow – impenetrable accent adds to our sense of alienation!) - speaking to her victims in formulaic sentences. There is a

shocking scene when she shows no emotional response to a toddler whose parents have just drowned. There are some subtle changes in her behaviour once she begins to respond to innocence(?) and kindness. Next thing, she's kissing and having sex with a stranger and you think, Whoa! How could that transformation have occurred? What kind of model of the emotions is that? It's a very naïve idea that we can learn all about the emotions from a couple of isolated experiences. She was not a robot, but an alien clothed in a human body. But I would want to contend that no species, alien or human, could engage in space travel, without having developed a co-operative culture similar or analogous to ours, based on shared understandings, a shared emotional life, empathy. (Maybe the film was meant to be more of a reflection of the loss of humanity in our lives – perhaps we can talk about it later. Blu-ray DVD of film in Silent Auction.)

These examples – our play, the film - are from science fiction, but what worries me is that some of our top scientists and philosophers (Stephen Hawking, Michio Kaku, Daniel Dennett) have equally naïve views. They see the human brain as a complex machine which could, in theory, be replicated. Some think we will eventually be able to build computers that will be superior to us (in intelligence) – they will take over the world and we will be lucky if they're prepared to keep us on as pets! Some imagine computing as a kind of oncoming rapture – (Ray Kurzweil *The Singularity is Near* 2005) – envisaging a "Utopian" (dystopian?) future where we shed our bodies and upload our minds into computers and live forever, virtual, disembodied, immortal.

To me, the fact that they think like this sends out alarm bells. It shows that they have no idea of what an amazing, wonderful, but complex creature the human being is; they have no appreciation of the subtlety of human interaction; no regard for human creativity; no understanding of the nature of our identity, of what makes us our authentic selves.

So I want to think about what it is that makes us human. We're often compared to animals – of course always coming out as superior – "animal" is a term of abuse used for murderers, rapists, torturers, etc, which is very unfair to animals. We are told in Genesis that God gave us dominion over other forms of life, including animals. (Animals, after all, aren't motivated to kill by a distorted ideology.) Well, we've made a right mess of being at the pinnacle of creation – riding roughshod over the natural environment, as if by right. We need a more

humble view of ourselves. *Yet* I think the Genesis creation myth illustrates a profound truth about the human condition – once we've eaten of the tree of knowledge, then we have choice/free will. Once we can think and talk about things, we have lost our animal innocence and entered the moral domain. We can make good or bad choices. We can choose to do right or wrong. And sometimes we face very real dilemmas where we don't know what is right, or situations are so complex, there aren't obvious goodies and baddies to side with. (Or other people, politicians/bankers/multi-national companies, do reckless and exploitative things in our name, in the name of our country or our culture and we experience collective guilt.)

So we are moral agents, able to make choices, but I would contend that this isn't a purely *rational* or *conscious* process; *unconscious* processing and *feelings* like empathy and compassion come into our moral decision making. It's important to consider how these come about – eg, what are the conditions that enable us to feel empathy and what prevents it from developing? How do we become brutalised? (That's beyond my scope for today, but we could return to this issue in our discussion this afternoon.) For the moment, I want to focus on **language,** as many of the examples of Artificial Intelligence we'll be looking at later in this talk are ones that simulate human *communication*. Language – in the sense of a symbolic system - is a good contender for what distinguishes us from other creatures. (Language/consciousness/free will linked in complex ways.)

I'm going to put my cards on the table and say that I don't think it will ever be possible to programme thoughts or feelings – consciousness – into a machine. These things emerge from **experience of being a body in the physical and social world**, with a **history of millions of years of evolution in our genes.** I'm going to argue that what being human is about is creativity, unpredictability, quirkiness, humour, fun, joy as well as our ability to empathise, to feel compassion.

This is not to say that a good *simulation* of human behaviour cannot be convincing - (In some cases, a "friendly" robot may be preferable to a nasty human being. Eg, Ishiguro's humanoid) – but this says more about us as human beings than it does about the robot. We need to bring in the notion of **projection** here. Human beings are meaning makers. Show someone a film of a couple of dots moving randomly on a screen and ask them what they

see, and you'll get a narrative: eg, that one's chasing that one; they had a row; that one's storming off; now they're making up. Add a different shape, eg a V and you've got a novel! We just can't help doing this. We do it when we meet someone on the internet and it can be dangerous. Apart from the fact that many people tell lies, eg, about their age, we have much less information about them that we would have if it were a live conversation in the real world– body language, facial expression, tone of voice, hormones like pheromones, and scent – all that information coming in through our senses, much of it not consciously processed, but processed nevertheless, often by nerves in the gut. However, when we don't have this sensory information, we make it up. When we have a conversation with someone on line, there is an awful lot of **projection** going on – they become our ideal: trustworthy, kind, generous, etc. So with minimal information, we find it very easy to project…but it worries me that some of our scientists, who ought to know better, make the elementary error of assuming that because a machine may be built that can produce something that looks like language, it really is performing an act of communication. They are projecting intentionality where there is none. I think this comes from adherence to a Materialist view of human nature, a sort of denial of the self.

This summer school is about The Authentic Self. I'm certainly happy talking about the self, or the mind, or the soul. These are all the same sorts of entities – immaterial entities – and I may use the words interchangeably. I'm not too worried about that. The point is that Materialists deny the existence of immaterial entities like the self and the mind and the soul; hence the title of philosopher Mary Midgley's latest book: *Are you an illusion?* in which she challenges the Materialist point of view.

Let's have a look at some models of human nature that illustrate different ways in which body and soul are linked:

Flipchart - Views of human nature

Dualism – Plato, Augustine, Descartes:

Humans = Body and Soul; soul is our authentic nature.
Soul is "in" the body (mysteriously linked)
Soul is immortal and superior to the body. (Original sin passed on through sexual act, etc)

Materialism – Michio Kaku, Stephen Hawking, Richard Dawkins, Daniel Dennett, etc

Science can tell us everything about the world and ourselves.

Immaterial souls/minds don't exist. The self is an illusion
We are nothing more than our physical bodies. Brain = machine
Mind = brain; Thoughts/feelings = physical processes in brain and body

Realism/Commonsense – Aristotle, Aquinas, Mary Midgley, me

Science can't tell us everything about the world and ourselves.

We are rational animals
Brain is part of the body
The self/soul/mind exists as a non-material entity, integrated with the body.
(John O'Donohue: Body in the soul)

EXERCISE – Post-it notes – 5'00"

Q: When do you feel at your most human? EG, making music, cooking, lovemaking, cooking, playing with children, reading a book, alone in countryside, communing with nature, laughing – (Maya Angelou)? Share with neighbour. Write on post-it if happy to have read by others – anonymously. Stick on flip chart on way out. I will look at them before this afternoon's session. "I feel at my most human when…."

"The Most Human Human"

My inspiration for this talk was a book with an intriguing title, written by a young American computer scientist, philosopher and poet, Brian Christian; it's called **The Most Human Human,** subtitle **A Defence of Humanity in the Age of the Computer.** It focuses on a key event that happens in the AI community every year, a competition called the Turing Test, named after the British mathematician and founder of computer science, Alan Turing. In 1950 Turing posed the question *Could a machine ever think?* ie, would it be possible to construct a computer so sophisticated that it could be said to be thinking, to be intelligent, to have a mind? And *How would we know?* Turing proposed this experiment: A panel of

judges poses questions to a pair of unseen correspondents, one a human, the other a computer program, and attempts to discern which is which. Turing predicted that by the year 2000, computers would be able to fool 30% of human judges after 5 mins of conversation. In 2008 competition, the winning programme convinced 29% of judges. The competition the book focuses on took place in 2009, in Brighton! [His own culture shock: No shower, Mock Turtle tea rooms, LET AGREED sign.] Brian Christian was a human competing against a machine. He won the coveted Most Human Human award and this book is a warm, witty and erudite account of how he did it. Through evaluating a range of programs, he asks ***what it is that makes us human***, and ***how we go about being the most human we can be, not just under the constraints of the test, but IN LIFE.***

Diagram to illustrate Turing Test

There are some very sophisticated programmes about. Brian Christian tells the cautionary tale of Robert Epstein, a psychologist and editor of an AI journal who subscribed to an on-line dating service and ended up having a four-month correspondence with a Russian woman, in which they declared their undying love for one another, before he began to suspect that something was amiss. You guessed it. Ivana was a computer program!

Beyond its use as a technological benchmark, beyond even the philosophical, biological and ethical questions it poses, the **Turing Test** is about the *act of communication*. It makes us focus on how we connect meaningfully with each other within the limits of *language* and *time*; how empathy works; what is the process by which somebody comes into our life and comes to mean something to us; what is the nature of intimacy. These are some of the most central questions of being human.

It is very interesting to consider the programmes – sometimes called chatbots or just bots – which are most successful. An early programme (1960s) called ELIZA simulated person-centred Rogerian counselling. For those of you unfamiliar, this is a non-directive, non-judgemental kind of therapy in which *reflection* plays a key role, eg, client says *I'm not feeling good today*, the therapist might reply, *So you're not feeling good. Would you like to say a little more about that?* Or maybe later, *Do go on.* The technique of fitting the user's statements into a set of predefined patterns and responding with a prescribed phrasing of its own – called *template matching* – was ELIZA's only capacity, but the results were

staggering. Many of the people who talked to E were convinced that they were having a genuine human interaction. Some medics wanted to use it as a therapeutic tool in hard-pressed psychiatric services suffering from staff shortages. Joseph Weizenbaum, E's creator, was horrified and pulled the plug on the project, becoming an outspoken critic of AI research. But the genie was out of the lamp, and subsequent programs all use the basic template matching approach introduced in ELIZA.

What had shocked Weizenbaum was the idea that psychiatrists were comfortable regarding ***technique*** or ***method*** as the crucial component of the therapeutic process. I quote: *What must a psychiatrist think he is doing while treating a patient, that he can view the simplest mechanical* ***parody*** *of a single interviewing technique as having captured anything of the* ***essence of a human encounter****?* I share his concern that we may be prepared to settle for a simulation rather than the real thing *in our lives.*

As we have seen, philosophers throughout history have considered the question of what it is that differentiates us from animals; the ability to think logically is one characteristic that has been a popular candidate from the time of Aristotle. *Reason* was elevated above *feeling*. Our education system remains focussed on *left brain* activities to do with conscious reasoning with language. But Brian Christian thinks it's about time we stopped what he calls this *fetishisation of analytical thinking* and the *denigration of the creatural* that goes with it, and adopted a healthier view of human intelligence. [Educationalist Ken Robinson's TED lecture about dance being as important as maths in schools.] 20C developments in computer technology should help us rethink and re-evaluate our skills profile. Computers are now so much better than we are at cold calculation. The 19C English mathematician George Boole worked out a system for describing logic in terms of conjunctions of three basic operations: AND, OR and NOT. In 1937 a young graduate student, Claude Shannon, at MIT, realised you could implement Boolean logic electrically. The rest is history. The question is, what is left in us that **is** quintessentially human?

Brian Christian looks at the strengths and weaknesses of the range of bots that have scored well against humans in the Turing Test. On creative writing courses, the golden rule is *Show, don't tell.* Telling is dull and wooden. A list of facts about a person does not capture their essence. On some *speed dating* events, to help the interaction along and maximise the

getting-to-know-you process, people are not allowed to say where they come from or what they do for a living. The Turing test is a bit like speed dating. On one famous occasion, 1991, several judges decided that an English literature professor was a bot because she could answer the most obscure, *factual* questions about Shakespeare. The bot she was competing against, by contrast, was quirky and whimsical; it made unpredictable replies, and random comments that seemed amusing. **We can learn from this that a person's *idiosyncrasies* are what make them feel *authentic* to us.** We can differentiate communication from our friends from spam on our computers because of their verbal *style*. (NB It has so far proved impossible for a computer to satisfactorily translate a novel.)

In 2005 a program called Cleverbot used the technique of challenging the judges, claiming that *they* were the computers. The *first user* of another programme (not in the competition) stayed on-line for nearly two hours – the nature of their "conversation"? Abuse! There is something all too predictable and ritualistic about arguments; their lack of site specificity makes them easy to replicate; it's very sad that 2hrs of hurling obscenities at an unknown victim can feel like being human, but it did. These bots have a huge database of real human responses-to-questions that they draw on. If the context is favourable, they can be very convincing. But, and it's a big but, they can be stymied by asking them specific questions about themselves – one got very confused over what gender it was when propositioned by a male human. Of course, this is because they don't have a *self*; they are a loose collection of thousands of snippets of talk, a "conversational puree", with no organising principle – what we might call the *self* – just a bank of frequently co-occurring utterances. While some of us are more fragmented than others, some degree of coherence of identity is the norm; we get very upset when people are inconsistent or in denial, or when they do things "out of character". In extreme cases, we diagnose mental health difficulties. **We are products of our life history, our culture. Computers have no experience.** Words like *memory* or *learning* when used of a computer are **metaphorical.** We should beware of mistakenly inferring *agency* or *intention* from *behaviour.*

Before Deep Blue, the chess-playing programme, beat world champion Garry Kasparov, in 1997, playing chess was seen as the highest form of human activity, a creative process on a par with music or poetry, an art that "draws intrinsically on central facets of the human condition" demanding "elusive abilities that lie close to the core of human nature itself".

(Douglas Hofstadter, Pulitzer Prizewinner,1980.) After the win, there seemed to be two main responses: 1) to acknowledge that intelligent machines had arrived and we had lost our supremacy over all creation (not very popular) or 2) to reframe our idea of chess and play down its status (Hofstadter: "I used to think chess required thought. Now I realise it doesn't". Music and literature require a soul, but chess "doesn't have deep emotional qualities to it".) However, there is a third response, which I would share with Garry Kasparov, and that is that Deep Blue **did not win the contest** because they were not doing the same thing. (It doesn't have a memory of all possible games (huge), but it has a repertoire of successful opening moves and endgames, so it's only in the middle section when things get unpredictable.) Deep Blue was using algorithms (procedures/rules) derived from a huge database, whereas GK was using intuition or "feel" for the game. Just as in conversation or letter writing, we have stylised, culturally determined openers and closes, but the middle bit is more personal and idiosyncratic, more creative and novel, more risky; so with chess. Of course we can experience conversations where it all feels predictable and never goes beyond the formalities or the conventions and we feel nothing meaningful has happened. And in general, it is these stylised conversations that the bots (or bot designers) want to be having in the Turing Test; it's the weakness that they best exploit. But human beings do not *only* engage in small talk; it has its uses but too much leaves us feeling disappointed.

The chess playing computer is not playing chess. The computer that produces text from speech, or speech from text is not reading or talking or writing. They "know" – another metaphor - rules, the rules of chess, the rules of grammar, rules for converting letter strings to phonemes (sounds) and they can do matching, and probability, working from huge databases, but there is no way that there is any *understanding* going on. For people who don't know anything about how computers work or how programmes are written, **it's easy to project intention and agency onto a machine** – it's part of our nature to read significance into whatever we perceive. What really worries me is the people who *do* understand these things believing that machines will soon have independent thought and even feelings. It's a form of **denial of our humanity**.

Our authentic nature stems from the fact that we are bodies, physical beings, who have evolved over millions of years. The human brain is not a recent add-on, although the neo-

cortex is a relatively recent development. The brain is part of the body, whose main function is to keep us alive. We throw off our ape ancestry at our peril. If we were completely rational in our decision making, we would probably die, like the donkey standing equidistant from two bales of hay, unable to choose which one to eat. **The emotions play a key role in judgement** – we talk of a "gut feeling" and that is not a metaphor; we don't always know what leads us to choose one thing, one person, one direction in life, rather than another. Things going on in our **bodies** - with hormones, enzymes, neurotransmitters, the immune system, all responding to sensory information - tell us how we feel before we become consciously aware of it. Intuition, guessing, inspiration, risk-taking – these are all faculties that a computer lacks but that are crucial to our survival. Some of us – and I put my hand up to this – have been seduced or deluded into placing the intellect in a privileged position above the senses and the emotions. It is a form of denial – a reluctance to engage with the messiness of everyday life. Brian Christian acknowledges that he wasted most of his adolescence in geeky AI activities, distrusting his senses and being terrified of his body – the result: a malnourished body with bad posture; a frustrated, proud and critical individual. Lower level processes, fulfilling our animal nature, are more important to our overall well being than higher level conscious processes. Our spirituality resides in what Brian Christian calls our ***mongrelism***; he compares us to ***lichen***, which is formed of two species, fungi and algae, living symbiotically; or we're like ***the robot and monkey holding hands***, an integrated system aware enough to apprehend its own limits and push at them to produce our best emotions: curiosity, enlightenment, wonder, awe. Scepticism about claims made for artificial intelligence systems can throw into focus our authentic human nature. Let us use our integrated hybrid nature as rational animals to be the best humans – the best friends, parents, teachers, artists, lovers – we can possibly be, ever-widening the range of our capacity for empathy and compassion.

May we use our gifts wisely.

Hymn: 163 (Purple) The peace of the earth be with you

Topics for discussion in afternoon session

- End of nature-nurture debate. Interaction. (Gene expression – twin studies.)
- Think about how empathy evolves – historically (rearing young?) and within individual. We learn feelings socially. Reciprocity leads to self-regulation. (Gerhardt)
- Language and thought are social phenomena with their roots in bodily experience. (Wittgenstein's argument against the possibility of a private language.)
- Dangers of giving robots too much responsibility? Drones. Driverless cars – you want empathy on the roads, eg, letting someone cross.
- Advantages of simulations – company for lonely people; sick – remind to take medication; obese – watch diet. Teaching autistic children – non-judgemental.
- We are less rational than we seem but this probably isn't such a bad idea. Rationality alone isn't enough to help us make moral decisions. Need feelings/empathy/compassion.
- A lot of unconscious information-processing goes on which is helpful for our survival.