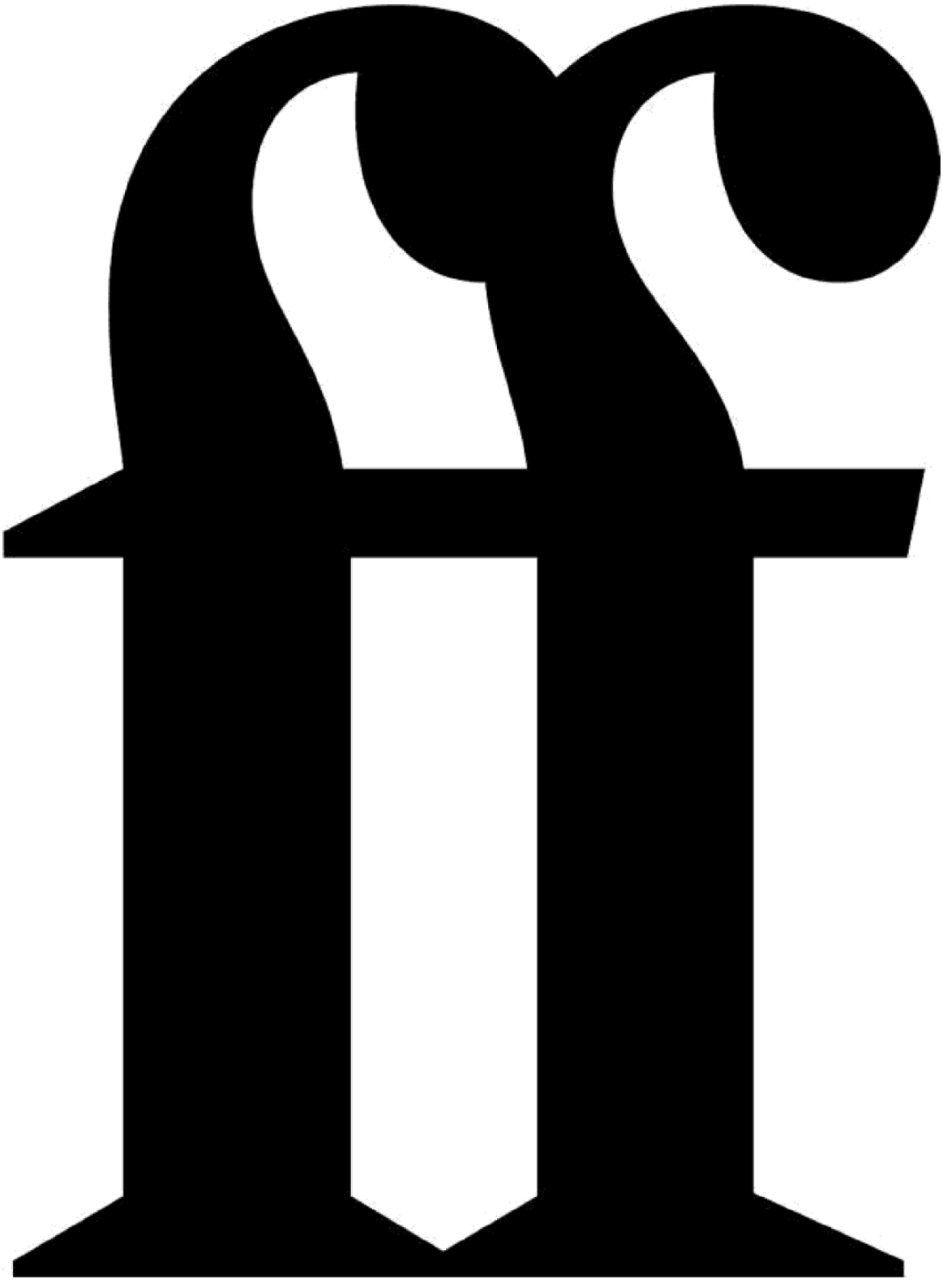

Contents

Title Page
Epigraph
Dedication
Prologue

I: LEVEL
1: The Real Problem
2: Measuring Consciousness
3: Phi

II: CONTENT
4: Perceiving from the Inside Out
5: The Wizard of Odds
6: The Beholder's Share

III: SELF
7: Delirium
8: Expect Yourself
9: Being a Beast Machine
10: A Fish in Water
11: Degrees of Freedom

IV: OTHER
12: Beyond Human
13: Machine Minds

Epilogue
Acknowledgements
References
Index
About the Author
Copyright

Prologue

Five years ago, for the third time in my life, I ceased to exist. I was having a small operation and my brain was filling with anaesthetic. I remember sensations of blackness, detachment, and falling apart ...

General anaesthesia is very different from going to sleep. It has to be; if you were asleep, the surgeon's knife would quickly wake you up. States of deep anaesthesia have more in common with catastrophic conditions like coma and the vegetative state, where consciousness is completely absent. Under profound anaesthesia, the brain's electrical activity is almost entirely quietened – something that never happens in normal life, awake or asleep. It is one of the miracles of modern medicine that anaesthesiologists can routinely alter people's brains so that they enter and return from such deeply unconscious states. It's an act of transformation, a kind of magic: anaesthesia is the art of turning people into objects.

The objects, of course, get turned back into people. So I returned, drowsy and disoriented but definitely *there*. No time seemed to have passed. Waking from a deep sleep, I am sometimes confused about the time, but there is always the impression that at least some amount of time has gone by, of a continuity between my consciousness then and my consciousness now. Under general anaesthesia, things are different. I could have been under for five minutes, five hours, five years – or even fifty. And 'under' doesn't quite express it. I was simply not there, a premonition of the total oblivion of death, and, in its absence of anything, a strangely comforting one.

General anaesthesia doesn't just work on your brain, or on your mind. It works on your consciousness. By altering the delicate electrochemical balance within the neural circuitry inside your head, the basic ground state of what it is to 'be' is – temporarily – abolished. In this process lies one of the greatest remaining mysteries in science, and in philosophy too.

Somehow, within each of our brains, the combined activity of billions of neurons, each one a tiny biological machine, is giving rise to a conscious experience. And not just any conscious experience, your conscious experience, right here, right now. *How does this happen? Why do we experience life in the first person?*

I have a childhood memory of looking in the bathroom mirror, and for the first time realising that my experience at that precise moment – the experience of *being me* – would at some point come to an end, and that 'I' would die. I must have been about eight or nine years old, and like all early memories it is unreliable. But perhaps it was at this moment that I also realised that if my consciousness could end, then it must depend in some way on the stuff I was made of – on the physical materiality of my body and my brain. It seems to me that I've been grappling with this mystery, in one way or another, ever since.

As an undergraduate student at Cambridge University in the early nineties, a

teenage romance with physics and philosophy broadened into a fascination with psychology and neuroscience, even though at the time these fields seemed to avoid, even outlaw, all mention of consciousness. My PhD research took me on a long and unexpectedly valuable detour through artificial intelligence and robotics, before a six-year stint at the Neurosciences Institute in San Diego, on the shores of the Pacific, finally delivered the chance to investigate the brain basis of consciousness directly. There, I worked with the Nobel Laureate Gerald Edelman – one of the most significant figures in bringing consciousness back into view as a legitimate scientific focus.

Now, for more than a decade, I've been Co-Director of a research centre – the Sackler Centre for Consciousness Science at the University of Sussex – nestled among the gentle green hills of the South Downs by the seaside city of Brighton. Our Centre brings together neuroscientists, psychologists, psychiatrists, brain imagers, virtual reality wizards and mathematicians, and philosophers, all of us trying to open new windows onto the brain basis of conscious experience.

—

Whether you're a scientist or not, consciousness is a mystery that matters. For each of us, our conscious experience is *all there is*. Without it there is nothing at all: no world, no self, no interior and no exterior.

Imagine that a future version of me, perhaps not so far away, offers you the deal of a lifetime. I can replace your brain with a machine that is its equal in every way, so that from the outside, nobody could tell the difference. This new machine has many advantages – it is immune to decay, and perhaps it will allow you to live forever.

But there's a catch. Since even future-me is not sure how real brains give rise to consciousness, I can't guarantee that you will have any conscious experiences at all, should you take up this offer. Maybe you will, if consciousness depends only on functional capacity, on the power and complexity of the brain's circuitry, but maybe you won't, if consciousness depends on a specific biological material – neurons, for example. Of course, since your machine-brain leads to identical behaviour in every way, when I ask new-you whether you are conscious, new-you will say yes. But what if, despite this answer, life – for you – is no longer in the first person?

I suspect you wouldn't take the deal. Without consciousness, it may hardly matter whether you live for another five years or another five hundred. In all that time *there would be nothing it would be like to be you.*

Philosophical games aside, the practical importance of understanding the brain basis of consciousness is easy to appreciate. General anaesthesia has to count as one of the greatest inventions of all time. Less happily, distressing disturbances of consciousness can accompany brain injuries and mental illnesses for the increasing number of us, me included, who encounter these conditions. And for each one of us, conscious experiences change throughout life, from the blooming and buzzing confusion of early life, through the apparent though probably illusory and certainly

not universal clarity of adulthood, and on to our final drift into the gradual – and for some, disorientingly rapid – dissolution of the self as neurodegenerative decay sets in. At each stage in this process you exist, but the notion that there is a single unique conscious self (a soul?) that persists over time may be grossly mistaken. Indeed, one of the most compelling aspects of the mystery of consciousness is the nature of *self*. Is consciousness possible without self-consciousness, and if so would it still matter so much?

Answers to difficult questions like these have many implications for how we think about the world and the life it contains. When does consciousness begin in development? Does it emerge at birth, or is it present even in the womb? What about consciousness in non-human animals – and not just in primates and other mammals, but in otherworldly creatures like the octopus and perhaps even in simple organisms such as nematode worms or bacteria? Is there anything it is like to be an *Escherichia coli*, or a sea bass? What about future machines? Here, we ought to be concerned not just about the power that new forms of artificial intelligence are gaining over us, but also about whether and when *we* need to take an ethical stance towards *them*. For me, these questions evoke the uncanny sympathy I felt when watching Dave Bowman destroy HAL's personality in the film *2001: A Space Odyssey*, by the simple act of removing its memory banks, one by one. In the greater empathy elicited by the plight of Ridley Scott's replicants in *Blade Runner* there is a clue about the importance of our nature as *living machines* for the experience of being a conscious self.

—

This book is about the neuroscience of consciousness: the attempt to understand how the inner universe of subjective experience relates to, and can be explained in terms of, biological and physical processes unfolding in brains and bodies. This is the project that has captivated me throughout my career, and I believe it has now reached a point at which glimmerings of answers are beginning to emerge.

These glimmerings already change, and change dramatically, how we think about conscious experiences of the world around us, and of ourselves within it. The way we think about consciousness touches every aspect of our lives. A science of consciousness is nothing less than an account of who we are, of what it is to be me, or to be you, and of why there is anything it is like to 'be' at all.

The story I will tell is a personal view, shaped over many years of research, contemplation, and conversation. The way I see it, consciousness won't be 'solved' in the same way that the human genome was decoded, or the reality of climate change established. Nor will its mysteries suddenly yield to a single eureka-like insight – a pleasant but usually inaccurate myth about how scientific understanding progresses.

For me, a science of consciousness should explain how the various properties of consciousness depend on, and relate to, the operations of the neuronal wetware inside our heads. The goal of consciousness science should not be – at least not primarily – to explain why consciousness happens to be part of the universe in the

first place. Nor should it be to understand how the brain works in all its complexity, while sweeping the mystery of consciousness away under the carpet. What I hope to show you is that by accounting for properties of consciousness, in terms of mechanisms in brains and bodies, the deep metaphysical whys and hows of consciousness become, little by little, less mysterious.

I use the word 'wetware' to underline that brains are not computers made of meat. They are chemical machines as much as they are electrical networks. Every brain that has ever existed has been part of a living body, embedded in and interacting with its environment – an environment which in many cases contains other embodied brains. Explaining the properties of consciousness in terms of biophysical mechanisms requires understanding brains – and conscious minds – as *embodied* and *embedded* systems.

In the end, I want to leave you with a new conception of the self – that aspect of consciousness which for each of us is probably the most meaningful. An influential tradition, dating back at least as far as Descartes in the seventeenth century, held that non-human animals lacked conscious selfhood because they did not have rational minds to guide their behaviour. They were 'beast machines': flesh automatons without the ability to reflect on their own existence.

I don't agree. In my view, consciousness has more to do with being alive than with being intelligent. We are conscious selves precisely *because* we are beast machines. I will make the case that experiences of *being you*, or of *being me*, emerge from the way the brain predicts and controls the internal state of the body. The essence of selfhood is neither a rational mind nor an immaterial soul. It is a deeply embodied biological *process*, a process that underpins the simple feeling of being alive that is the basis for all our experiences of self, indeed for any conscious experience at all. *Being you* is literally about your body.

This book is divided into four parts. In the first part, I explain my approach to the scientific study of consciousness. This part also deals with the question of conscious 'level' – of how conscious someone or something can be – and with progress in attempts to 'measure' consciousness. The second part takes on the topic of conscious 'content' – of what you are conscious of, when you are conscious. Part three turns the focus inwards, to the self, and to all the varied experiences that conscious selfhood entails. The fourth and final part – 'other' – explores what this new way of understanding consciousness can say about other animals, and about the possibility of sentient machines. By the end of the book, you'll understand that our conscious experiences of the world and the self are forms of brain-based prediction – 'controlled hallucinations' – that arise with, through, and *because of* our living bodies.

—

Despite his tarnished reputation among neuroscientists, Sigmund Freud was right about many things. Looking back through the history of science, he identified three 'strikes' against the perceived self-importance of the human species, each marking a

major scientific advance that was strongly resisted at the time. The first was by Copernicus, who showed with his heliocentric theory that the Earth rotates around the sun, and not the other way around. With this dawned the realisation that we are not at the centre of the universe; we are just a speck somewhere out there in the vastness, a pale blue dot suspended in the abyss. Next came Darwin, who revealed that we share common ancestry with all other living things, a realisation that is – astonishingly – still resisted in some parts of the world even today. Immodestly, Freud's third strike against human exceptionalism was his own theory of the unconscious mind, which challenged the idea that our mental lives are under our conscious, rational control. While he may have been off target in the details, Freud was absolutely right to point out that a naturalistic explanation of mind and consciousness would be a further, and perhaps final, dethronement of humankind.

These shifts in how we see ourselves are to be welcomed. With each new advance in our understanding comes a new sense of wonder, and a new ability to see ourselves as less *apart from*, and more *a part of*, the rest of nature.

Our conscious experiences are part of nature just as our bodies are, just as our world is. And when life ends, consciousness will end too. When I think about this, I am transported back to my experience – my *non*-experience – of anaesthesia. To its oblivion, perhaps comforting, but oblivion nonetheless. The novelist Julian Barnes, in his meditation on mortality, puts it perfectly. When the end of consciousness comes there is nothing – really *nothing* – to be frightened of.

Notes

meditation on mortality: Barnes (2008).


image
not
available


image
not
available

human animals lack consciousness. And complex abstract thinking is just one small part – though possibly a distinctively human part – of being conscious.

Some prominent theories in the science of consciousness continue to emphasise function and behaviour over phenomenology. Foremost among these is the 'global workspace' theory, which has been developed over many years by the psychologist Bernard Baars and the neuroscientist Stanislas Dehaene, among others. According to this theory, mental content (perceptions, thoughts, emotions, and so on) becomes conscious when it gains access to a 'workspace', which – anatomically speaking – is distributed across frontal and parietal regions of the cortex. (The cerebral cortex is the massively folded outer surface of the brain, made up of tightly packed neurons.[†]) When mental content is broadcast within this cortical workspace, we are conscious of it, and it can be used to guide behaviour in much more flexible ways than is the case for unconscious perception. For example, I am consciously aware of a glass of water on the table in front of me. I could pick it up and drink it, throw it over my computer (tempting), write a poem about it, or take it back into the kitchen now that I realise it's been there for days. Unconscious perception does not allow this degree of behavioural flexibility.

Another prominent theory, called 'higher-order thought' theory, proposes that mental content becomes conscious when there is a 'higher-level' cognitive process that is somehow oriented towards it, rendering it conscious. On this theory, consciousness is closely tied to processes like *metacognition* – meaning 'cognition about cognition' – which again emphasises functional properties over phenomenology (though less so than global workspace theory). Like global workspace theory, higher-order thought theories also emphasise frontal brain regions as key for consciousness.

Although these theories are interesting and influential, I won't have much more to say about either in this book. This is because they both foreground the functional and behavioural aspects of consciousness, whereas the approach I will take starts from phenomenology – from experience itself – and only from there has things to say about function and behaviour.

The definition of consciousness as 'any kind of subjective experience whatsoever' is admittedly simple and may even sound trivial, but this is a good thing. When a complex phenomenon is incompletely understood, prematurely precise definitions can be constraining and even misleading. The history of science has demonstrated many times over that useful definitions evolve in tandem with scientific understanding, serving as scaffolds for scientific progress, rather than as starting points, or ends in themselves. In genetics, for example, the definition of a 'gene' has changed considerably as molecular biology has advanced. In the same way, as our understanding of consciousness develops, its definition – or definitions – will evolve too. If, for now, we accept that consciousness is first and foremost about phenomenology, then we can move on to the next question.

—

How does consciousness happen? How do conscious experiences relate to the biophysical machinery inside our brains and our bodies? How indeed do they relate to the swirl of atoms or quarks or superstrings, or to whatever it is that the entirety of our universe ultimately consists in?

The classic formulation of this question is known as the 'hard problem' of consciousness. This expression was coined by the Australian philosopher David Chalmers in the early 1990s, and it has set the agenda for much of consciousness science ever since. Here is how he describes it:

> It is undeniable that some organisms are subjects of experience. But the question of how it is that these systems are subjects of experience is perplexing. Why is it that when our cognitive systems engage in visual and auditory information-processing, we have visual or auditory experience: the quality of deep blue, the sensation of middle C? How can we explain why there is something it is like to entertain a mental image, or to experience an emotion? It is widely agreed that experience arises from a physical basis, but we have no good explanation of why and how it so arises. Why should physical processing give rise to a rich inner life at all? It seems objectively unreasonable that it should, and yet it does.

Chalmers contrasts this hard problem of consciousness with the so-called easy problem – or easy problems – which have to do with explaining how physical systems, like brains, can give rise to any number of functional and behavioural properties. These functional properties include things like processing sensory signals, selection of actions and the control of behaviour, paying attention, the generation of language, and so on. The easy problems cover all the things that beings like us can do and that can be specified in terms of a function – how an input is transformed into an output – or in terms of a behaviour.

Of course, the easy problems are not easy at all. Solving them will occupy neuroscientists for decades or centuries to come. Chalmers' point is that the easy problems are easy to solve in principle, while the same cannot be said for the hard problem. More precisely, for Chalmers there is no conceptual obstacle to easy problems eventually yielding to explanations in terms of physical mechanisms. By contrast, for the hard problem it seems as though no such explanation could ever be up to the job. (A 'mechanism' – to be clear – can be defined as a system of causally interacting parts that produce effects.) Even after all the easy problems have been ticked off, one by one, the hard problem will remain untouched. '[E]ven when we have explained the performance of all the functions in the vicinity of experience – perceptual discrimination, categorization, internal access, verbal report – there may still remain a further unanswered question: *Why is the performance of these functions accompanied by experience?*'

The roots of the hard problem extend back to ancient Greece, perhaps even earlier, but they are particularly visible in René Descartes' seventeenth-century sundering of the universe into mind stuff, *res cogitans*, and matter stuff, *res extensa*.

This distinction inaugurated the philosophy of dualism, and has made all discussions of consciousness complicated and confusing ever since. This confusion is most evident in the proliferation of different philosophical frameworks for thinking about consciousness.

Take a deep breath, here come the 'isms'.

My preferred philosophical position, and the default assumption of many neuroscientists, is *physicalism*. This is the idea that the universe is made of physical stuff, and that conscious states are either identical to, or somehow emerge from, particular arrangements of this physical stuff. Some philosophers use the term *materialism* instead of physicalism, but for our purposes they can be treated synonymously.

At the other extreme to physicalism is *idealism*. This is the idea – often associated with the eighteenth-century Bishop George Berkeley – that consciousness or mind is the ultimate source of reality, not physical stuff or matter. The problem isn't how mind emerges from matter, but how matter emerges from mind.

Sitting awkwardly in the middle, *dualists* like Descartes believe that consciousness (mind) and physical matter are separate substances or modes of existence, raising the tricky problem of how they ever interact. Nowadays few philosophers or scientists would explicitly sign up for this view. But for many people, at least in the West, dualism remains beguiling. The seductive intuition that conscious experiences *seem* non-physical encourages a 'naïve dualism' where this 'seeming' drives beliefs about how things actually are. As we'll see throughout this book, the way things seem is often a poor guide to how they actually are.

One particularly influential flavour of physicalism is *functionalism*. Like physicalism, functionalism is a common and often unstated assumption of many neuroscientists. Many who take physicalism for granted also take functionalism for granted. My own view, however, is to be agnostic and slightly suspicious.

Functionalism is the idea that consciousness does not depend on what a system is made of (its physical constitution), but only on what the system does, on the functions it performs, on how it transforms inputs into outputs. The intuition driving functionalism is that mind and consciousness are forms of information processing which can be implemented by brains, but for which biological brains are not strictly necessary.

Notice how the term 'information processing' sneaked in here unannounced (as it also did in the quote from Chalmers a few pages back). This term is so prevalent in discussions of mind, brain, and consciousness that it's easy to let it slide by. This would be a mistake, because the suggestion that the brain 'processes information' conceals some strong assumptions. Depending on who's doing the assuming, these range from the idea that the brain is some kind of computer, with mind (and consciousness) being the software (or 'mindware'), to assumptions about what information itself actually *is*. All of these assumptions are dangerous. Brains are very different from computers, at least from the sorts of computers that we are familiar with. And the question of what information 'is' is almost as vexing as the question of

what consciousness is, as we'll see later on in this book. These worries are why I'm suspicious of functionalism.

Taking functionalism at face value, as many do, carries the striking implication that consciousness is something that can be *simulated* on a computer. Remember that for functionalists, consciousness depends only on what a system does, not on what it is made of. This means that if you get the functional relations right – if you ensure that a system has the right kind of 'input–output mappings' – then this will be enough to give rise to consciousness. In other words, for functionalists, *simulation* means *instantiation* – it means coming into being, in reality.

How reasonable is this? For some things, simulation certainly counts as instantiation. A computer that plays Go, such as the world-beating AlphaGo Zero from the British artificial intelligence company DeepMind, *is actually playing Go*. But there are many situations where this is not the case. Think about weather forecasting. Computer simulations of weather systems, however detailed they may be, do not get wet or windy. Is consciousness more like Go or more like the weather? Don't expect an answer – there isn't one, at least not yet. It's enough to appreciate that there's a valid question here. This is why I'm agnostic about functionalism.

There are two more 'isms', then we're done.

The first is *panpsychism*. Panpsychism is the idea that consciousness is a fundamental property of the universe, alongside other fundamental properties such as mass/energy and charge; that it is present to some degree everywhere and in everything. People sometimes make fun of panpsychism for claiming things like stones and spoons are conscious in the same sort of way that you or I are, but these are usually deliberate misconstruals designed to make it look silly. There are more sophisticated versions of the idea, some of which we will meet in later chapters, but the main problems with panpsychism don't lie with its apparent craziness – after all, some crazy ideas turn out to be true, or at least useful. The main problems are that it doesn't really explain anything and that it doesn't lead to testable hypotheses. It's an easy get-out to the apparent mystery posed by the hard problem, and taking it on ushers the science of consciousness down an empirical dead end.

Finally, there's *mysterianism*, which is associated with the philosopher Colin McGinn. Mysterianism is the idea that there may exist a complete physical explanation of consciousness – a full solution to Chalmers' hard problem – but that we humans just aren't clever enough, and never will be clever enough, to discover this solution, or even to recognise a solution if it were presented to us by super-smart aliens. A physical understanding of consciousness exists, but it lies as far beyond us as an understanding of cryptocurrency lies beyond frogs. It is cognitively closed to us by our species-specific mental limitations.

What can be said about mysterianism? There may well be things we will never understand, thanks to the limitations of our brains and minds. Already, no single person is able to fully comprehend how an Airbus A380 works. (And yet I'm happy to sit in one, as I did one time on the way home from Dubai.) There are certainly things which remain cognitively inaccessible to most of us, even if they are understandable

by humans in principle, like the finer points of string theory in physics. Since brains are physical systems with finite resources, and since some brains seem incapable of understanding some things, it seems inescapable that there must be some things which are the case, but which no human could ever understand. However, it is unjustifiably pessimistic to pre-emptively include consciousness within this uncharted domain of species-specific ignorance.

One of the more beautiful things about the scientific method is that it is cumulative and incremental. Today, many of us can understand things that would have seemed entirely incomprehensible *even in principle* to our ancestors, maybe even to scientists and philosophers working just a few decades ago. Over time, mystery after mystery has yielded to the systematic application of reason and experiment. If we take mysterianism as a serious option we might as well all give up and go home. So, let's not.

These 'isms' provide different ways of thinking about the relationship between consciousness and the universe as a whole. When weighing their merits and demerits, it's important to recognise that what matters most is not which framework is 'right' in the sense of being provably true, but which is most useful for advancing our understanding of consciousness. This is why I tend towards a functionally agnostic flavour of physicalism. To me, this is the most pragmatic and productive mindset to adopt when pursuing a science of consciousness. It is also, as far as I am concerned, the most intellectually honest.

—

Despite its appeal, physicalism is by no means universally accepted among consciousness researchers. One of the most common challenges to physicalism is the so-called 'zombie' thought experiment. The zombies in question here are not the brain-munching semi-corpses from the movies – our zombies are 'philosophical zombies'. But we need to get rid of them all the same, since otherwise the prospect of a natural, physicalist explanation of consciousness is dead in the water before we get started.

A philosophical zombie is a creature that is indistinguishable from a conscious creature, but which lacks consciousness. A zombie Anil Seth would look like me, act like me, walk like me and talk like me, but there would be nothing it is like to *be* it, no inner universe, no felt experience. Ask zombie Anil if he is conscious, and he will say, 'Yes, I'm conscious.' Zombie Anil would even have written various essays on the neuroscience of consciousness, including some thoughts about the questionable relevance of philosophical zombies to this topic. But none of this would involve any conscious experience whatsoever.

Here's why the zombie idea is supposed to provide an argument against physicalist explanations of consciousness. If you can imagine a zombie, this means you can conceive of a world that is indistinguishable from our world, but in which no consciousness is happening. And if you can conceive of such a world, then consciousness cannot be a physical phenomenon.

make great progress in shedding light on the properties and nature of conscious experiences without it being necessary to explain how or why they happen to be part of the universe in which we live.

Nor should we necessarily expect scientific explanations always to be intuitively satisfying. In physics, quantum mechanics is notoriously counterintuitive but is nonetheless widely accepted as providing our current best grip on the nature of physical reality. It could equally be that a mature science of consciousness will allow us to explain, predict, and control phenomenological properties without ever delivering the intuitive feeling that 'yes, this is right, *of course* it has to be this way!'

Importantly, the real problem of consciousness is not an admission of defeat to the hard problem. The real problem goes after the hard problem indirectly, but it still goes after it. To understand why this is so, let me introduce the 'neural correlates of consciousness'.

—

It still amazes me how disreputable consciousness science was, even just thirty years ago. In 1989, one year before I started my undergraduate degree at Cambridge University, the leading psychologist Stuart Sutherland wrote: 'Consciousness is a fascinating but elusive phenomenon. It is impossible to specify what it is, what it does, or why it evolved. Nothing worth reading has been written on it.' This damning summary appeared in no lesser place than the *International Dictionary of Psychology*, and it captures the attitude to consciousness that I often encountered in my first steps into academia.

Elsewhere, far away from Cambridge and though I did not know it at the time, the situation was more promising. Francis Crick (the co-discoverer, with Rosalind Franklin and James Watson, of the molecular structure of DNA) and his colleague Christof Koch, who were both based in San Diego in California, were setting out what would become the dominant method in the rise of consciousness science – the search for the neural correlates of consciousness.

The gold-standard definition of a neural correlate of consciousness, or NCC, is 'the minimal neuronal mechanisms jointly sufficient for any one specific conscious percept'. The NCC approach proposes that there is some specific pattern of neural activity that is responsible for any and every experience, such as the experience of 'seeing red'. Whenever this activity is present, an experience of redness will happen, and whenever it isn't, it won't.

The great merit of the NCC approach is that it offers a practical recipe for doing research. To identify an NCC, all you need to do is concoct a situation in which people sometimes have a particular conscious experience, and at other times do not, while making sure that these conditions are otherwise as closely matched as possible. Given such a situation, you then compare activity in the brain between the two conditions, using brain imaging methods such as functional magnetic resonance imaging (fMRI) or electroencephalography (EEG).[‖] The brain activity specific to the 'conscious' condition reflects the NCC for that particular experience.

The phenomenon of 'binocular rivalry' offers a helpful example. In binocular rivalry, a different image is shown to each eye – perhaps a picture of a face to the left eye and a picture of a house to the right eye. In this situation, conscious perception doesn't settle on a weird face-house chimera. It flips back and forth between the face and the house, dwelling for a few seconds on each. First you see a house, then a face, then a house again ... and so on. What's important here is that conscious perception changes even though the sensory input remains constant. By looking at what happens in the brain, it's therefore possible to distinguish brain activity that tracks conscious perception from activity that tracks whatever the sensory input happens to be. The brain activity that goes along with the conscious perception identifies the NCC for that perception.

The NCC strategy has been impressively productive over many years, delivering reams of fascinating findings, but its limitations are becoming apparent. One problem is that it is difficult, and perhaps in the end impossible, to disentangle a 'true' NCC from a range of potentially confounding factors, the most important of which are those neural happenings that are either prerequisites for, or consequences of, an NCC itself. In the case of binocular rivalry, brain activity that goes along with the conscious perception may also track upstream (prerequisite) processes like 'paying attention' and, on the downstream side, the verbal behaviour of 'reporting' – of saying that you see a house or a face. Although related to the flow of conscious perception, the neural mechanisms responsible for attention and verbal report – or other prerequisites and downstream consequences – should not be confused with those that are responsible for the conscious perception itself.

The deeper problem is that *correlations* are not *explanations*. We all know that mere correlation does not establish causation, but it is also true that correlation falls short of explanation. Even with increasingly ingenious experimental designs and ever more powerful brain imaging technologies, correlation by itself can never amount to explanation. From this perspective, the NCC strategy and the hard problem are natural bedfellows. If we restrict ourselves to collecting correlations between things happening in the brain and things happening in our experience, it is no surprise that we will always suspect an explanatory gap between the physical and the phenomenal. But if we instead move beyond establishing correlations to discover explanations that connect properties of neural mechanisms to properties of subjective experience, as the real problem approach advocates, then this gap will narrow and might even disappear entirely. When we are able to predict (and explain, and control) why the experience of redness is the particular way it is – and not like blueness, or like jealousy – the mystery of how redness happens will be less mysterious, or perhaps no longer mysterious at all.

The ambition of the real problem approach is that as we build ever sturdier explanatory bridges from the physical to the phenomenological, the hard-problem intuition that consciousness can never be understood in physical terms will fade away, eventually vanishing in a puff of metaphysical smoke. When it does we will have in our hands a satisfactory and fully satisfying science of conscious experience.

Copyright

This ebook edition first published in the UK in 2021
by Faber & Faber Ltd
Bloomsbury House
74–77 Great Russell Street
London WC1B 3DA

Graphs and line illustrations by Steve West at LazyChief

Cover design by Faber
Cover image from Shutterstock

ISBN 978–0–571–33773–6