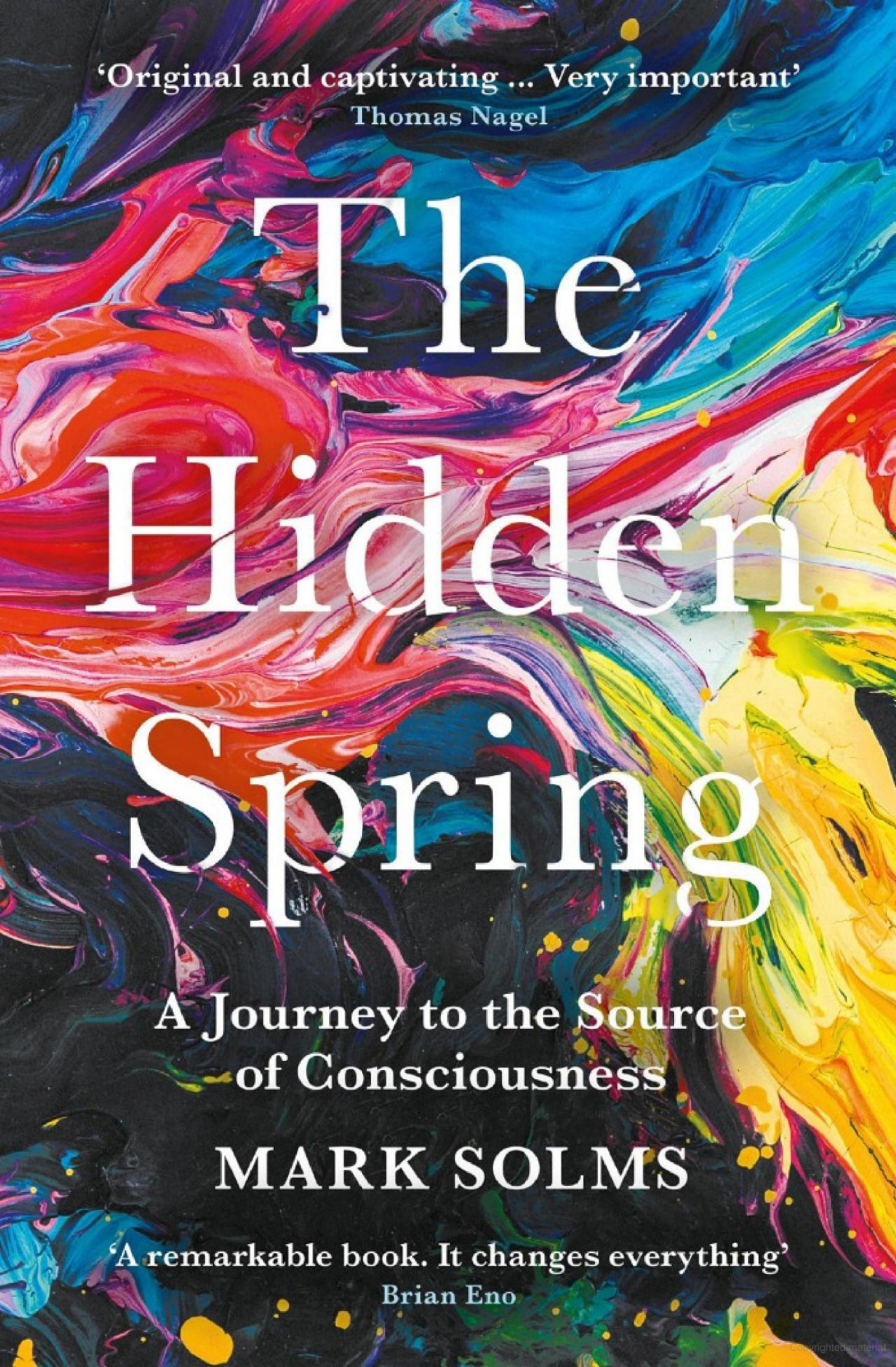

This paperback edition first published in 2022

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

Typeset in Sabon by MacGuru Ltd

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

ISBN 978 1 78816 284 5
eISBN 978 1 78283 571 4
Export ISBN 978 1 78816 762 8

Contents

List of Figures

While every effort has been made to contact copyright-holders of illustrations, the author and publishers would be grateful for information about any illustrations where they have been unable to trace them, and would be glad to make amendments in further editions.

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 imposing 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 that memories were encoded by the hippocampus, stored in the neocortex and retrieved by frontal-limbic mechanisms.

Was the brain really no different from the stomach and lungs? The obvious thing that set it apart was the fact that there is 'something it is like' to *be* a brain. This did

not apply to any other part of the body. The sensations that we locate in other bodily organs are not felt by the organs themselves; nerve impulses arising from them are felt only when they reach the brain. Surely this highly distinctive property of brain tissue – the capacity to sense, feel and think things – existed for a reason. This property appeared to do something. And if it did – if subjective experience had causal effects upon behaviour, as it seems to when we spontaneously decide to make a cup of tea – then we would be led badly astray if we omitted it from our scientific accounts. Yet that is precisely what was happening in the 1980s. At no point did my lecturers say anything about what it is like to comprehend speech or retrieve a memory, let alone why it feels like anything at all.

Those who did take the subjective perspective into account were not taken seriously by proper neuroscientists. I am not sure how many people know that Sacks's publications were widely derided by his colleagues. One commentator went so far as to call him 'The man who mistook his patients for a literary career'. This caused him a good deal of distress. How can you describe the inner life of human beings without telling their stories? As Freud had lamented a century before in relation to his own clinical reports:

> It still strikes me as strange that the case histories I write should read like short stories and that, as one might say, they lack the serious stamp of science. I must console myself with the reflection that the nature of the subject is evidently responsible for this, rather than any preference of my own.[4]

Sacks was delighted when I sent him this quotation.[5] For my own part, when I first read these lines, I realised that I was not alone in having entered neuropsychology with the hope that it would enable me to learn how the brain generates subjectivity. One is quickly disabused of this notion. You are warned not to pursue such intractable questions – they are 'bad for your career'. And so, most students of neuroscience gradually forget why they entered the field, and come to identify with the dogma of cognitivism, which approaches the brain as though it were no different from a mobile phone.

The one aspect of consciousness that *was* a respectable scientific topic in the 1980s was the brain mechanism of wakefulness versus sleep. In other words, the 'level' of consciousness was a respectable topic but not its 'contents'. So, I decided to focus my doctoral research on an aspect of sleep. In particular, I chose to study the subjective aspect of sleep, namely the brain mechanisms of dreaming. Dreaming, after all, is nothing but a paradoxical intrusion of consciousness ('wakefulness') into sleep. Amazingly, there was a huge gap in the literature on this topic: nobody had systematically described how damage to different parts of the brain affected dreaming. So, this is what I set out to do.

What makes dreaming tricky to study is precisely its subjective nature. Mental phenomena in general can be witnessed only introspectively by a single observer

and then reported to others indirectly, through words. But dreams are even more problematic: they can be reported only retrospectively, once the dream is over and the dreamer has woken up. Everyone knows how unreliable our memory for dreams is. What kind of 'data' are those?[6] Which is why, from the middle of the twentieth century onwards, dreams were a significant front in the transition from behaviourism to what would later become cognitive neuroscience.

The electroencephalogram was first applied to the study of sleep in the early 1950s by two neurophysiologists, Eugene Aserinsky and Nathaniel Kleitman. They hypothesised that the level of brain activity would decrease as we fall asleep and increase when we wake up, and therefore predicted that the amplitude of our brainwaves (which is one of the things that electroencephalography measures) would increase and their frequency (the other thing it measures) would decrease as we fall asleep; and that the opposite would happen when we wake up (see Figure 10 on p. 127).

When the brain descends into what is now called 'slow wave' sleep, we see exactly what Aserinsky and Kleitman predicted. Their hypothesis was confirmed. The surprise is what happens next: within about ninety minutes of drifting off (and roughly every ninety minutes thereafter, in regular cycles) the brainwaves speed up again, almost reaching waking levels, even though the person from whom the recordings are being obtained remains asleep.[7] Aserinsky and Kleitman named these curious states of brain activation 'paradoxical sleep' – the paradox being that the brain is physiologically aroused despite being fast asleep.

Various other things happen in this peculiar state. The eyes move rapidly (which is why paradoxical sleep was later renamed 'rapid eye movement' or REM sleep), yet the body below the neck is temporarily paralysed. There are dramatic autonomic changes, too, such as reduced control of core body temperature and engorgement of the genitals leading to visible erections in men. How science managed not to notice all this until 1953 is mind-boggling.

On the basis of these observations, Aserinsky and Kleitman formulated a further, not-unreasonable hypothesis: that REM sleep is the physiological basis of the psychological state called dreaming. Accordingly, they predicted that awakenings from REM sleep would elicit dream reports while awakenings from slow-wave (non-REM) sleep would not. Together with the unfortunately named William Dement, they tested this prediction and confirmed it: whereas approximately 80 per cent of awakenings from REM sleep produced dream reports, fewer than 10 per cent of awakenings from non-REM sleep did so. From that moment onward, REM sleep was considered to be synonymous with dreaming.[8] Excellent news! The field no longer had to bother with dreaming, because now we had an objective marker of it, which enabled neuroscientists to do proper science without having to contend with the methodological complications introduced by retrospective, single-witness, verbal reports of fleeting subjective experiences.

There was another reason to be grateful for getting rid of dreams. This was the embarrassing role that they had played in the establishment of psychoanalysis. Unlike the mainstream responses to mind/body metaphysics that characterised mental science in the second half of the twentieth century, psychoanalysts had no qualms about treating introspective reports as data. In fact, reports elicited by 'free association' (unstructured sampling of the stream of consciousness) were the primary data of psychoanalytic research. Using this method, Sigmund Freud came to the conclusion that, despite the nonsensical appearance of 'manifest' dream experiences, their 'latent' content (the underlying story, which he inferred from the dreamer's free associations) revealed a coherent psychological function. This function was *wish-fulfilment*.

According to Freud, dreaming is what happens when the biological needs that generate waking behaviour are released from inhibition during sleep. Dreams are attempts to meet those needs, which continue to make demands upon us even when we sleep. However, dreams do so in a hallucinatory fashion, and thereby enable us to stay asleep (rather than wake up in order to really satisfy our drives). Since hallucinations are a core feature of mental illness, Freud in his seminal book *The Interpretation of Dreams* (1900) used this theory to paint a broad-brushstroke model of how the mind as a whole works, in health and disease.

As Freud put it: 'psychoanalysis is founded upon the analysis of dreams'.[9] But dreams, as we have seen, are incredibly difficult things to study empirically, and so the behaviourists ruled them out of science. What was more, the theoretical edifice that Freud built upon dreams was no better than its foundations. The great philosopher of science Karl Popper declared psychoanalytic theory 'pseudoscientific', because it did not give rise to experimentally falsifiable predictions.[10] How do you falsify the hypothesis that dreams express the latent desires that Freud inferred? If the desires do not have to appear in the manifest (reported) dream, then any dream can be 'interpreted' to suit the requirements of the theory. Not surprisingly, therefore, when the discovery of REM sleep made it possible for neuroscientists to shift from the ephemeral stuff of dream reports to their concrete physiological correlates, the dreams themselves were dropped like slippery fish.

The discovery of REM sleep in the 1950s triggered a race to identify its neurological basis, since the function of REM sleep could reveal the *objective* mechanism of dreams, whose elucidation would place the psychiatry of the time on a more respectable scientific footing. (This research was made easier by virtue of the fact that REM sleep occurs in all mammals.) The race was won by Michel Jouvet, in 1965. In a series of surgical experiments on cats, he demonstrated that REM sleep was generated not by the forebrain (which includes the cortex, the upper part of the brain that is so impressively large in humans and partly for that reason is considered the organ of the mind) but rather by the brainstem, a supposedly much

humbler structure of exceedingly ancient evolutionary origin.[11] Jouvet came to this conclusion by observing that progressive slices through the brain, starting at the top and working downwards, only produced loss of REM sleep once the cutting had reached the level of a 'lowly' brainstem structure known as the pons (see Figure 1).[12]

It fell to Jouvet's student Allan Hobson to wrap up the details. Hobson identified precisely which assemblies of pontine neurons generated REM sleep and therefore dreams. It became apparent by the mid-1970s that the whole sleep/waking cycle – including all the phenomena of REM sleep enumerated above, as well as those of the different stages of non-REM sleep – were orchestrated by a small number of brainstem nuclei interacting with each other.[13] Those controlling REM sleep resembled a simple on/off switch. The neurons that switch REM *on* are found in the mesopontine tegmentum (see Figure 1). They release a neurochemical called acetylcholine throughout the forebrain. Acetylcholine causes arousal: it increases the 'level' of consciousness (for example, it is boosted by nicotine, which thereby helps you concentrate). The brainstem neurons that switch REM sleep *off* are located deeper within the pons, in the dorsal raphe and locus coeruleus complex (again, see Figure 1). They release serotonin and noradrenaline respectively. Like acetylcholine, these neurochemicals modulate different aspects of the level of consciousness.

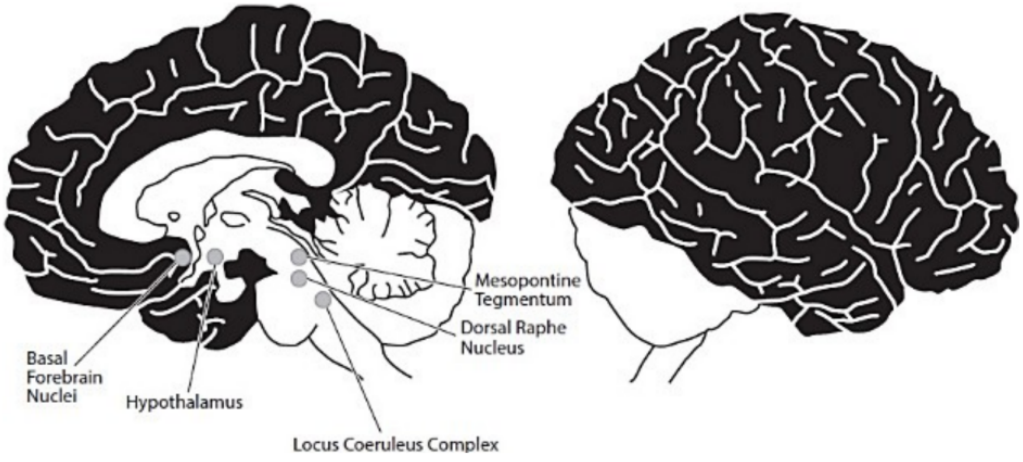

Figure 1 The image on the left is a medial view of the brain (cut through the middle) and the one on the right is a lateral view (seen from the side). The figure shows the cortex (black) and brainstem (white). Only those brainstem nuclei considered important for REM sleep control are indicated, namely the mesopontine tegmentum, dorsal raphe nucleus and locus coeruleus complex. Also shown are the location of the basal forebrain nuclei (underneath the cortex) and the hypothalamus, the relevance of which becomes apparent later.

Combining these findings with the fact that REM sleep switches on and off automatically, roughly every ninety minutes, like clockwork, Hobson wasted no time in drawing the inevitable conclusion: 'The primary motivating force for dreaming is not psychological but physiological since the time of occurrence and duration of dreaming sleep are quite constant, suggesting a preprogrammed, neurally determined genesis.'[14]

Because REM sleep arises from the cholinergic brainstem, an ancient and lowly part of the brain far from the majestic cortex where all the action of human psychology presumably takes place, he added that dreaming could not possibly be motivated by wishes; it was 'motivationally neutral'.[15] Therefore, according to Hobson, Freud's view that dreams were driven by latent desires must be completely wrong. The meaning that Freud saw in dreams was no more intrinsic to them than it is to inkblots. It was projected onto them; it was not in the dream itself. From the scientific point of view, dream interpretation was no better than reading tea leaves.

Because the whole of psychoanalysis was grounded on the method that Freud used to study dreams, the entire body of theory that he derived in this way could be dismissed. Following Hobson's demolition job on the idea that dreams might mean anything at all, psychiatry could at last turn away from its historical reliance on introspective reports and base itself instead upon objective neuroscientific (especially neurochemical) methods of research and treatment. In consequence, whereas in the 1950s it was almost impossible to become a tenured professor of psychiatry at a leading American university unless you were a psychoanalyst, today

the opposite is true: it is almost impossible to become a professor of psychiatry if you *are* a psychoanalyst.

None of this particularly struck me at the time. The question at the heart of my doctoral research seemed fairly straightforward, and not at all implicated in the battles over the legacies of Freudianism and behaviourism. All I wanted to know was this: how did damage to different parts of the forebrain and its cortex affect the actual experience of dreaming? After all, if the forebrain was where the action was, psychologically speaking, surely it must do *something* in dreams.

The department of neurosurgery at the University of the Witwatersrand had wards in two teaching hospitals – Baragwanath Hospital and Johannesburg General Hospital. Baragwanath was a sprawling ex-military hospital, set in the 'non-European' township of Soweto. Bearing in mind that this was during the height of apartheid in South Africa, it was a sea of human misery. The Johannesburg General Hospital, by contrast, which was reserved for 'Europeans', was a state-of-the-art academic hospital; a monument to racial inequality. The neurosurgery department also had beds in the Brain and Spine Rehabilitation Unit at Eden-vale General Hospital, which was in an old colonial building set in Johannesburg's suburbia. Starting in 1985 I worked across all three sites, examining hundreds of patients per year. I included 361 of them in my doctoral research, which extended over the next five years.

After learning how to use electroencephalographic and related technology and to recognise the characteristic brainwaves associated with the different stages of sleep, I was able to wake people up during REM, when they were most likely to be dreaming. I also asked neurological patients at the bedside about changes in their dreams, and then followed them up over days, weeks and months. This is how I proceeded to investigate whether the content of dreams was systematically affected by localised damage to different parts of the brain. Despite the dubious reputation of dream reports, I assumed that if patients with damage to the same brain area claimed the same change in dream content, there was every reason to believe them. This method is called 'clinico-anatomical correlation': by probing the psychological capacities of patients clinically, you observe how a mental function has been altered by damage to a part of the brain; then you correlate that alteration with the site of the damage, in this way discovering clues about the function of the damaged brain structure, which leads to testable hypotheses. The method had been systematically applied decades before to all the major cognitive functions, such as perception, memory and language, but it had not yet been applied to dreaming.

At first, I was a little uneasy about talking to such seriously ill people about their dreams. Many of them were facing, or had just undergone, life-threatening brain surgery, and in the circumstances I feared they might consider my questions frivolous. But my patients were surprisingly willing to describe the changes in their mental life that neurological diseases had brought about.

By the time I began my research, several case reports had been published in which the same effect observed in experimental animals was shown to occur in human beings: namely that REM sleep was obliterated by damage to the mesopontine tegmentum (see Figure 1). But, astonishingly, nobody had bothered to enquire about changes in these patients' *dreams*. This is as clear an example as one can get of the prejudice against subjective data in neuroscience.[16]

In my research, I expected to find the obvious: that patients with damage to the visual cortex would experience non-visual dreams; that patients with damage to language cortex would experience non-verbal dreams; that patients with damage to somatosensory and motor cortex would experience hemiplegic dreams; and so on. These are the ABCs of brain/behaviour correlation. This was the gap I wanted to fill; and, happily, I did.[17]

To my amazement, however, alongside all the obvious things I observed, I found also that patients with damage to the part of the brain that generates REM sleep *still experienced dreams*. Moreover, patients in whom dreaming *was* abolished had damage to a completely different part of the brain. Dreaming and REM sleep were therefore what we call 'doubly dissociable' phenomena.[18] They were correlated with each other (i.e. they usually happened at the same time) but they were not the same thing.[19]

For a period of almost fifty years, in the whole field of sleep science, brain researchers had been confusing correlation with identity. As soon as they had established that dreaming accompanied REM sleep, they leapt to the conclusion that they were one and the same – then jettisoned the troublesome subjective side of the correlation. Thereafter, with very few exceptions, they studied REM sleep alone, mainly in experimental animals, which cannot provide introspective reports. The error came to light only when I began to take neuroscientific interest in the *experience* of dreams in neurological patients.

When, in the early 1990s, I first reported that dreaming was obliterated by damage in a different part of the brain from the part that generates REM sleep, I took pains to stress that the critical area was not in the brainstem.[20] This was because I wanted to emphasise the mental nature of dreaming, and we all knew that mental functions reside in the cortex.

In fact, I found two areas of damage that caused loss of dreaming with preservation of REM sleep. The first was in the cortex, in the inferior parietal lobule (see Figure 2). That finding was not surprising, as the parietal lobe is important for short-term memory. If a patient cannot hold the contents of their memory in the buffer of consciousness, how can they experience a dream? Far more interesting was the second brain area, namely the white matter of the ventromesial quadrant of the frontal lobes, which connects the frontal cortex to various subcortical structures. This finding was totally unexpected; nothing about the functions of this part of the brain is obviously connected with the manifest experience of dreaming, and yet it

namely 'major tranquillisers'. What these drugs did, and modern 'antipsychotics' still do, was block the neurochemical dopamine at the terminals of a brain circuit known as the mesocortical-mesolimbic dopamine system (see Figure 2). Since this circuit is cut through by prefrontal leucotomy, as it was in my nine patients with naturally occurring damage, I hypothesised that this might be the system that generates dreams.

Further experiments confirmed my hypothesis. It had already been established that pharmacological stimulation of this circuit increased the frequency, length and intensity of dreams, without commensurate effects on REM sleep.[23] The drug in question was levodopa, the very same drug that Oliver Sacks had used to 'awaken' his post-encephalitic patients. Neurologists using dopamine stimulants for the treatment of Parkinson's disease have long known that they must be careful not to push their patients into psychosis, like Sacks did; and the onset of unusually vivid dreams is often the first sign of this side effect.[24] The crucial subsequent observations were that the neurons that constitute this circuit (the cell bodies of which are located in the ventral tegmental area) fire at maximum rates during dreaming sleep,[25] and at the same time deliver dopamine in maximum quantities to their targets in the nucleus accumbens (see Figure 2).[26] It is therefore now widely accepted that dreaming can occur independently of REM sleep and that the mesocortical-mesolimbic dopamine circuit is indeed the major driver of dreaming.[27]

Damage to cholinergic pathways in the ventromesial quadrant of the frontal lobes (which arise from the basal forebrain nuclei, see Figure 1) produces the opposite effect to what happens when dopaminergic pathways are damaged, namely *more* dreaming rather than less. Hobson had claimed that acetylcholine was the motivationally neutral generator of dreams, but the same thing occurs if you block acetylcholine pharmacologically as happens when its pathways are damaged. Anticholinergic drugs – acetylcholine blockers – are now widely known to cause *excessive* dreaming.[28] In other words, blocking of the neural system that Hobson claimed was responsible for dreaming has the opposite effect to what his theory predicted.

It rapidly became clear that neuroscience owed Freud an apology. If there is one part of the brain that might be considered responsible for 'wishes', it is the mesocortical-mesolimbic dopamine circuit. It is anything but motivationally neutral. Edmund Rolls (and many others) calls this circuit the brain's 'reward' system.[29] Kent Berridge calls it the 'wanting' system. Jaak Panksepp calls it the SEEKING system – and foregrounds its role in the function of *foraging*.[30] This is the brain circuit responsible for 'the most energised exploratory and search behaviours an animal is capable of exhibiting'.[31] It is also the circuit that drives dreaming.[32]

Hobson was not amused. He invited me to present my findings to his research group in the department of neurophysiology at Harvard. Initially he accepted them, and published a favourable review of the book I wrote on the topic in 1997, noting

that my clinico-anatomical findings were confirmed down to the last detail by Allen Braun's neuroimaging studies (see Figure 3, p. 37).[33] Then he realised that these developments might vindicate a broadly Freudian outlook on dreams, at which point he wrote to me saying that he was willing to endorse my findings publicly only on the condition that I did not claim they supported Freud. So much for the supposed objectivity of neuropsychology.

Yet there was one other very surprising aspect to my discovery. When I first stumbled upon it, I did not pay much attention to the fact that the neurons which drive this circuit are located in the *brainstem* (like those of the circuits that generate REM sleep). As I said, I wanted to emphasise the mental nature of dreaming. My oversight had to be politely pointed out to me by Allen Braun, the neuroimager just mentioned. In the context of the scientific disagreement between myself and Hobson as to which brain circuits drive the dream process (dopaminergic or cholinergic), Braun wrote:

> The curious thing is that, after making a case that forebrain structures must play a critical role in the dream system, Solms ends up by suggesting that it is the dopaminergic afferents to these regions that [generate dreams] – *thereby placing the dream instigator back in the brainstem*.[34]

Braun concluded: 'It sounds to me like these gentlemen are approaching common ground.'[35] In the 1990s, in common with the rest of neuropsychology, I thought the cortex was where all the psychological action was, so I focused on the fact that the white matter tracts that interested me were in the frontal lobes, which is where the damage in my nine cases was located. But all the core nuclei of the brainstem send long axons upwards into the fore-brain (see Figure 2). The cell bodies of these neurons are located in the brainstem, although their output fibres (the axons) terminate in the cortex. This underpins the main *arousal* function of these brainstem nuclei, known collectively as the reticular activating system. It was these activating pathways that were damaged in my nine patients, and in the hundreds of documented non-dreaming leucotomy patients who preceded them.

From 1999 onwards, partly prompted by Braun's comments about the implications of my discovery, I directed my attention to the other arousal systems of the brainstem. The most interesting work in this area was being done by Jaak Panksepp, whose encyclopaedic book *Affective Neuroscience* (1998) laid out in exquisite detail a vast array of evidence for his view that these supposedly mindless systems, responsible for regulating only the 'level' of consciousness, generated a 'content' of their own.

This would turn out to be highly significant.

2

Before and After Freud

In 1987 I made another decision that put me at odds with the rest of my field. I decided to train as a psychoanalyst.[1] My emerging dream-research findings had convinced me that subjective reports had a vital role to play in neuropsychology, and that the field's opposition to Freud had led it into error in more ways than one. But my research findings weren't the deciding factor.

What made up my mind was a seminar that I attended at the University of the Witwatersrand, in the mid-1980s, led by a professor of comparative literature named Jean-Pierre de la Porte. The seminar concerned *The Interpretation of Dreams*, which I was curious about in light of my doctoral research. Like everybody else in those days, I was sceptical about Freud. I had learnt since my undergraduate years that psychoanalysis was 'pseudoscience'. Nobody in the hard sciences took Freud seriously any more, which is presumably why the seminar took place in a humanities department. The reason I attended was Freud had been willing to talk about the *content* of dreams, the topic of my research.

De la Porte explained that one could not understand the theoretical conclusions Freud reached without first digesting an earlier manuscript of his, written in 1895 but published only in the 1950s, after his death. This manuscript was titled 'Project for a scientific psychology'.[2] In it, Freud attempted to place his early insights about the mind on a neuroscientific footing.

In doing so, he was following in the footsteps of his great teacher, the physiologist Ernst von Brücke, a founding member of the Berlin Physical Society. The mission of this society was formulated as follows by Emil du Bois-Reymond in 1842:

> Brücke and I pledged a solemn oath to put into effect this truth: 'No other forces than the common physical and chemical ones are active within the organism. In those cases which cannot currently be explained by these forces one has either to find the specific way or form of their action by means of the physical-mathematical method or to assume new forces equal in dignity to the chemical-physical forces inherent in matter, reducible to the forces of attraction and repulsion.'[3]

Their beloved teacher, Johannes Müller, had asked how and why organic life differs from inorganic matter. He concluded that 'living organisms are fundamentally different from non-living entities because they contain some non-physical element or are governed by different principles than are inanimate

things'.[4] In short, according to Müller, living organisms possess a 'vital energy' or 'life force', which physiological laws cannot explain. He held the view that living creatures cannot be reduced to their component physiological mechanisms because they are indivisible wholes with *aims* and *purposes*, which Müller attributed to the fact that they possess a soul. Considering that the German word *Seele* can be translated as either 'soul' or 'mind',[5] the disagreement between Müller and his students bears a striking resemblance to the debate raging in our own time between philosophers like Thomas Nagel and Daniel Dennett as to whether *consciousness* can be reduced to physical laws (Nagel claims it cannot, Dennett claims it can).

The surprise for me, upon attending De la Porte's seminar, was to learn that Freud – the pioneering investigator of human subjectivity – had aligned himself not with the vitalism of Müller but rather with the physicalism of Brücke. As he wrote in the opening lines of his 1895 'Project': 'The intention is to furnish a psychology that shall be a natural science: that is, to represent psychical processes as quantitatively determinate states of specifiable material particles.'[6]

I hadn't realised that Freud was a neuroscientist. Now I learnt that he had only reluctantly abandoned neurological methods of enquiry when it became clear to him, somewhere between 1895 and 1900, that the methods then available were not up to the task of revealing the physiological basis of mind.

Freud's change of heart brought ample compensation, though. It forced him to look more closely at psychological phenomena in their own right, and to elucidate the functional mechanisms that underpinned them. This gave rise to the psychological mode of investigation that he went on to call 'psychoanalysis'. Its fundamental assumption was that manifest (nowadays called 'explicit' or 'declarative') subjective phenomena have latent (nowadays called 'implicit' or 'non-declarative') causes. That is, Freud argued that the erratic train of our conscious thoughts can be explained only if we assume implicit intervening links of which we are unaware. This gave rise to the notion of latent mental functions and, in turn, to Freud's famous conjecture of 'unconscious' intentionality.

Since no methods were available at the turn of the nineteenth century to investigate the physiology of unconscious mental events, their mechanisms could be inferred only from clinical observation. What Freud learnt in this way gave rise to his second fundamental claim. He observed that patients adopted a far-from-indifferent attitude to their inferred unconscious intentions; it appeared to be more a matter of being *unwilling* rather than unable to become aware of them. He called this tendency variously 'resistance', 'censorship', 'defence' and 'repression', and observed that it prevents emotional distress. This in turn revealed the pivotal role that *feelings* play in mental life, how they underpin all sorts of self-serving biases. These findings (obvious today) showed Freud that some of the major motivating forces in mental life are entirely subjective but also unconscious. Systematic investigation of those forces led him to his third fundamental claim. He concluded

that what ultimately underpinned feelings were bodily needs; that human mental life, no less than that of animals, was *driven* by the biological imperatives to survive and reproduce. These imperatives, for Freud, provided the link between the feeling mind and the physical body.

Freud took a remarkably subtle approach to the mind/body relationship. He realised that the psychological phenomena he studied were not straightforwardly reducible to physiological ones. As early as 1891 he argued that it was not possible to attribute psychological symptoms to neurophysiological processes without first reducing the relevant psychological and physiological phenomena (both sides of the equation) to their respective underlying *functions*. As noted earlier with reference to information processing, functions can be performed on various substrates.[7] It was only upon the common ground of function, Freud argued, that psychology and physiology can be reconciled. His goal was to explain psychological phenomena by means of 'metapsychological' functional laws (the term means 'beyond psychology').[8] Trying to skip over this functional level of analysis, jumping directly from psychology to physiology, is nowadays called the localisationist fallacy.[9]

Clearly, for Freud if not his followers, psychoanalysis was meant to be an interim step. Although his quest from the first had been to discern the laws underpinning our rich inner life of subjective experience, nevertheless mental life remained a *biological* problem for him.[10] As he wrote in 1914: 'all our provisional ideas in psychology will presumably someday be placed on an organic foundation'.[11] He therefore enthusiastically anticipated the day when psychoanalysis would once again join up with neuroscience:

> Biology is truly a land of unlimited possibilities. We may expect it to give us the most surprising information, and we cannot guess what answers it will return in a few dozen years [...] They may be of a kind which will blow away the whole of our artificial structure of hypothesis.[12]

This was not the wildly speculative Freud that I had learnt about as an undergraduate student. The 'Project' was a revelation to me, as it had been to Freud himself. He wrote to his friend Wilhelm Fliess at the time:

> In the course of a busy night [...] the barriers were suddenly raised, the veils fell away, and it was possible to see through from the details of the neuroses to the determinants of consciousness. Everything seemed to fit together, the gears were in mesh, the thing gave one the impression that it was really a machine and would soon run of itself.[13]

But the euphoria lasted only a short time. A month later he wrote: 'I can no longer understand the state of mind in which I hatched the "Psychology"; I cannot make out how I came to inflict it on you.'[14] Devoid of appropriate neuroscientific methods, Freud relied upon 'imaginings, transpositions and guesses' to translate his


image
not
available

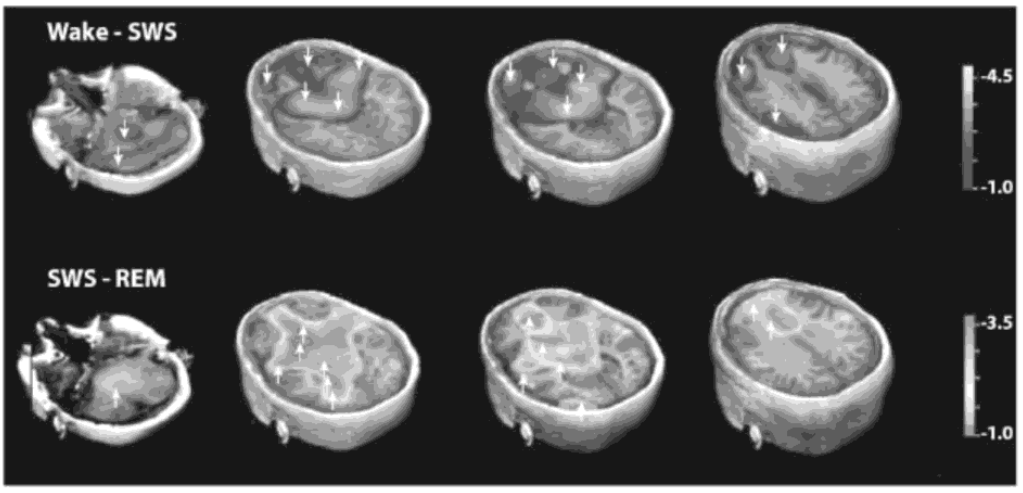

Figure 3 The horizontal rows show progressively higher slices through the brain (from left to right). The top row shows the difference between the awake and sleeping brain, with the shaded area depicting *decreased* cortical activation with sleep onset; the bottom row shows the difference between REM and non-REM (slow-wave) sleep, with the highlighted area depicting *increased* subcortical activation with REM onset. The area of greatest activation is where the SEEKING system is located.

In the early 1990s a neurosurgical colleague in South Africa referred to me Mr S, a patient on whom, ten months previously, he had performed an operation to remove a tumour that was growing under the frontal lobes of his brain and displacing his optic nerves. Mr S had suffered a small haemorrhage during the operation, which had interrupted the blood supply to the basal forebrain (see Figure 1). Basal forebrain nuclei transmit acetylcholine to various cortical and subcortical structures involved in retrieving long-term memories. These cholinergic pathways are thought to interact with dopaminergic pathways (see Figure 2), with the latter being the so-called reward system that activates 'search' behaviours, not only in relation to physical actions in the external world but also in relation to the inner world of representations, the imaginary actions that arise in thinking and in dreams.[26] As a result of his haemorrhage, Mr S awoke from surgery with a profound amnesic syndrome, known as Korsakoff psychosis, the central feature of which is a dream-like state called confabulation. His memory for recent events was profoundly disordered in such a way that he constantly retrieved false recollections. This *search* deficit is disabling enough, but in confabulatory amnesia it is compounded by the fact that patients do not adequately *monitor* the reliability of the memories they wrongly retrieve, and therefore treat them as if they are true when they patently are not.

For example, Mr S thought he was in Johannesburg (his home town) but he had in fact just travelled to London to consult me. He had no memory of the journey. When I corrected him on that score, he insisted that he could not possibly be in London. I therefore asked him to look out of the window, since it was snowing, which never happens in Johannesburg. Initially he looked shocked, but then he composed himself

and retorted: 'No, I *know* I'm in Jo'burg; just because you're eating pizza, it doesn't mean you're in Italy.'

Mr S was a fifty-six-year-old electrical engineer. I saw him in my daily out-patient clinic, six times per week, in an attempt to orientate him and help him gain some insight into the ways in which his memory was failing him. Although I saw him at the same time and place every day, he never recognised me as his therapist from one session to the next. He apparently knew my face, but routinely mistook me for someone else he knew in a different context – usually an engineering colleague who was working with him on some electronic problem, or a client seeking his professional assistance. In other words, Mr S treated me as if I were in need of his help, rather than the other way round. Another frequent misconception of his was that we were both university students, having a drink together after some sporting activity (either a rowing contest or a rugby match). I was young enough at the time for this to be plausible, but Mr S had not been a student for more than thirty years.

After each clinical session, I had a consultation with his wife in order to contextualise his misrememberings and attempt to establish their meaning. This was the main difference between the approach I was taking and the more traditional approach my colleagues took to 'cognitive rehabilitation'. Whereas neuropsychologists conventionally concern themselves with the *degree* of memory disorder, measured from the third-person viewpoint, I was more interested in the subjective *content* of Mr S's errors, understood from the first-person perspective. I started from the assumption that the personal significance of the events that compulsively came to his mind, in place of the target memories that he was searching for, would cast some light on the mechanism of these confabulations – and thereby open new paths to influencing them. So, in my meetings with his wife, for example, I wanted to know whether Mr S really did belong to rowing and rugby teams when he was a student and whether he really did provide professional help with electronic problems.

Two facts that I learnt in this way are relevant to understanding his confabulations. These were, first, that he had once suffered from chronic problems with his teeth – problems which had eventually been treated (successfully) using dental implants – and, second, that he suffered from cardiac arrhythmia, which was being controlled by a pacemaker.

I have selected a short transcription from an audio recording of the first few minutes of the tenth session I had with Mr S. I have chosen this particular snippet because, when I went to collect him from the waiting room that day, he appeared briefly (for the first time) to recognise who I was and why he was consulting me. As I entered the waiting room, he touched the craniotomy scar on the top of his head and said: 'Hi, Doc.'

I was hoping to build upon this glimmer of insight, if that is what it was, as we sat down in my office.

Me: You touched your head when we met in the waiting room.
Mr S: I think the problem is that a cartridge is missing. We must ... we just need the specs. What was it? A C49? Should we order it?
Me: What does a C49 cartridge do?
Mr S: Memory. It's a memory cartridge; a memory implant. But I never really understood it. In fact, I haven't used it for a good five or six months now. It seems we don't really need it. It was all chopped away by a doctor. What's his name? Dr Solms, I think. But it seems I don't really need it. The implants work fine.
Me: You are aware that something is wrong with your memory, but ...
Mr S: Yes, it's not working one hundred per cent, but we don't really need it – it was just missing a few beats. The analysis showed that there was some C or C09 missing. Denise [his first wife] brought me here to see a doctor. What's his name again? Dr Solms or something. And he did one of those heart transplant things, and now it is working fine; never misses a beat.
Me: You're aware that something is amiss. Some memories are missing, and, of course, that's worrying. You hope I can fix it, just like those other doctors fixed the problems with your teeth and your heart. But you want it so much that you are having difficulty accepting that it's not fixed already.
Mr S: Oh, I see. Yes, it's not working one hundred per cent. [He touches his head.] I got knocked on the head. Went off the field for a few minutes. But it's fine now. I suppose I shouldn't go back on. But you know me; I don't like going down. So, I asked Tim Noakes [a renowned South African sports physician] – because I've got the insurance, you know, so why not use it, why not go to the best – and he said: 'Fine, play on.'

I will break off the vignette there. It should be fairly easy to recognise the purely cognitive disorders of memory search and monitoring that I mentioned above. When Mr S saw me entering the waiting room for that tenth session, my appearance evoked a swarm of associations in him – to do with doctors, his head, missing memory, surgical procedures and the like. But in each of these instances, he did not retrieve the precise target memory he was searching for; instead, he came up with what might be characterised as near misses – memories that were in the same broad semantic categories as the targets, but which were mislocated in space and time. Thus, the idea of a 'doctor' evoked associations concerning the neurosurgeon and a famous sports physician instead of its target, me; the idea 'head' evoked a concussion incident instead of a brain tumour; 'missing memory' evoked an electronic cartridge instead of his amnesia; 'surgical procedures' evoked his earlier dental and cardiological procedures instead of the recent brain surgery, and so on. It is equally easy to see the monitoring deficit: Mr S accepted the veracity of his mistaken memories far too readily. The fact that he experienced himself as being a

twenty-something student on a rugby field (despite all the evidence to the contrary) is an obvious example of this. Likewise his belief that he was still in Johannesburg.

But when Mr S's confabulations are considered from the *subjective* point of view, additional facts emerge. Imagine what it *feels* like to suddenly realise that you do not recognise the clinician who just walked into the room, although he seems to be responsible for your care; that you do not know what room (or even which city) you are in; that you have a huge scar over the top of your head, and you do not know where it comes from; that – in fact – you do not remember what happened just two minutes ago, let alone over the days and months preceding the present moment. You would probably feel something like *panic*, wondering whether this doctor might have performed an operation on your head, as a result of which you no longer remember anything from one moment to the next. This is what missing memory search and monitoring mechanisms feel like to the intentional subject of the mind – to the living I.

Now, notice what Mr S did in consequence of having these feelings (in other words, notice what causal effects they had on his cognition). Upon realising that his 'memory cartridge' was missing, he (delusionally) reassured himself that *one can simply order a new one.* Not entirely convinced by his own reassurance, he changes his mind. In fact, *one does not really need the cartridge*, one manages fine without it, and he has done so for months already. He then makes a link between the missing cartridge and the craniotomy scar: apparently something has been chopped away by a doctor. He hopes that this is not the doctor sitting before him, and moreover he hopes that the operation has not been botched. At this point, Mr S recalls that his equivalent dental and cardiological operations were successful and he (delusionally) conflates these procedures with the present one: *it was a success*, *the implants work fine* and he 'never misses a beat'. When I introduce some doubt on this score, he changes tack. He agrees that it is not working 100 per cent, but he simultaneously decides that what has happened to his head was not surgery after all, *it was merely concussion*; he is suffering the temporary effects of a minor sporting accident. Accordingly, he has been sent off the field for a few minutes. But, happily, with access to the best sports physician money can buy, he is once again reassured: *he may play on.* All will be fine.

Considering Mr S's confabulations from the first-person perspective clearly reveals something new about them: the content of his misrememberings is tendentiously *motivated.* These are far from being random search errors. They contain a clear self-serving bias; they have the aim and purpose of recasting his anxiety-ridden situation into a reassuring, safe and familiar one. So, just as Freud inferred in the case of dreams, confabulations are motivated. The mental processes in confabulatory amnesia are *wishful.* But this fact becomes apparent only when the emotional context and personal meaning (experienced by Mr S alone) of dental implants ('the implants work fine') and cardiac pacemakers ('it never misses a beat') are taken into account – as a psychoanalyst would do. This is what

neuropsychologists fail to see when they aim to be entirely objective; as Sacks put it, when they exclude the psyche.

The first-person observational perspective I have just described also reveals something new about the *mechanism* of confabulation, something that is overlooked from the third-person viewpoint. It tells us that confabulation occurs not solely due to deficits in strategic search and source monitoring (i.e. missing 'memory cartridges') but also due to the release from inhibition of more *emotionally* mediated forms of recall, much as a child's memory might work. This psychodynamic mechanism has implications for the treatment of confabulation, and, of course, for the question of which brain processes are involved in it. Accurate memory search and monitoring functions turn out to depend in part upon the cholinergic basal forebrain circuits, which constrain the 'reward' mechanisms of the mesocortical-mesolimbic dopamine circuit in memory retrieval. As it happens, a similar unfettering of dopaminergic search occurs in dreams.[27] That is why I reported the case of Mr S to my colleagues under the title, 'The man who lived in a dream'.

This enabled me, as it had with dreams, to tentatively link the unconstrained dopaminergic 'reward' or 'wanting' or 'SEEKING' mechanism with Freud's notion of 'wish-fulfilment'[28] – a meta-psychological concept that was closely linked with his concept of 'drive'.[29] Conversely, the functions of the cholinergic forebrain nuclei can be linked in some respects with the inhibitory influences of 'reality-testing'.[30] In this way, I began to translate Freud's inferences about the functional mechanisms of subjectivity with their physiological equivalents.

These were my first steps. Naturally, such broad generalisations cannot be based on purely clinical evidence in a single case. Having formulated my impression of Mr S, therefore, I enlisted 'blind' raters (colleagues unfamiliar with my hypothesis) to measure, on a seven-point Likert scale, the degree of pleasantness versus unpleasantness in a continuous unselected sample of 155 of his confabulations. The results were statistically (highly) significant: when compared to the target memories they replaced, Mr S's confabulations substantially improved his situation from the emotional point of view.[31] Next, my research collaborators and I demonstrated the same strong effect in studies involving numerous other patients with confabulations. In subsequent empirical studies, the mood-regulating effects of confabulation that I inferred clinically in the case of Mr S were statistically validated.[32] This programme of research opened a whole new approach to the neuropsychology of confabulation,[33] and related disorders such as anosognosia.[34] It also laid the foundations for a novel approach to common psychiatric disorders such as addiction and major depression.[35] I have spent the last three decades developing this 'neuropsychoanalytic' approach to mental illness, trying to return subjectivity to neuroscience.[36]

As I accumulated clinical experiences of the kind just described during my psychoanalytic training in London, I was invited to report my findings in a series of