

<http://www.amazon.co.uk/exec/obidos/ASIN/0349108536/instituteofid-21>  The Cambridge Quintet

The Cambridge Quintet by John L. Casti Reviewed by Charles Brickdale

April 2011

This review article was solicited to form part of some background readings for a discussion on Artificial Intelligence and Human Consciousness organised by the Manchester Salon to coincide with the Manchester Science Festival.

The Cambridge Quintet by John L. Casti is not about chamber music or yet another batch of undergraduates recruited by the KGB. It concerns one of those slow-burning science stories that has been smouldering quietly away, occasionally flaring up and generating some light and a fair amount of heat, in the backgrounds of our lives for many decades.

Artificial intelligence has broken through into popular culture; think of the film *AI*, *Data*, the android with feelings, in *Star Trek* and Philip K. Dick's novel *Do Androids Dream of Electric Sheep?*, but has not, so far, been viewed as an issue requiring much attention by lawmakers or the general public.

In 1950 Alan Turing, the mathematical genius who helped launch the computer age, published a paper called *Computing Machinery and Intelligence* which began: *I propose to consider the question, "Can machines think?"*

Turing's answer, in essence, was yes. He saw no great conceptual or theoretical barriers to the emergence, one day, of computers that possessed the ability to think about problems in ways strongly analogous, at least, to the thought processes of human beings.

Turing committed suicide in 1954 but his ideas remained very much alive and, despite several false dawns, continue to have resonance today. A conference in 1956 at Dartmouth College in the USA considered two rival positions on how to create AI. One view held that human intelligence could be replicated by finding ways of representing the symbols that constitute thought and of combining and manipulating them in the patterns that, on this view, constitute mental processes. The main alternative view argued for the need to mimic the brain's neuronal structure.

Over the next twenty five years each theory was to have its place in the sun. For a time, AI researchers gave the impression that thoughtful robots who might engage in a Socratic dialogue or two would shortly be joining the universe of discourse. According to Herbert Simon in 1964: *Machines will be capable, within twenty years, of doing any work a man can do.*

By 1997 researchers had devised a programme capable of defeating the chess grand master Garry Kasparov and in February of this year another programme was on its way to winning the US quiz show *Jeopardy*. These are, one suspects, achievements that fall some way short of what the pioneers had in mind.

So what went wrong with Turing's dream and what does it suggest are the key issues in understanding the human mind and the nature of thought? In an interview with the Financial Times IBM's Arvind Krishna suggests that the early researchers did not fully understand the subtlety and complexity of human interactions with the world (IBM designed the programme that defeated Kasparov): *People hoped they could lay down an algebra but you can't model it as a set of rules.*

The mechanised chess champion provides a clue to the rather more modest, yet potentially very

positive, turn that research into machine intelligence is now taking. Rule-based processes and systems with computable possibilities and identifiable possible outcomes are well-tested terrain on which to try out the potential of machines that can engage with aspects of human intelligence. Human minds remain in charge and they do the seriously difficult thinking; indeed, the more mundane tasks that can be outsourced to machines the more we are free to push forward the frontiers of human knowledge, creativity and achievement. Again, not the flowering of new kinds of self-conscious intelligence envisaged by the proponents of **strong AI**.

John Casti's contribution to the debate in *The Cambridge Quintet* is to construct an imaginary dinner party in the Cambridge lodgings of C.P. Snow in which four of the leading thinkers of the time the year is 1949 - debate the feasibility of creating artificial intelligence in the sense described by the researcher Howard Gardner: **a pattern of output that would be considered intelligent if displayed by human beings.**

Casti imagines Snow, in his capacity as one of the Establishment's safe pairs of hands, being asked by the Atlee Government to assemble a group of experts to probe the possibility of machines being produced that are conscious and can think. It's a delightful image: never mind the Cold War, post-war reconstruction and what to do about Germany, what do we think of talking robots?

In a sense, Casti places Snow and his guests outside time. In their contributions to the discussion Turing, the geneticist J.B.S Haldane, the physicist Erwin Schrodinger and Ludwig Wittgenstein reflect ideas which emerged after their deaths as well as state opinions they were known to hold at the time. Casti's purpose is to disclose the main themes in the debates about artificial intelligence both as they were at the time and as they have subsequently developed. Many of the developments have, after all, built on thoughts first expressed by these four thinkers especially Turing and Wittgenstein.

The mathematician and the philosopher of language games represent the two poles of the argument; Turing, of course, convinced he could show that thinking machines were possible, Wittgenstein scornful of the notion. Haldane is unconvinced arguing that **there is something very special about the human brain**. Schrodinger thinks it technically possible but can't see the point. On the technical feasibility he is lukewarm, going no further than conceding that a machine could **fool us into believing that it is thinking like a human.**

Much of the debate over the Cambridge dinner table revolves around the credibility and efficacy of a test, the Imitation Game, proposed by Turing in his 1950 paper (there is a neat twist on this idea in Dick's dystopia, the Voigt-Kampff Empathy Test **Altered Scale**, used to flush out runaway androids) which in itself leads to intense probing of what we mean when we use words like **thinking**, **consciousness** and **awareness**. Intriguingly, Turing implies that his interlocuters have misunderstood both what he is trying to achieve and what it means to think: **Put simply, my interest is in duplicating human thought processes, not human physiology.**

This, precisely, is the problem for Haldane. Without the deep background of a fully human life (food, sex, relationships, music, sport) he cannot see how **thinking** in any strong or meaningful sense can take place. He is unpersuaded by Turing's insistence that this can be so even though the range of activities pursued by intelligent machines would be, at least initially, quite constrained: **areas like chess-playing, cryptography and mathematics are good candidates, since they require little contact with the outside world.**

This looks, to this non-specialist reviewer, very like the delegation of rule-based, computational **thinking** to machines that seems, currently, to be the main thrust of research into machine assistance for human intellectual effort. To argue that it

is anything more than that, say the skeptics, let alone that the processes involved might be tending in the direction of self-aware cognition, implies the emergence of a subject with access to what Ray Tallis calls *an unrestricted domain of awareness.* He goes on to quote Wittgenstein on the subject: "*A picture held us captive. And we could not get out of it for it lay in our language and language seemed to repeat it to us inexorably.*"

The picture in question would, one feels, look very like the diagrams of computational thinking processes drawn by Turing at Casti's imagined dinner party to illustrate his descriptions of the ways in which the brain's neuronal activity mirrors that of computers.

Wittgenstein's objection to this view of thinking, as rendered by Casti, is fundamental: *meaning resides in social practice, not in logic.* He uses the fish course as an example of what he is getting at: *The naming of this piece of protein we call fillet of sole can take place only within the context of a developed language, one in which there already exist rules for picking out objects, using names and doing operations. The criteria for this are not in the logic of machines, tapes and codes but in the actual practice of a language community.*

If true, how much more, one might argue, would such considerations apply to the examples of suitable activities for thinking machines cited by Turing. Can a machine programmed to compute the possible moves and combinations of moves in a game of chess meaningfully be described as intelligent, let alone need to be conscious in any sense of the word?

According to Casti Garry Kasparov thought so and Casti appears to agree with him. Kasparov said that he could detect an *alien intelligence* in the computer that defeated him. Casti glosses this to suggest that *to Kasparov the program has become a kind of person.*

It's difficult not to sympathise. To devote one's life to being one of the world's greatest chess-players and then lose, in the full glare of publicity, to a machine must have been a demoralizing and disorientating experience. Difficult then to admit that one's opponent possessed neither intelligence nor the ability to care whether it won or lost. An uncharitable reflection, perhaps, but Russia's Grand Master was only human.

Casti's gloss on Kasparov's nemesis is the starting point of the argument which concludes his provocative and entertaining book. He gives advances in machine translation as an example of the developments which will lead, he believes, to the emergence of machines that think but not like humans: *After the current, but brief, interregnum machines and humans will go their separate ways, much as humans and dolphins parted company many millennia ago.*

There are several questions that might be asked about this assertion.

Arguably, translation would seem to fit the category occupied by the three activities itemised by Turing and would, therefore, be open to the same questions about intention and the nature of the *intelligence* required. Appealing, as Casti does, to Chomsky's theory of a universal grammar embedded in the human mind just seems to strengthen the point. Moreover, it is hard to see how machines would make the jump from being programmed to carry out aspects of such quintessentially human activities as processing language to constituting a radically *alien* and separate species of intelligence.

Implicit in Casti's analogy is the question of animal consciousness: which one would evolved machines more resemble? Us or the dolphins? And why would we bother making this possible?

As for machine evolution and self-reproduction, they are raised in the prandial dialogue. Evolution would arise through, for example, mistakes in the transmission of instructions for the assembly of each succeeding generation. Presumably, too, if the machines were equal or superior to us in intelligence, they

could reach the point we have of being able consciously to direct the course of their own evolution.

Such speculation, enticing though it is, bumps up against one or two hard objections. One is raised by Ray Tallis and is an example of a difficulty that lies at the heart of this issue: the ways we use language, especially figurative and analogical language, to frame and debate concepts in science and philosophy. Tallis uses the example of the word *information* and the multiplicity of meanings with which it is invested: *the computational meaning of information, has little to do with the word as it is used in everyday life. It should not be confused with ordinary usage, which refers to knowledge consciously communicated between conscious human beings.*

He goes on to point out that the word is increasingly used now in an analogical or metaphorical sense akin to the idea that some structures retain a *memory* of their own shape and function. A similar point might be made about the encoding of *information* in DNA and, of course, the instructions passed on by one machine to its successors. What matters is whether there is any reason to suppose that a conscious, purposeful agent has caused that information to be passed on. On that possibility, with regards to machines, it seems hard to disagree with Casti's Haldane that it is (at the very best) *not proven*.

The value of Casti's book lies in its very fair and well-informed coverage of the main competing viewpoints on AI and the light it throws on the relationship between science and philosophy (someone should give it to Stephen Hawking). Its value as an introduction to the issue is enhanced by Casti's willingness, in the end, to lay his cards on the table. By doing so and by raising deeply speculative questions about the future of machine *minds* he thereby requires us to sharpen up our thinking about language, the mind, our relationship to technology and what it means to be human. By inviting public engagement with science and philosophy at this level he has done us all a great service.

**Editor's note:** If you like this subject matter, click on this [Artificial Intelligence and Human Consciousness](artificial-intelligence-human-consciousness-oct-2011.html "AI and Human Consciousness") link to read about the Salon discussion on Tuesday 25 October 2011.