Psychiatric Bulletin (1993), **17**, 100–103

Computers in psychiatry

4. On artificial intelligence (AI): is there a ghost in the machine?

BANKOLE AKINDEINDE JOHNSON, Wellcome Research Fellow, Oxford University Department of Psychiatry, Psychopharmacology Research Unit, Littlemore Hospital, Littlemore, Oxford OX4 4XN. E-mail: Kole @ UK.AC.Oxford.Vax

"I am superior in many ways but would gladly give it up to be human."
Lieutenant-Commander Data – on android and serving officer on the Starship Enterprise Circa the 25th century at *Encounter Farpoint*. (See Fig. 1).

Striking advances in artificial intelligence (AI) have brought to life philosophical debates about the concept of the mind, and of consciousness. Strong advocates of AI suggest the symbiosis of man and computers is the next step in our evolution, and to control this process, we need to address the "Big Question": what is the meaning of life, and specifically, what are we? Understanding what we are is of great interest to us all; for psychiatrists, philosophers, and psychologists it is their life's work. In this paper, I shall introduce some of the main ideas about AI under the following headings. Can computers have a mind, improve our understanding of, or be integrated with our mental processes?

Can computers have a mind?

Many people are unwilling to consider that computers may, at some time, be able to think, and even, become sentient. The human mind, they would argue, is essentially these qualities: it is what gives us consciousness, separates us from animals (presumably), other beings, and inanimate objects, allows us to exercise freewill and judgement, and is at the heart of, perhaps, our most treasured attributes – insight, inspiration, intuition, and creativity. It is separate from the mechanical processes of our bodies which can be simulated by machines. No matter how efficient a computer becomes at solving problems, it will always be no more than a machine because it has no understanding of what it is doing. This mind–body dualism promulgated by Descartes (1596–1650) suggests there is a distinction between the mechanical process of problem solving – a physical event which is "extended" – and understanding which is a property of the "unextended" mind. He stated that these two entities, mind and body, made

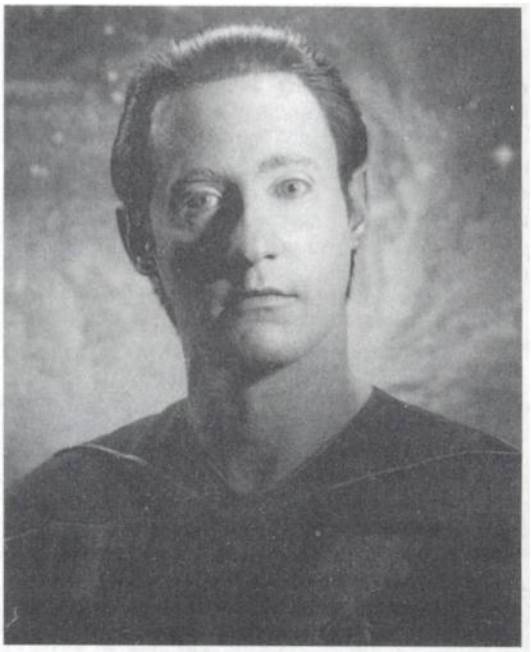

FIG. 1. Lieutenant-Commander Data by courtesy of Simon & Schuster International Group, Herts.

some form of contact in the pineal gland. Descartian dualism has been heavily criticised. Ryle (1990), in his influential book, *The Concept of Mind*, perhaps the definitive rebuttal of dualism, called it the belief of the "Ghost in the Machine". Put simply, Descartes could not explain satisfactorily how the mind and body interacted within one another: it is neither plausible to ascribe a purely "mental" role to the pineal gland, as it is part of the brain (i.e. a "physical" organ), nor is it clear how mental events could produce physical changes – for example, dualism cannot explain how thinking of lifting a limb (a "mental" event) could lead to its movement (a "physical" event). Thomas Hobbes (1588–1679), who was also a critic of Descartes, countered with

what is now known as Cartesian materialism. He argued that mental and physical events were similar; that is, they were alternatives not opposites, each with influence over the other – this view is similar to our modern concept of mental illness whereby a disorder such as depression can have physical consequences (i.e. abnormality of aminergic transmission in the brain); and, both cognitive and pharmacological interventions can be therapeutic.

Spinoza (1632–77), the founder of modern ethics, expanded on Cartesian materialism. He suggested that: (a) all mental events had a physical parallel, like opposite sides of the same coin; (b) the mind was subject to the principle law of nature – physics; (c) God was nature; (d) Man was a machine, and (e) that all thinking machines could possess a mind. Spinoza's view of God as nature, with no personal qualities, was unpopular with the religious leaders of his Jewish community and he was excommunicated in 1656. His ideas questioned the concept of freewill. If we are only an assembly of mental processes, some deemed to be wrong and others right, who or what is really responsible for our actions? Are we capable of completely independent thought and action? Is thinking the only requirement of a mind? While most Libertarians accept that the majority of an individual's actions are determined by who he or she is – either by nature or by a nurture – they assert that everyone has some degree of freewill of which they are aware. This capacity to deviate from a predictable course of events is commonly seen in nature, and is known as Heisenberg's uncertainty principle – for example, light travels as both a particle and a wave; hence, by attempting to define its velocity there is a loss in accuracy of determining its position. Thus, as we observe an event we influence that observed: the fabric of our reality or consciousness is woven into the subjectivity of nature.

In 1950, Albert Turing suggested a method, the Turing test, for determining whether a machine could think. In this paradigm, a computer and a human volunteer are both shielded from another person who poses them questions and receives their replies through a keyboard. If the inquisitor is unable to tell the difference between the computer and the volunteer, the computer is judged to have passed the test. For a computer to perform this feat it has to decide what an appropriate response of a human would be. For instance if the investigator were to ask it to multiply two 13 digit numbers, and a correct reply was forthcoming within a few seconds, the computer would be found out.

There has been some success in creating machines which, under specific conditions, can simulate human responses. As early as the 1960s K. M. Colby devised a computer program which successfully simulated a psychotherapist; indeed, some clients preferred the machine. These Expert Systems are under continuing development, and some have scored notable successes in treating certain phobic disorders.

There have, however, been two main problems with conceptualising a machine that can respond to any question

(a) Computers solve problems using algorithms – hierarchies of probable responses each of which is processed singly (serially) until a conclusion is reached and the machine halts until it receives a further input. Because computers carry out a large number of instructions at the same time its apparent "oneness of mind" is not more than singularity of purpose. In contrast, the human mind is a parallel processor with many different functions being pursued at any given time; most of these are pre-conscious. While it is possible to tackle an algorithm by running several serial computers simultaneously, and in effect, mimic parallel processing, this is wasteful, and does not convey the essence of human thinking. Humans are often able to reach an answer to a problem by a partial solution of the alternatives. Dennett (1992) saw this drawing together of ideas, albeit virtual, as what gives individuals the same singularity of thought as a computer; the different configurations in which these impressions can be brought together correspond to "multiple drafts" of consciousness. Penrose (1989) argued that a computer designed to mimic the brain not only has to work in parallel but has to obey the laws of quantum mechanics – different parts of it will need to exist in alternate states but only one, the sum of the component parts, would be perceived to have taken place. Advocates of AI suggest there is some evidence of quantum processing in the brain. Retinal cells are able to perceive a single photon of light; the response is theorised to cause enough disturbance (technically one graviton) to surrounding neurones as to alter their state. Thus, both neural networks (see Fig. 2) and digital computers in this paradigm may be able to overcome the limitations of their usual on/off mode of transmission – for transmission, nerves use action potentials, and computers employ electrical pulses. This allows for an enormous degree of complexity. Nevertheless, Penrose (1989) is not convinced that the case for quantum parallelism in the brain is proven because isolated phenomena are likely to get lost in the electrical "noise" of the brain. As yet, there is no good evidence to demarcate where a site of central quantum activity might be. The reticular formation, a potential site, appears to function in much the same way as an on/off light switch. My guess, if I may be granted the indulgence, is that continuous consciousness may lie where there is an abundance of spontaneously firing neurones – to generate the quantum field, a "reinforcement" driven gate to propel the process may be necessary, and a link between pre-conscious and conscious processing must be forged. Potential sites may include

the median forebrain bundle – ventral tegmental area – hippocampal – thalamic axis including their projections to the cortex; it is, at least, hypothetically possible for quantum parellelism to be generated in these areas by spontaneously firing neurones.

An extension of the multiple draft or alternate state hypothesis is the concept of the teleportation machine of science fiction. All atoms are the same – the ones in our bodies are like those in any other form of matter; the difference lies in their configuration. In computer metaphor, the hardware is essentially the same; it is the software that makes us what we are. If an individual is duplicated exactly at the atomic level, teleported (transferred as a beam of light) to another destination, and then reassembled, the original having been destroyed, the copy retains self-awareness. Importantly, to comply with the laws of quantum parallelism the original has to be destroyed by the copying process; that is, alternate states existed – only one version became manifest. If the person is copied, but instead of being transported he or she is stored and then re-configured on the appropriate magnetic material of a computer, the person retains all their properties and can communicate in this different form with the device. While, teleportation must, for now, remain a myth it is estimated that within the next 50 years there will be a computer sophisticated enough to store all the experiences and characteristic responses of an individual and simulate a personalised environment to convey self-awareness; thus, long after the individual has died it will be possible to ask the person's computer image questions and obtain replies – this has biblical overtones of the Resurrection.

(b) The mathematical problem of constructing an algorithm complex enough to simulate the brain has also been questioned. It is known as the *Entscheidungsproblem* – is it possible to devise a general mechanical procedure which can, in principle, solve all mathematical problems? Gödel showed a true solution required human determinism, and that a formal procedure could never be a substitute. In other words, a computer cannot solve a non computable problem because it has no insight; this supported Kant's earlier view that, in both physics and mathematics, the mind's ability to determine the truth of a statement goes beyond a purely logical analysis of the problem. In gainsay, despite its practical drawbacks, Russell's paradox – a complex explanation of how a set of characteristics can be used to describe itself – is sometimes cited as a refutation of Gödel's theorem.

In sum, at an estimated million-fold increase in computing complexity every 20 years it is conceivable that, before the end of the next century, computers will have the processing power to simulate the brain of lower vertebrates – but what will their personalities be like?

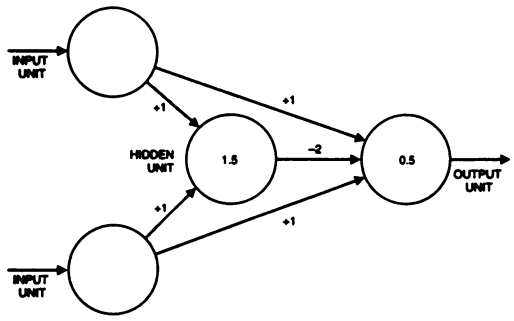

FIG. 2. What are Artificial Neural Networks? Neural Networks (NN), like biological neurones, are able to detect and interpret incoming data, and to learn the relationship between the input (stimulus) and its consequence (i.e. the output or response). They are made up of *units*. Links between units serve as its axons and dendrites; the unit itself is analogous to the cell body. Links conduct the summed *weight* of the input signals to the output units. Changes in this weighting factor alter response; in effect, variations in neuronal firing patterns adjust the strength of synaptic connection or *learning*. Various types of neural learning strategies have been developed. Of, perhaps, greatest relevance to my interest is *reinforcement* learning which might provide mechanisms for exploring addictive behaviours; others include unsupervised learning – which has been suggested to be a model for visual recognition and primary cortical processing, and supervised learning. At its simplest, NN consist of three layers of units: input, "hidden", and output units. In the worked example above of a "threshold" system, the threshold value of the hidden unit is 1.5. If the summed inputs exceed this value, the impulse is propagated; importantly, inputs can bypass the hidden units (analogous to neuromodulators) and activate the output units directly. While in biological neurones the production of an action potential has threshold characteristics, the output has a sigmoidal function. NN "learn" by making smaller errors each time they are presented with the same problem. In this way, it is possible to train a NN to learn the relationship between a stimulus and its response without having any knowledge of what the weights should be in the first place. In psychiatry and psychology, this may offer new insights into the pathophysiological functioning of the brain which may bear little resemblance to its observed neuronal structure. That is, we will be able to observe the phenomena as they are rather than as at present – how we interpret them to be. The details given below of how to train a NN might interest someone wishing to do so. In back-propagation, a training technique for NN, the weights of the units must be adjusted so that there is minimal difference between the desired and actual outputs. That is, the rate of change in the error of the network as the "weighting" between units change must be calculated. For a network with a representative output unit j, and i a characteristic unit in the previous layer, find the total weighted input x_j; $x_j = y_i W_{ij}$, where y_i is the activity level of the unit in the previous layer and w_{ij} the weight of the connection between the previous and output layers. The activity level of the output unit is given by $y_i = 1/(1+e^{-x})$, and the network error $E = \frac{1}{2}\Sigma (y_j - d_j)^2$, where d_j is the desired output. All that is left is to calculate the rate of change in the error (error derivative) for the different layers in the unit: $\sigma E/\sigma w_{ij} = y_i y_j(1-y_j)\beta_j$, where $\beta_j = y_j - d_j$ for the output units and $\beta_j = \Sigma_k w_{jk} y_k (1-y_k)\beta_k$ for the previous (including hidden) units; k is the number in the next layer j is coupled with. In essence you find the forward pass value of β_j for the output unit, and then obtain the back-propagation value for all layers from the last to the first $\beta_j = \Sigma_k w_{jk} y_k (1-y_k)\beta_k$. To change the weights in the network use $-\Delta w_{ij} = -\sigma y_i y_j (1-y_j)\beta_j$.

The figure is by courtesy of *Drew Van Camp* and *Scientific American*.

https://doi.org/10.1192/pb.17.2.100 Published online by Cambridge University Press

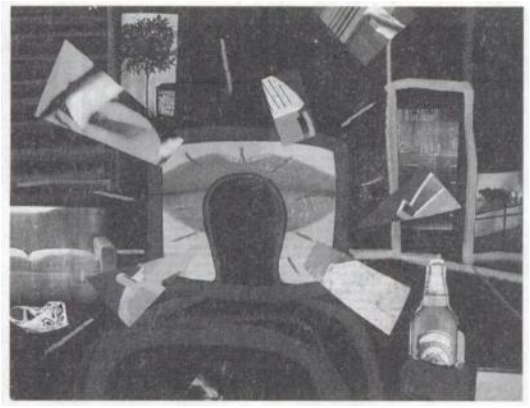

FIG. 3. Virtual Reality by courtesy of Huntley/Muir Design Studio, London.

Can computers improve our understanding of, or be integrated with our mental processes?

The first robots were designed to operate in extreme environments such as the inside of nuclear reactors. However, their perceptual skills were crude, and the manipulations lacked the degree of dexterity that could be afforded by a human operator. In 1979, Marvin Minsky invented the term "telepresence"; he suggested that what was needed was a way to use the highly developed cognitive and perceptual skills of humans to fine tune the control of robots. Obviously, the retina would be the link up point between the robot and the "mind's eye". The technical problems were enormous. The robot's scanning device had to: (a) be of such high resolution as to produce a uniform background – a difficulty which has been overcome partly by using fibre-optic, and soon, laser micro scanners; (b) use its two dimensional images to produce three dimensional representations in the eye; (c) be able to follow the tracking movements of the eye and head rotation and (d), the actions of the robot had to be felt and correctly interpreted by the human operator.

The combined solution of these problems has resulted in what we now know as a virtual reality (VR) helmet and suit (see Fig. 3). The next step is to provide the robot with three dimensional vision which is processed with brain-like parallelism; thus, it becomes not only an extension of the individual's field of vision but he or she is also contained within projected image – cyberspace. For example, by locating a surgeon's sight at the end of such a device located within a blood vessel it would be as if he or she were actually there, and surgery can be performed at almost a cellular level of precision.

Psychology will benefit from a greater understanding of how the brain processes visual information. An obvious application will be to provide people who have lost the use of their limbs – either through physical handicap, disease, or an accident – not only with working limbs but ones which feel like a part of their own body. Another would be to help individuals following a cerebrovascular accident to recognise the part of their body they are neglecting.

In psychiatry there are potential uses and abuses of this technology. Psychotic disorders often produce perceptual disturbances which may be related to abnormalities of information processing. VR will not only allow psychiatrists to test this hypothesis, but its ability to create personalised environments may enable experiences such as hallucinations to be filtered out or over-ridden by more suitable images. In addition, accelerated learning may be possible for those with mental handicap. Anxieties about the misuse of VR will be familiar to those who have experienced the almost hypnotic effect of computer games. VR offers even more. It allows its user to enter into a personalized "trip" in much the same way as someone who has used an hallucinogen – in effect, electronic LSD. On a larger scale, some may argue that VR is not a new science, but simply, the ultimate scientific application of an old one – "mind control". Will addictive behaviour be the final frontier of psychiatry? Additionally, with the advent of teledildonics – the ability to experience erotic feelings over telephone wires, society will, perhaps, need to review its concepts of mortality and freewill.

Conclusion

There is little doubt that AI and VR will bring about great advances in medicine and psychology. Paradoxically, these advances are likely to re-establish that Man is central to, and not an accidental by-product of, the forces of nature. The concept of the mind and its links with mental processes has, rightly, achieved scientific credibility. Our human subjectivity, it appears, may turn out to be what gives reality meaning.

Acknowledgements

I wish to thank Dr L. T. Wells.

The Starship Enterprise features in the science fiction series *Star Trek, The Next Generation* created by the late Gene Roddenberry and released by Paramount Pictures.

References and further reading

CHURCHLAND, P. S. & SEJNOWSKI, T. J. (1992) *The Computational Brain*. England: MIT Press/Bradford Books.

DENNETT, D. C. (1992) *Consciousness Explained*. London: Allen Lane.

PENROSE, R. (1989) *The Emperor's New Mind*. Oxford: Oxford University Press.

POPKIN, R. & STROLL, A. (1990) *Philosophy – Made Simple Books*. Oxford: Heinemann Professional.

RHEINGOLD, H. (1991) *Virtual Reality*. London: Quality Paperbacks Direct.

RUMELHART, D. E., HINTON, G. E. & WILLIAMS, R. J. (1986) Learning representations by back-propagating. *Nature*, **323**, 533–536.

RYLE, G. (1990) *The Concept of Mind*. London: Penguin Books.