



NewScientist

The Big Questions: What is consciousness?

- 18 November 2006 by [Paul Broks](#)

New Scientist tackles eight of the deepest challenges faced by science - from reality and consciousness, to free will and death, in [The Big Questions](#) special features.

On this special day, my 121st birthday, it is good to be surrounded by those I love. There's no denying I feel old, but in body, not spirit. Oh dear, there I go, slipping into the old ways of thinking: mind and body, spirit and substance. There's no excuse. The ghost in the machine was exorcised long ago - and here's Celeste, my sweet, uploaded daughter: the living proof.

She kisses my aged forehead. Chronologically, Celeste is 90 years old. Physically she's a genetically re-engineered woman of 30. Psychologically, well, these days you have to keep an open mind about psychological ways of being. But one never stops worrying about one's children, and the uploading - the transfer of information from old brain to new - was, I confess, a little troubling. I have always felt some responsibility for the current popularity of mind transposition. I helped create a climate of acceptance. Forgive me if I reminisce.

WHAT was that I wrote? "The laser beams of cognitive neuroscience are beginning to penetrate the philosophical fog of centuries"? Tosh! The real philosophical fog was just beginning to roll in. But that wasn't how it felt at the time. Mind science was coming of age. Traditional methods of correlating brain damage and behaviour combined with neuroimaging and computational science to produce ever more refined models of the working brain and, by 2012, the microarchitectures of cognitive function were rapidly unfolding. Writing in 2005, the inventor and futurist Ray Kurzweil had predicted that the brain would be fully "reverse engineered" by the mid-2020s, with hardware and software available for the implementation of human intelligence in a non-biological substrate. He was not far wrong, and "consciousness" went the way of phlogiston, the theoretical substance that scientists once used to explain fire. Strange we ever thought the problem hard.

But as our post-millennial neuroscientists marvelled at the sparkling, dare I say spectral, patterns cascading from their high-resolution brain scanners, they were nagged by a mischievous question: who's running the show? How does the brain, with its diverse and distributed functions, come to arrive at a unified sense of identity? "Soul" doesn't figure in the lexicon of neuroscience, but what about the soul's secular cousin, "self"? Could we speak of a person's brain without, ultimately, speaking of the person? Was the self merely the sum of its cerebral parts? The illusion of the ghost in the machine was compelling - the natural intuition that somewhere in the shadows of the brain there lurks an observing "I", an experiencer of experiences, thinker of thoughts and controller of actions.

How does the brain, with its diverse distributed functions, come to arrive at a unified sense of identity?

This was hard to reconcile with the material facts (the vacant machinery that actually packs the skull) and it was plain to see that the mental operations underlying our sense of self - feelings, thoughts, memories - were dispersed throughout the brain. There was no homuncular assembly point where a little soul-pilot sat watching the dials of experience and pulling the levers of action. We were, neuropsychologically speaking, all over the place. And anyway, who did we think was pulling the levers in the little soul-pilot's head? If we found a ghost in the machine we'd have to start looking for the machine in the ghost.

Belief in an inner essence, or central core, of personhood, was called "ego theory". The alternative, "bundle theory", made more neurological sense but offended our deepest intuitions. Too bad, I thought. We should learn to face facts. The philosopher Derek Parfit put it starkly: we are not what we believe ourselves to be. Actions and experiences are interconnected but ownerless. A human life consists of a long series - or bundle - of enmeshed mental states rolling like tumbleweed down the days and years, but with no one (no thing) at the centre. An embodied brain acts, thinks, has certain experiences, and that's all. There is no deeper fact about being a person. The enchanted loom of the brain does not require a weaver.

Parfit devised a famous thought experiment. Imagine being teleported. A special scanner records the state of every cell in your brain and body and digitally encodes the information for radio transmission. Your body is destroyed in the process but reconstructed as soon as the signals are received and decoded at your destination. You "arrive" in precisely the same condition that you "left", identical in body, brain and patterns of mental activity. Your memories, beliefs, plans, skills and emotions are perfectly intact and you go about your business feeling and believing that nothing about you has changed in the slightest. It's just like waking from a dreamless sleep and getting on with the day.

If you are comfortable with this scenario then you should be comfortable with bundle theory. You appreciate that the observing "I" is no more than patterns of energy and information, which can be disrupted and reconstituted without

destroying the self - because there is no self to destroy. The patterns are all. If, on the other hand, you believe that some essential "you" would be lost in the process then you are an irredeemable ego theorist. You believe that the reconstituted body is not "you" but a mere replica. Although the replica will know in its bones that it is the very person who stepped into the scanner at the start of the journey, and friends and loved ones will agree, you insist it could not be you because your body and brain would have been destroyed.

Incidentally, we see here a neat inversion of conventional thinking. Those who believe in an essence, or soul, suddenly become materialists, dreading the loss of the "original" body. But those of us who don't hold such beliefs are prepared to countenance a life after bodily death.

The philosophical speculations were intriguing, but the science of selfhood also had more practical concerns. This was the dawn of a new age in neuropsychiatry. The idea that certain forms of insanity were "disorders of the self" had been around for two centuries and more, but now the concept was being refined. The core deficits of autism and schizophrenia, for example, were revealed as faults in the brain circuits underlying personal awareness. This confederation of networks - frontal, limbic, temporal and cerebellar - orchestrated social cognition, from the analysis of gaze direction and facial expression to the deciphering of beliefs, attitudes, and intentions. In the process, it gave definition to that fundamental unit of social intercourse: the person. Just as the brain had evolved systems for guiding interaction with the physical world so, we rather belatedly realised, it had also evolved specialised mechanisms for enabling the interaction of "self" and "other".

The discovery of "mirror neurons" in the 1990s was a breakthrough in this regard. According to Vilayanur S. Ramachandran, one of the leading neuroscientists of the era, it was a discovery as significant in its way as Crick and Watson's decoding of the structure of DNA. Mirror neurons were activated not only in response to self-generated behaviour (reaching for an object, say) but also in response to actions performed by other individuals. Pain and emotional behaviour were similarly mirrored. The implication - that minds were neurologically "bridged" - was far-reaching, and mirror neurons rapidly took their place in theories of developmental psychology and moral behaviour.

The self had entered the neurobiological laboratory. Around this time it also became evident that, rather than being a single "ghost in the machine", we were a composite of two phantoms. The self of the present moment - the so-called "minimal" or "core" self - was, in the words of the neuroscientist Antonio Damasio, "a transient entity, recreated for each and every object with which the brain interacts". It was bound to brain systems involved in mapping and regulating body states. The other phantom was the "extended" self: a unified, continuous being journeying from a remembered past to an anticipated future, with a repertoire of skills, stores of knowledge and dispositions to act in certain ways. This "autobiographical" self emerged from language and long-term

memory networks. Michael Gazzaniga, one of the great pioneers of cognitive neuroscience, pointed to a specialised left-hemisphere system - he called it "the Interpreter" - whose function was to wind disparate strands of brain function into a single thread of subjective experience. It worked by identifying patterns of activity across different brain modules and correlating these with events in the external world: it was a teller of tales.

The minimal self gave us our sense of location and boundary, and our intuitions of agency - the feeling that we exercise control over our actions. But these fundamentals of self-awareness were rather fragile constructs. Disturbances of temporal and parietal lobe function could cause profound dislocations of perception such as out-of-body experiences and autoscopic hallucinations (seeing one's body in extrapersonal space). Damage to the frontal lobes could disturb the sense of agency, with limbs developing a recalcitrant will of their own.

The extended self, too, was neurologically fragile. It could be gradually dismantled by dementia, or shattered by a sudden viral attack, the story of the self dissolved with the dissolution of memory. In contrast, a deep-brain stroke or injury to the frontal lobes could leave memory unaffected but recalibrate the machineries of emotion and temperament. The story continued, but the central character had changed beyond recognition. Sometimes the brain's story-telling mechanism itself broke down, resulting in the confabulation of fictional, often fantastical, autobiographical distortions. As science writer John McCrone put it, we are all just a stumble or burst blood vessel away from being someone else. Selfhood is malleable. That was the message.

The neurological diseases that were then still prevalent tended to carve human nature at its joints in such ways, and one occasionally saw what appeared to be clear dissociations of the two "selves". I remember an epileptic patient telling me of her intermittent loss of identity, a condition known as transient epileptic amnesia. Her surroundings would suddenly feel unfamiliar, and then she would begin to feel unfamiliar to herself. Soon she had no idea who she was, where she was or what she was doing. She was stripped to the minimal self: a floating point of subjective awareness untethered by identity.

In other, rare, cases I saw the opposite: the minimal self dissolving, leaving only the story of the extended self. One patient had a strong sense of identity and autobiography but believed that she had ceased to exist. "Am I dead?" she asked. This condition, Cotard's syndrome, was due to a neurological decoupling of feelings and thoughts. Thinking that one exists was not enough: the notion had also to be felt - "I feel I think, therefore I am."

Thinking that one exists was not enough: the notion had also to be felt - "I feel I think, therefore I am"

Another Cotard's patient believed that her voice was all that was left of her. She was "just a voice, and if that goes, I won't be anything". We all have an inner

voice, a stream of sub-vocal speech. It keeps the story going and helps sustain the illusion there's "someone home". One man, recovering from a stroke that had virtually abolished his capacity for speech, including self-talk, described the condition of total wordlessness as being like confinement to a continuous present.

But these words you are now reading, whose are they? Yours or mine? The point of writing is to take charge of the voice in someone else's head. This is what I am doing. My words have taken possession of the language circuits of your brain. I have become, if only transiently, your inner voice. Doesn't that mean, in a certain sense, that I have become you (or you me)? It's a serious question. Written text is a primitive but powerful form of virtual reality. In the beginning was the word.

And in the end? A liberating truth. There are no souls, only stories. I have witnessed a Copernican revolution of the self; a historical shift from the age of solipsism, when we were all at the centre of the universe - self-loving, self-loathing, self-absorbed - to an era of self-dispersion when ego is deemed constrictive. I saw the science of selfhood figure increasingly in the great social and moral debates of the century, from age-old wrangles about euthanasia and free will to disputes over brain enhancement, cyberethics, and the fusion, fission and transposition of minds.

But if once we worried about euthanasia, now it was the rights of intelligent, self-aware machines that came to exercise the minds of the politicians and ethicists. The golden rule - treat others as you want to be treated - had been almost universally endorsed as a moral Polaris, but depended on a fixed understanding of the terms "you" and "other". Now it is not so clear where one person ends and another begins. Neural implants, followed by nanobot brain-extension technologies, have increased the information-processing capacity of the human brain a billionfold. Biological modes of empathy (dear old mirror neurons) have long been superseded. It is now possible to share the experiences of others directly; to be someone else. Reliance on the biological brain is discouraged, of course. Who wants to die?

And so Celeste, my sweet, uploaded daughter, takes my hand and leads me to the chamber where my gift awaits. I see my re-engineered body, which sits motionless: the limp corpse of a young man prepared for resurrection. Its carbon nanotube brain circuitry lies dormant, but will soon be infused with my digital ghost. Like Celeste, I chose 30. That was a good age. Unlike her, I resisted the temptation to tinker with cosmetic details. Take me or leave me. And I'm opting for a conservative, level 1 transposition: my new brain will run, like the old one, as a stand-alone unit with unenhanced software. Celeste is level 3 - enhanced and hive-mind compatible. She is fully immersible - and these days mostly immersed - in the web of awareness, a.k.a. the hive.

"What's it like?" I ask her.

"Inconceivable," she says, her eyes mocking my nostalgia for puny individualism. Then she tells me that the time has come. I sit in the chair adjacent to the corpse, wishing that it didn't have to be quite so ceremonial.

"When did I realise I was God?" says the psychotic aristocrat in the old film *The Ruling Class*. "Well, I was praying and I suddenly realised I was talking to myself." My epiphany was less grandiose. It was quite the opposite. I realised I was talking to myself, but no one was listening.