The reel of consciousness

Nearly 40 years ago, the Princeton psychologist Julian Jaynes opened his extraordinary book, *The origins of consciousness in the breakdown of the bicameral mind*, with a lavish item of purplish prose on the phenomenon of higher order human consciousness:

'O. what a world of unseen visions and heard silences, this insubstantial country of the mind! What ineffable essences, these touchless rememberings and unshowable reveries! And the privacy of it all! A secret theatre of speechless monologue and prevenient counsel, an invisible mansion of all moods, musings, and mysteries, an infinite resort of disappointments and discoveries. A whole kingdom where each of us reigns reclusively alone, questioning what we will, commanding what we can...' (Jaynes, 1976)

The intervening decades have seen a prodigious circuit of academic monographs and popular books, as well as thousands of articles in periodicals such as *The Journal of Consciousness*, attempting to explain what philosophers would describe as the ‘hard’ problem. How can objective descriptions encapsulate subjective states? The many stratagems—from brutal frontal attacks to stealthy tunnelling—have been reminiscent of a prolonged siege on an impregnable citadel.

The expanding discipline of neuroscience, aided by new scanning technologies, and experimental work on Artificial Intelligence, gave fresh impetus to the task, especially during the late 1980s and the 1990s, the so-called ‘decade of the brain’. At the same time, a new breed of neuroscientifically informed philosophers of mind had arisen, their qualitative conclusions prompting remarkable contradictions, antagonisms and best-selling books. They can roughly be divided into those who insist that consciousness remains an intractable mystery, resistant to scientific probing; and those who maintain that consciousness is really no big deal—that we have pretty well cleared up most of the difficulties and could even replicate the phenomenon in an artefact.

Now come two prominent veterans, a scientist and a philosopher, offering their considered afterthoughts in the light of protracted peer exchanges down the years. Has time encouraged them to alter their earlier cherished opinions? Christof Koch, in *Consciousness: confessions of a romantic reductionist*, remains as adamantly reductionist as he was 20 years ago, when collaborating with Francis Crick on the correlation between consciousness and 40 Hz oscillation of cortical neurons. Yet he allows himself to wander into the realms of sentimentality on the subject of his beloved pet dogs, and a decidedly scientific form of mysticism about the sentient nature of the cosmos. Koch’s work is a 180-page laconic notebook, the prose at times reminiscent of a rap session (he even quotes Eminem and refers to his own writing as a ‘riff’). He tells us that he has deliberately broken away from the ‘tired third person accounts of consciousness enquiry’ in order to place his theorizing within the story of his life.

Daniel Dennett, in *Intuition pumps and other tools for thinking*, also remains doggedly attached to the reductionist explanation...
that he first mooted more than 20 years ago in Consciousness explained. But something has happened to his prose and manner of exposition: there is an impatience—an economy reminiscent of telegraphese—less like reading a book than a series of newspaper columns or emails.

Despite the intransigence of their convictions, our authors' mode of writing draws attention to the distance travelled these past four decades in the varieties and quality of prose and literary forms employed by exponents of consciousness studies. What kinds of book has the consciousness problem generated since Jaynes? Has it produced a genre of original writing of high quality and permanent value? What do these writings tell us about the minds (and hearts) of those that write them?

Jaynes, who died in 1985, proposed a solution to the problem of consciousness with a weird theory about the split mind of ancient man—akin to schizophrenia. He argues that early man had no such thing as consciousness, but was zombie-like. In moments of stress a man would hallucinate voices, which he took to be the word of a god. Jaynes's book is an ambitious sprawling epic, ranging through nearly 500 pages of poetry, pre-history, ancient history, languages, myth, theology, psychology, psychiatry and anthropology. His genre is a fragmented, speculative history of ideas, given unity by an air of literary dandyism, and what he calls 'my personal narratizations', in other words—memoir (Jaynes, 1976). The book's sheer bombast and outlandish hypotheses guaranteed its doom as a literary form to be emulated.

The books that followed in its wake demonstrated the subject's scope for wide-ranging literary treatment of varying quality. The first, and still the most surprising rendering, was the expansive (close to 600 pages) conversation between Karl Popper and John Eccles entitled The self and its brain: an argument for interactionism. Eccles was a neurobiologist, chiefly remembered today for his initial conviction that synaptic transmission is primarily electrical, and his later acknowledgement that chemistry was also involved. In 1977, he collaborated with the philosopher Karl Popper in an attempt to demonstrate their belief in a theory that went even further than Cartesian dualism. They defended a hypothesis of 'trialism': the interaction between sense perception, thought and the soul, as if there were three substances, or 'worlds', as they called them, involved in consciousness. In the series of 12 dialogues at the book's conclusion, Popper, more than Eccles, concedes the importance of the 'limitations of explanation [of consciousness]: the fact that we can never achieve explanations which are fully satisfactory in the sense of being ultimate' (Popper and Eccles, 1977). All the same, both men, incapable of accepting the purely material nature of mind, or the possibility of a straight correlation between mind and brain, were unrepentant defenders of the existence of the spiritual soul. Eccles, it has to be said, was less happy with Popper's apparent agnosticism on the score of ultimate explanations, possibly because he was a Catholic. Koch, incidentally, who is militantly allergic to dualistic, let alone trialistic thinking nevertheless instructs his readers not to scorn them for such heterodoxy. 'They are not', he writes, and here you get a flavour of his prose, 'your typical kooks spouting quantum flapdoodle about Shrödinger's cat, entanglement and the interconnectedness of things.'

Their attempt, however, to reinstate the soul appeared more of a final gasp of outmoded Cartesianism than a new beginning. Not only had dualism, or, as some put it, 'belief in the spooky stuff', been seen off by such empiricists as Gilbert Ryle, A.J. Ayer and others of the analytical school of philosophy, but the dominant explanatory model, or metaphor, for the brain had opted for the computer (hardware/software), discarding an earlier reductionist preference for telephone exchanges. The situation was eloquently, perhaps tragically, described by the American philosopher William Barrett in his The death of the soul (1987). His elegant book-length essay pleaded with scientists, philosophers and the common reader to reject models of the mind-brain based on any kind of machine. No machine however large and complex, he argued, has the capacity to engage in such activities as poetry, and in any case:

'a machine cannot mature, for it is not an organic body, growing and ripening through time. As a piece of equipment, it becomes used and defective, its wires frayed and its tubes burned out, and shortly ready for the scrap heap. That might be metaphorical description of some human lives, but only a very nihilistic and reductive one' (Barrett, 1987).

In the same year as Barrett's lament for the demise of the soul in late 20th century Western culture, the scientist Gerald Edelman, whose early work on the immune system had won him a Nobel prize, produced a book that met Barrett half-way. Edelman's The remembered present: a biological theory of consciousness (1987) postulated a mind-brain theory that invoked a metaphor closer to a Brazilian rain-forest, a living bio-mass of birth, death and regeneration, than any kind of machine. To overcome the mysteries of mind, spiritual and poetic, Edelman cited the prodigious numerosity involved in neuronal connections (more than all the particles in the Universe!), and an unfamiliar theory which looked like feedback loops, but more complex, called 're-entrant signalling' (in a later, more accessible, book he likened the process to the action of a symphony orchestra). Written in highly technical language, with specialized figures and diagrams, Edelman's contribution to the debate carried the stamp of a scientific manual or text-book; yet, he concedes an important, wholly accessible perspective in his final chapter, revealing sympathy with Barrett: 'While science can serve as our best means for verification', he wrote, 'it cannot replace either our commonsense view or philosophy, whatever the initial errors we risk in either of these enterprises' (Edelman, 1990).

If Edelman was more 'human', more open to aesthetic sensitivities, reviewers found his prose maddening, and he paid for it in unfriendly reviews. Francis Crick picked on his neologistic phrase 'neural Darwinism', and dubbed it 'neural Edlemanism.'

As if in response to Edelman's point about numerosity, the Rouse Ball Professor of Mathematics at Oxford, Roger Penrose, entered the fray with a book that belonged squarely in the genre of history and philosophy of physics and mathematics. It was also something of a campaigning text. He was anxious about the proposition that we might one day become slaves to the machines of our own making, and was bent on exploring whether this might be even a remote possibility. Penrose's The emperor's new mind (Oxford, 1989) was a tour de force against simplistic notions about the relationship between mind-brains and digital computers.
At its heart, Penrose argues that humans are capable of engaging in kinds of mathematics that would defy the capabilities of digital computers however large or complex. Citing Kurt Gödel’s famous proof, which demonstrates that no non-trivial axiomatic system can be both complete and consistent, he suggests that the brain must be a kind of biological quantum computer that mediates somehow between the physics of the very small and the very large. Edelman was to reject this proposition as surrogate Cartesianism. Others questioned whether living brain matter was the right temperature for quantum effects.

Penrose was clearly less preoccupied with the nature of human consciousness than a related question: can consciousness be replicated artificially? The question harked back to Alan Turing’s thought experiment about how one might tell the difference between an actual person and a well-programmed computer sending messages from behind a screen. Penrose, in alliance with the philosophers John Searle (of the Chinese Restaurant argument), William Barrett, and Colin McGinn, were convinced that consciousness had something to do with the fact that brains were made of ‘living’ stuff rather than transistors. And beyond that notion lay another conviction: to accept that machines could be conscious was to believe that we ourselves are no more than mechanical. As Marvin Minsky had put it, the brain is just a ‘meat machine’. As if to demonstrate that mechanized consciousness seemed perfectly feasible, even to a biologist, the geneticist John Maynard Smith sided with the mechanists in his review article of The emperor’s new mind in the New York Review of Books (13 March 1990): ‘As a geneticist, I am prejudiced against the idea that the peculiarities of living organisms arise because of the special nature of living material.’

Also reviewing Penrose, Dennett opined that consciousness understood as a ‘complex computation mechanically performed’ is a ‘beautiful idea’. To others, he went on, the idea ‘is deeply repellent: philistine, reductionistic (in some bad sense), as incredible as it is offensive. His sympathies were entirely with the former’, he went on. Popular culture, moreover, seemed to be on the side of the machines: 1990 was the year in which a computer program, Deep Blue, defeated several Grand Masters at chess. The following year saw the release of Terminator 2: Judgment Day, in which Sarah Connor admires cyborg Schwarzenegger’s care of her son John, since the algorithm-based machine appears more kind, reliable, more ‘human’ and paternal, than any biological father. That year also saw the publication of Dennett’s Consciousness explained.

Highly readable, ebullient, irreverent, running to more than 500 pages, Dennett’s ‘explanation’ takes within its ambit an enormous array of reading—in literature, psychoanalysis, behavioural psychology, philosophy of mind, neurobiology and general history of ideas. Citation, however, is no guarantee of acknowledgement: ‘I have sometimes deliberately ignored what these thinkers consider the best parts of their theories, and have mixed together ideas drawn from “hostile camps”, but I suppressed these messy details in the interests of clarity and vividness’ (Dennett, 1991). A central, clear and vivid contention of his great ramble through the highways and byways of myriad notions and rationincisions of consciousness is his argument that there is no central location within the mind of understanding and will: or as he puts it ‘we are shedding our bad habits of thought…the demise of the Central Meaner is more generally the demise of the Central Intender, but the Boss still lives on in other disguises.’

The curious thing about Dennett’s book as a piece of literature is that it owes less to philosophy of mind (his proper field of endeavour) than to encounters with once fashionable literary theories of a French kind, e.g. Jacques Derrida. The conscious self, like the deconstructionist’s illusory author, is no more than a series of multiple drafts, in Dennett’s view, a notion originally propounded in Paris and transplanted to English departments in British and American universities in the 1970s. ‘Most of these fragmentary drafts of “narrative”,’ he writes, ‘play short-lived roles in the modulation of current activity but some get promoted to further functional roles, in swift succession, by the activity of the virtual machine in the brain.’ The self does not really exist, you do not really exist: you are no more than a ‘Center of narrative gravity’.

Dennett himself, or rather his narrative centre of gravity, pays homage in respect of the self’s evanescent status, or human identity as photocopying machine, not to science or philosophy, nor even the realms of the original literary theorists, but to a novelist’s satirized version of them found in David Lodge’s Nice work (1988). Robyn the female English lecturer character, an academic pain-in-the-neck, high on deconstructionist jargon, expatiates to her students that ‘there is no such thing as the “Self” on which capitalism and the classic novel are founded…there is only a subject position in an infinite web of discourses…’. Thus, 20 years ago, Dennett was espousing a ‘theory’ of human identity already scorned by one of academe’s noted critics of deconstructionist gobbledegook.

Today, in his new book, Dennett shows no readiness to return to commonsense convictions about consciousness and the self. In the section that deals with consciousness, ‘Tools for thinking about consciousness’, Dennett does his level best to persuade readers that they are not to trust to their intuitions about the self and consciousness as special to human nature. He has always been forthright, but I have never come across passages so routinely dismissive of his readers: ‘some folks hate the idea of our uncovering the secrets of how the conscious mind works…they have a way of inflating their notions of what consciousness must be, of bamboozling themselves’. Such folk, he avers, are suffering from ‘impoverished imagination’. So what is this big mistake that Dennett gets so worked up about? It is nothing less than his insistence not only that conscious robots are possible, but that we ourselves are just a sort of meat robot. He is right of course to suspect that there are people who don’t like this idea; but why so angry? He addresses the bamboozled reader directly: ‘What you mean is that you won’t conceive of a conscious robot. You think it silly and preposterous to try to take the idea seriously.’ Underpinning this petulant prose is his old denial that there might be something special that it is like to be an aware robot. You think it silly and preposterous to try to take the idea seriously."

Christof Koch is no philosopher, and makes no reference whatsoever to Dennett, even in his bibliography. By the same token, there is only one brief mention of Koch in Dennett’s text, as a side-kick of Crick. Such Soviet style eliminations are common in the consciousness wars. Yet, in common with Dennett, Koch
insists that rather than focusing on the ‘hard’ problem of consciousness, scientists and philosophers should concentrate on ‘studying which bits of the brain are important for it.’ This allows no provision for precise brain correlates with the phenomenon of ‘qualia’: what it is like to experience the world—the actual nub of consciousness. Koch, however, is prepared in the final stages of his book to take us into realms where few theologians (ever since the Vatican relieved Father Teilhard de Chardin SJ of his teaching licence for similar ruminations), let alone students of consciousness, ever venture. Perhaps, Koch speculates, consciousness ‘is a fundamental feature of the universe, rather than emerging out of simpler elements’. He goes on: ‘the hypothesis that all matter is sentient to some degree is terribly appealing for its elegance, simplicity and logical coherence’. Perhaps the entire cosmos, he hazards, ‘is suffused with sentience.’

This reminded me of a peculiar event that occurred in Cambridge after World War II. Ludwig Wittgenstein’s friend Norman Malcolm once wrote of the great philosopher’s complaint that science in schools was taught in too abstract a fashion (Malcolm, 1958). He took, as an example, the practice of teaching astronomy through equations about forces, which are noted and remembered by pupils without a felt sense of the physical reality of gravitation from a phenomenological point of view: that there is something that it is like, in other words, for the sun, earth and moon to be in orbital progress through space. Illustrating his point, to the amazement of passers-by, while proceeding along the street, he assigned to Malcolm’s wife the role of the sun. She was instructed to walk sedately along, while Malcolm walked quickly around her, and Wittgenstein, in turn, ran frantically around Malcolm.

The lunatic exercise constitutes a peculiar Wittgensteinian game in which the pathetic fallacy is deliberately employed to illustrate the nature of unconscious, physical forces and their relationships. Dennett’s attempt to subordinate the conscious to the unconscious—to imagine ourselves as selfless, choiceless, zombies whose notion of consciousness is an illusion—turns the pathetic fallacy inside out. It asks us to suspend our sense of human identity and to see ourselves as composed of the identical blind, determinist, material forces that regulate the rest of the material cosmos. As for Koch’s notion that we inhabit a sentient, conscious cosmos: elegant as it might seem to some (or flapdoodle to others) that, surely, was not at all what Wittgenstein had in mind as he and his companions performed their strange reel through the university town.

John Cornwell
Jesus College, Cambridge

References