THE HIDDEN SPRING

A Journey to the Source of Consciousness

MARK SOLMS

PROFILE BOOKS

First published in Great Britain in 2021 by
Profile Books Ltd
29 Cloth Fair
London
EC1A 7JQ

www.profilebooks.com

1 3 5 7 9 10 8 6 4 2

Printed and bound in Great Britain by Clays Ltd, Elcograf S.p.A.

A CIP catalogue record for this book is available from the British Library.

ISBN 978 1 78816 283 8
eISBN 978 1 78283 571 4

Contents

Introduction

When I was a child, a peculiar question occurred to me: how do we picture the world as it existed before consciousness evolved? There was such a world, of course, but how do you *picture* it – the world as it was before picturing things became possible?

To give you a sense of what I mean, try to imagine a world in which *a sunrise cannot occur*. The earth has always revolved around the sun, but the sun only rises over the horizon from the viewpoint of an observer. It is an inherently perspectival event. The sunrise will forever be trapped in experience.

This obligatory perspective-taking is what makes it so difficult for us to comprehend consciousness. If we want to do so, we need to elude subjectivity – to look at it from the outside, to see things as they really are as opposed to how they appear to us. But how do we do that? How do we escape our very selves?

As a young man, I naïvely visualised my consciousness as a bubble surrounding me: its contents were the moving pictures and sounds and other phenomena of experience. Beyond the bubble, I assumed there lay an infinite blackness. I imagined this blackness as a symphony of pure quantities, interacting forces and energies and the like: the true reality 'out there' that my consciousness represents in the qualitative forms that it must.

The impossibility of any such imagining – the impossibility of representing reality without representations – illustrates the scale of the task that is tackled in this book. Once again, all these years later, I am trying to peek behind the veil of consciousness, to catch a glimpse of its actual mechanism.

The book you hold in your hands, then, is unavoidably

perspectival. In fact, it is even more perspectival than the paradox I have just described requires it to be. To help you see things from my point of view, I decided to tell a part of my own history. Advances in my scientific ideas about consciousness have often emerged from developments in my personal life and clinical work, and though I believe that my conclusions stand alone, it is much easier to grasp them if you know how I came to them. Some of my discoveries – for example, the brain mechanisms of dreaming – happened largely by serendipity. Some of my professional choices – for example, to take a detour from my neuroscientific career and train as a psychoanalyst – paid off more handsomely than I could reasonably have hoped. In both cases, I will explain how.

But to the extent that my quest to understand consciousness has been successful, my greatest stroke of luck has been the brilliance of my collaborators. In particular, I had the profound good fortune to work with the late Jaak Panksepp, a neuroscientist who, more than any other, understood the origin and power of *feelings*. Pretty much everything that I now believe about the brain was shaped by his insights.

More recently, I have been able to work with Karl Friston, who, among his many excellent qualities, bears the distinction of being the world's most influential living neuroscientist. It was Friston who dug the deepest foundations for the theory I am about to elaborate. He is best known for reducing brain functions (of all kinds) to a basic physical necessity to minimise something called *free energy*. That concept is explained in Chapter 7, but for now, let me just say that the theory that Friston and I have worked out joins with that project – so much so that you may as well call it the free energy theory of consciousness. That's what it is.

The ultimate explanation for sentience is a puzzle so difficult it is nowadays referred to reverentially as 'the hard problem'. Sometimes, once a puzzle is solved, both the question and its answer cease to be interesting. I will leave it to you to judge whether the ideas I set out here shed new light on the hard problem. Either

way, I am confident they will help you to see *yourself* in a new light, and to that degree they should remain interesting until such time as they are superseded. After all, in a profound sense, you *are* your consciousness. It therefore seems reasonable to expect a theory of consciousness to explain the fundamentals of why you feel the way you do. It should explain why you are the way you are. Perhaps it should even clarify what you can do about it.

That last topic, admittedly, transcends the intended scope of this book. But it is not beyond the scope of the theory. My account of consciousness unites in a single story the elementary physics of life, the most recent advances in both computational and affective neuroscience and the subtleties of subjective experience that were traditionally explored by psychoanalysis. In other words, the light this theory sheds ought to be light you can use.

It has been my life's work. Decades on, I am still asking myself how the world might have looked before there was anyone around to see it. Now, better educated, I imagine the dawn of life in one of those hydro-thermal vents. The unicellular organisms that came into being there would surely not have been conscious, but their survival prospects would have been affected by their ambient surrounds. It is easy to imagine these simple organisms responding to the biological 'goodness' of the energy of the sun. From there, it is a small step to imagine more complex creatures actively striving for such energy supplies and eventually evolving a capacity to weigh the chances of success by alternative actions.

Consciousness, in my view, arose from the experience of such organisms. Picture the heat of the day and cold of the night from the perspective of those first living beings. The physiological values registering their diurnal experiences were the precursors of the first sunrise.

Many philosophers and scientists still believe that sentience serves no physical purpose. My task in this book is to persuade you of the plausibility of an alternative interpretation. This requires me to convince you that feelings are part of nature, that they are

not fundamentally different from other natural phenomena, and that they *do* something within the causal matrix of things. Consciousness, I will demonstrate, is about feeling, and feeling, in turn, is about how well or badly you are doing in life. Consciousness exists to help you do better.

The hard problem of consciousness is said to be the biggest unsolved puzzle of contemporary neuroscience, if not all science. The solution proposed in this book is a radical departure from conventional approaches. Since the cerebral cortex is the seat of intelligence, almost everybody thinks that it is also the seat of consciousness. I disagree; consciousness is far more primitive than that. It arises from a part of the brain that humans share with fishes. This is the 'hidden spring' of the title.

Consciousness should not be confused with intelligence. It is perfectly possible to feel pain without any reflection as to what the pain is about. Likewise, the urge to eat – a feeling of hunger – need not imply any intellectual comprehension of the exigencies of life. Consciousness in its elemental form, namely raw *feeling*, is a surprisingly simple function.

Three other prominent neuroscientists have taken this approach: Jaak Panksepp, Antonio Damasio and Bjorn Merker. Panksepp led the way. He (like Merker) was an animal researcher; Damasio (like me) is not. Many readers will be horrified by the animal research findings I report here, precisely because they show that other animals feel just as we do. All mammals are subject to feelings of pain, fear, panic, sorrow and the like. Ironically, it was Panksepp's research that removed any reasonable doubt on that score. Our only consolation is that his findings made it impossible for such research to continue unabated.

I was drawn to Panksepp, Damasio and Merker because they believed, as I do, that what is lacking in the neuroscience of our time is a clear focus on the embodied nature of *lived experience*. It could be said that what unites us is that we have built, sometimes

unwittingly, upon the abandoned foundations that Freud laid for a science of the mind that prioritises feelings over cognition. (Cognition is mostly unconscious.) This is the second radical departure of this book; it returns us to Freud's 'Project' of 1895 – and it attempts to finish the job. But I do not overlook his many mistakes. For one thing, like everyone else, Freud thought that consciousness was a cortical function.

The third and last major departure of this book is that it comes to the view that consciousness is engineerable. It is artificially *producible*. This conclusion, with its profound metaphysical implications, arises from my work with Karl Friston. Unlike Panksepp, Damasio and Merker, Friston is a computational neuroscientist. Therefore, he believes that consciousness is ultimately reducible to the laws of physics (a belief that, surprisingly, was shared by Freud). But even Friston largely equated mental functions with cortical ones before we began our collaboration. This book takes his statistical-mechanical framework deeper, into the most primitive recesses of the brainstem …

These three departures make the hard problem less hard. This book will explain how.

Mark Solms
Chailey, East Sussex
March 2020

1

The Stuff of Dreams

I was born on the Skeleton Coast of the former German colony of Namibia, where my father administered a small South African-owned company called Consolidated Diamond Mines. The holding company, De Beers, had created a virtual country within a country, known as the *Sperrgebiet* ('prohibited area'). Its sprawling alluvial mines extended from the sand dunes of the Namib Desert down to the Atlantic Ocean floor, several kilometres out to sea.

This was the peculiar landscape that moulded my imagination. As small children, my older brother Lee and I used to play at diamond mining, using toy earth-moving machines, recreating in our garden the impressive engineering feats we witnessed at our father's side when he took us to see the open-cast mines in the desert. (We were, of course, too young to know about the less impressive aspects of his industry.)

One day in 1965, when I was four years old, my parents were yachting at the Cormorant Yacht Club, as they often did, and I was left playing in the clubhouse with Lee, aged six. The early morning mists had burned away. I wandered from the cool interior of the three-storey clubhouse down to the water's edge. Wading there in the heat, I watched tiny shimmering fishes scatter from my feet as Lee and some friends of his clambered onto the roof from the back of the building.

What I remember next are three snapshots. First, the sound of something like a watermelon cracking open. Next, the image

of Lee lying on the ground whimpering about a sore leg. Last, my aunt and uncle telling me that they would be looking after my sister and me while our parents travelled to the hospital with Lee. The bit about a sore leg must be a confabulation: the medical records state that my brother lost consciousness upon impact with the concrete paving.

Lee needed specialist care of a kind that our local hospital could not provide. He was flown by helicopter to Groote Schuur Hospital in Cape Town, 800 km away. The neurosurgery department was then housed in an impressive block built in the Cape Dutch style, the very building in which I now work as a neuropsychologist. Lee's skull had fractured and he had suffered an intracranial haemorrhage. When such haematomas expand, they present a life-threatening emergency requiring surgical intervention. My brother was lucky: his resolved over the next few days and he was eventually discharged home.

Apart from the fact that he had to wear a helmet after the accident to protect his fractured skull, Lee looked no different. As a person, however, he was profoundly altered. There is a German word for the feeling this aroused in me, *Unheimlichkeit*, for which there is no adequate English equivalent. Literally, it means 'unhomeliness' but it translates better as 'eeriness' or 'the uncanny'.

The most obvious way in which he was changed was that he lost his developmental milestones. For a time, he even lost reliable bowel control. What I found more disturbing was the fact that he seemed to *think* differently from before. It felt as if Lee was simultaneously there and not there. He seemed to have forgotten many of the games we played. Now our diamond-mining game became simply digging holes. Its imaginative and symbolic aspects no longer spoke to him. He was no longer Lee.

He failed that year at school – his first. The thing I remember most from those early days after the accident was trying to reconcile the dichotomy that my returned brother looked the same

but was not the same. I wondered where the earlier version of him had gone.

Over the ensuing years, I fell into a depression. I remember not being able to muster the energy to put on my shoes in the morning, to go to school. This was about three years after the accident. I couldn't find the energy to do these things because I couldn't see the point of them. If our very being depended upon the functioning of our brains, then what would become of me when *my* brain died, with the rest of my body? If Lee's mind was somehow reducible to a bodily organ then, surely, mine was too. This meant that I – my sentient being – would exist only for a relatively short period of time. Then I would disappear.

I have spent my whole scientific career thinking about this problem. I wanted to understand what happened to my brother, and what would in time happen to all of us. I needed to understand what, in biological terms, our existence as experiencing subjects amounted to. In short: to understand consciousness. That is why I became a neuroscientist.

Even in retrospect, I don't believe I could have taken a more direct route to the answers I sought.

The nature of consciousness may be the most difficult topic in science. It matters because you *are* your consciousness, but it is controversial because of two puzzles that have bedevilled thinkers for centuries. The first is the question of how the mind relates to the body – or, for those of a materialist bent (which is almost all neuroscientists), how the brain gives rise to the mind. This is called 'the mind/body problem'. How does the physical brain produce your phenomenal experience? Equally confoundingly, how does the non-physical stuff called consciousness control the physical body?

Philosophers have assigned this problem to what they call 'metaphysics', which is a way of saying they don't think it can be resolved scientifically. Why not? Because science depends upon

empirical methods, and 'empirical' implies 'derived from sensory evidence'. The mind is not accessible to sensory observation. It cannot be seen or touched; it is invisible and intangible, a subject, not an object.

The question of what we can know about minds from the outside – how we can even tell when they are present, for that matter – is the second puzzle. It is called 'the problem of other minds'. Simply put: if minds are subjective, then you can only observe your own. How, then, can we know whether other people (or creatures, or machines) have one at all, let alone discern any objective laws governing how minds in general work?

Over the past century, these questions have elicited three major scientific responses. Science relies upon experiments. One thing in our favour is that the experimental method does not aspire to ultimate truths, but rather to what may be described as best guesses. Starting from observations, we offer conjectures as to what might plausibly explain the observed phenomena. In other words, we formulate hypotheses. Then we generate *predictions* from our hypotheses. These take the form: 'if hypothesis *X* is correct, then *Y* should happen when I do *Z*' (where there is a reasonable chance that *Y* will not happen under some other hypothesis). This is the experiment. If *Y* does not happen, then *X* is inferred to be false and is revised in accordance with the new observations. Then the experimental process begins again, until it gives rise to falsifiable predictions that are confirmed. At that point, we hold the hypothesis to be *provisionally* true, until and unless further observations contradict it. In this way, we do not expect to attain certainty in science; we aspire only to less uncertainty.[1]

Starting in the first half of the twentieth century, a school of psychology called 'behaviourism' began systematically to apply the experimental method to the mind. Its starting point was to disregard everything except empirically observable events. The behaviourists threw out all 'mentalistic' talk of beliefs and ideas, feelings and desires, and restricted their field of study to the

subject's visible and tangible responses to objective stimuli. They were fanatically uninterested in subjective reports about what was going on inside. They treated the mind as a 'black box', whose inputs and outputs were all that could be known of it.

Why did they take such an extreme stance? Partly, of course, it was an attempt to navigate around the problem of other minds. If they refused to countenance any talk of minds in the first place, it stood to reason that their theories could not be afflicted by the philosophical doubts endemic to psychology. In effect, they excluded the psyche from psychology.

That may seem like a high price to pay. But behaviourism was from the outset a revolutionary doctrine. The behaviourists weren't chasing epistemological purity for its own sake: they were also trying to dethrone the incumbent power in psychology at the time. Freudian psychoanalysis had dominated the science of the mind since the start of the century. By closely examining the curious features of introspective testimonies, Sigmund Freud had sought to develop a model of the mind considered, as it were, from the inside out. The resulting ideas set the agenda for treatment and research for half a century, spawning institutions, accredited experts and a cadre of prominent intellectual champions. Yet in the judgement of the behaviourists, all Freud's theories were just so many cloud castles, erected on the vaporous foundations of subjectivity. Freud had run headlong into the problem of other minds and dragged the rest of psychology after him. It was up to the behaviourists to pull it back again.

Despite the austerity of their programme, they were in fact able to infer causal relations between certain types of mental stimuli and responses. Not only that: they could also manipulate the inputs to elicit predictable changes in the outputs. In doing so, they discovered some of the fundamental laws of learning. For example, when the trigger of an involuntary behaviour is paired repeatedly with an artificial stimulus, then the artificial stimulus will come to trigger the same involuntary response as the innate

stimulus. So, if the sight of food is paired repeatedly with the ringing of a bell (in animals that naturally salivate when they see food, as dogs do), then the sound of the bell alone will come to trigger salivation. This is called 'classical conditioning'. Likewise, if a voluntary behaviour is accompanied repeatedly by rewards, that behaviour will increase, and if the same behaviour is accompanied by punishments, it will decrease. So, if a dog that jumps on visitors is hugged, it will jump on them more; if it is smacked, it will jump on them less. This is called 'operant conditioning' – also known as the Law of Effect.

Such discoveries were no small achievement; they showed that the mind is subject to natural laws, like everything else. But there is a lot more to the mind than learning, and even learning is influenced by factors other than external stimuli. Imagine thinking to yourself: 'after I have read this page, I will make myself a cup of tea'. This type of thinking influences your behaviour all the time. Yet the behaviourists did not consider such introspective reports to be acceptable scientific data, because thoughts are not externally observable. In consequence, they could not know what caused you to make your cup of tea.

The great neurologist Jean-Martin Charcot once said: 'theory is good, but it doesn't prevent things from existing'.[2] Since internal mental events clearly do exist and causally influence behaviour, the behaviourist approach was gradually eclipsed in the second half of the twentieth century by another approach. It was called 'cognitive' psychology, which was able to accommodate internal mental processes – in a manner of speaking.

The impetus behind the cognitive revolution was the advent of computers. Behaviourists considered the internal workings of the mind to be an inscrutable 'black box' and focused instead on its inputs and outputs. But computers are not unfathomable. It would have been impossible for us to invent them without thoroughly understanding their inner workings. By treating the mind as though it were a computer, therefore, psychologists felt

emboldened to formulate models of the *information processing* that went on within it. Their models were then tested using artificial simulations of mental processes, combined with behavioural experiments.

What is information processing? I will say a lot about it later, but the most interesting thing for our present purposes is that it can be implemented with vastly different kinds of physical equipment. This casts new light on the physical nature of the mind. It suggests that the mind (construed as information processing) is a *function* rather than a structure. On this view, the 'software' functions of the mind are implemented by the 'hardware' structures of the brain, but the same functions can be implemented equally well by other substrates, such as computers. Thus, both brains and computers perform *memory* functions (they encode and store information) and *perceptual* functions (they classify patterns of incoming information by comparing them with stored information) as well as *executive* functions (they execute decisions about what to do in response to such information).

This is the power of what came to be called the 'functionalist' approach, but it is also its weakness. If the same functions can be performed by computers, which presumably are not sentient beings, then are we really justified in reducing the mind to mere information processing? Even your phone has memory, perceptual and executive functions.

The third major scientific response to mind/body metaphysics developed in tandem with cognitive psychology, but by the end of the last century it had grown to overshadow it. I am referring to an approach that is broadly termed 'cognitive neuroscience'. It focuses on the hardware of the mind, and it arose with the development of a plethora of physiological techniques that make it possible for us to observe and measure the dynamics of the living brain directly.

In behaviourist times, neurophysiologists were limited to a single such technique: they could record the brain's

electrical activity from the outer surface of the scalp using an electroencephalogram (EEG). Nowadays we have many more tools at our disposal, such as functional magnetic resonance imagery (fMRI) to measure the rates of haemodynamic activity in different parts of the brain while it is performing specific mental tasks, and positron emission tomography (PET), with which we can measure differential metabolic activity for single neurotransmitter systems. This enables us to identify precisely which brain processes generate our different mental states. We can also visualise the detailed functional-anatomical connectivity between those different brain regions using diffusion tensor tractography. And by using optogenetics we can see and activate the circuits of neurons comprising individual memory traces as they light up during cognitive tasks.

These techniques render the inner workings of the organ of the mind plainly visible – thereby realising the wildest empiricist dreams of the behaviourists without limiting the scope of psychology to stimuli and responses.

The state of neuropsychology in the 1980s when I entered the field explains why behaviourists made such a seamless transition from learning theory to cognitive neuroscience. The neuropsychology of that time might as well have been called neurobehaviourism. The more I was taught about functions like short-term memory, which was said to provide a 'buffer' for holding memories in consciousness, the more I realised that my lecturers were talking about something other than what I had signed up for. They were teaching us about the functional tools used by the mind, rather than the mind itself. I was dismayed.

The neurologist Oliver Sacks, in his book *A Leg to Stand On* (1984), aptly described the situation I found myself in:

> Neuropsychology, like classical neurology, aims to be entirely objective, and its great power, its advances, come from just this. But a living creature, and especially a human being, is

> first and last active – a subject, not an object. It is precisely the subject, the living 'I', which is being excluded. Neuropsychology is admirable, but it excludes the psyche – it excludes the experiencing, active, living 'I'.[3]

That line 'Neuropsychology is admirable, but it excludes the psyche' captured my disappointment perfectly. Upon reading it, I entered into a correspondence with Oliver Sacks that continued until his death in 2015. What drew me to him was the fact that he took so seriously the subjective reports of his patients. This was evident already in his 1970 book *Migraine*, and even more so in his extraordinary *Awakenings* (1973). The second book recorded in exquisite detail the clinical journeys of a group of chronic 'akinetic-mute' patients with encephalitis lethargica. This disease was also known as 'sleeping sickness', although the patients were not literally asleep, rather they showed no spontaneous initiative or drive. Sacks 'awakened' them by giving them levodopa, a drug that increases the availability of dopamine. Following the return of active agency, however, they rapidly became excessively driven, manic and eventually psychotic. Shortly after I read *A Leg to Stand On*, which described Sacks's own subjective experience of a nervous-system injury, he published *The Man Who Mistook His Wife for a Hat* (1985) – a series of case studies that provided enlightening insights into neuropsychological disorders from the perspective of *being* a neurological patient. This brought Sacks lasting fame.

These books were quite unlike my neuropsychological textbooks, which dissected mental functions as we would the functions of any bodily organ. For example, I learnt that language was produced by Broca's area in the left frontal lobe, that speech comprehension took place in Wernicke's area, a few centimetres further back, in the temporal lobe, and that the ability to repeat what is said to you was mediated by the arcuate fasciculus, a fibre tract that connects these two regions. Likewise, I learnt