AISB Book Review - In Our Own Image - George Zarkadakis, 2015

Robert H. Wortham

September 18, 2015

Zarkadakis' new book presents the human quest for artificial intelligence through three lenses; the historical, the philosophical and the technical. Is AI a future existential threat to humanity, a uniquely new and powerful Pandora's Box that we should not open? Or is it merely another smart piece of technology for us to exploit, no more or less dangerous than the many powerful technologies we have already harnessed? This book neatly presents humankind's quest to understand both the world and itself by creating artefacts and narrative. It argues that in order to properly understand questions about the nature of intelligence, intentionality and consciousness, we need first to understand the origin and development of the stories embedded within human culture that give us tools to explain the world to ourselves. For example, in ancient times it was believed that humans were created from mud, brought to life by some unseen force (often a God or Gods working together). The Greeks believed that we were driven by a mixture of four humours (blood, yellow bile, black bile & phlegm) and remarkably this idea persisted until the 19th century. In recent times we have come to believe that our bodies are machines, and our brain is like a computer, the brain being the computer hardware, and our mind being the software that runs on it. Zarkadakis usefully reminds us that these are all narratives, built using metaphor. We can only understand and explain new knowledge using terms from existing knowledge. Protons are not electrically charged billiard balls, although sometimes that metaphor is useful. Similarly brains are not *actually* computers. It is a sobering fact to remember that all science is still story-telling, although we can argue that our stories are based on empirical investigation and that they have much greater practical utility (our technology, based on science, works!).

Throughout his book, Zarkadakis often revisits the idea that since the time of the Greek Philosophers Plato and Aristotle there has been an ongoing philosophical battle between form and matter for the foundation of existence. Cartesian dualism is essentially Platonic, and we still see a strong dualistic approach to the understanding of mind today, including ideas that the mind will one day be uploadable to a computer and we will cease to need our human bodies at all! The Aristotelian approach conversely places emphasis on our physical substance, from which our 'self' emerges. There can be no human 'self' without a human body.

Humans are clearly very interested in how we came to be, what it means to be a 'self', how we come to be able to recursively experience ourselves experiencing things (qualia), and whether we can create artificial selves.

The exploration of the possibility of creating a person from inanimate matter is not new. In Greek mythology Pygmalion was a sculptor who fell in love with his statue, and successfully asked the goddess Aphrodite to bring it to life. Zarkadakis takes us on a journey through history showing how the Greek Pygmalion myth is retold by Shakespeare in A Winter's Tale, George Bernard Shaw in his novel Pygmalion, Mary Shelley's Frankenstein, and then through films such as Metropolis, Blade Runner, Star Trek (Commander Data), Bicentennial Man and most recently Ex Machina. We have always been fascinated by the possibility of creating a being in our own image.

Zarkadakis spends some time exploring cybernetics, a term coined in 1948 by Norbert Weiner as "the scientific study of control and communication in the animal and the machine", and the related idea of homeostasis – the cybernetic state of dynamic equilibrium. He then takes us through the 20^{th} Century early evolution of AI covering the work of Turing, Shannon, McCulloch and von Neumann, and the significant influence of Crick and Watson and their discovery of DNA. He also points out that wide ranging research is today much less favourably funded than narrower research that might more likely yield quick results, papers and better still, economic return. Thus cybernetics as a research discipline has been relegated to the history books. Zarkadakis hints that this might be a mistake, and I would agree with him here.

According to Zarkadakis, referencing the work of Douglas Hofstadter, Stanislas Dehaene and Kurt Gödel in particular, consciousness arises as a result of 'reflexivity'. That is, multiple levels of feedback loops within the brain that allow it to sense itself, and sense itself sensing itself, and so on recursively. He argues that this self-referencing mechanism *is* the 'self'.

Returning for another history lesson, the book the covers the US military funding of AI, in particular the Dartmouth conference of 1956. The goal was to program computers to perform human mental functions such as learning, solving logical problems and using language. Over the following decades, much was learned about how hard these problems really were, but also how a great deal of human intelligence is really about everyday 'common sense' thinking – something that is also very hard for computers. After the 'AI winter' in the 1970's, the book brings us up to date, covering IBM's Deep Blue & subsequently Watson, the DARPA challenge and Google. Interestingly there is little mention here of the substantial work on AI and robotics carried out at MIT by Rodney Brooks, Cythia Breazeal and many others. Perhaps Zarkadakis sees Brooks' then revolutionary idea that one can have intelligent behaviour without representation as very separate to the other strongly symbolic approaches. The work at MIT is however referenced in a very useful timeline provided in an appendix. This starts at 65,000 BC and ends 4 pages later in 2015! This timeline also serves as a quick reference for Zarkadakis' many and varied influences.

In the final chapters of the book Zarkadakis considers the future. He discusses the likelihood of many white collar jobs being displaced by robots and AI, and whether over time this will lead to better standards of living for us all. He is sceptical that the current route towards AI using von Neumann computing architectures, present day programming languages & algorithmic approaches and relying on Moore's law will achieve human-like intelligence or support artificial consciousness. Zarkadakis appears however to be a convert to the idea that neuromorphic computers will indeed enable superintelligent, conscious and self-willed artificial beings to evolve. Zarkadakis describes how coupling memristors

with capacitors can create an electrical circuit capable of simulating the spiking behaviour of biological neurons. An architecture based on this technology would be essentially analogue, not digital. He extrapolates an implementation based on 'liquid electronics' and simulating the spiking neurons in the human brain. Such a brain, if suitably embodied, would be able to learn like a child and become a unique and conscious individual. Over time, the ability of such a system to self-replicate would harness evolutionary forces and achieve superintelligence. I found this to be a startling transformation within a few pages from the approach of a well informed, scientific skeptic, to that of a hopeful believer. Time will tell whether this futurology is justified.

George Zarkadakis has an engaging and likable writing style, and I certainly would wish to recommend this book, particularly for its excellent coverage of the human search to understand the minds of others and the pitfalls of the inevitable reliance on metaphorical narrative, even within modern science. We would do well to remember that this necessarily introduces a similar problem when we tell stories of what AI does today, and might do in the future.

Rob Wortham
PhD candidate
University of Bath, UK