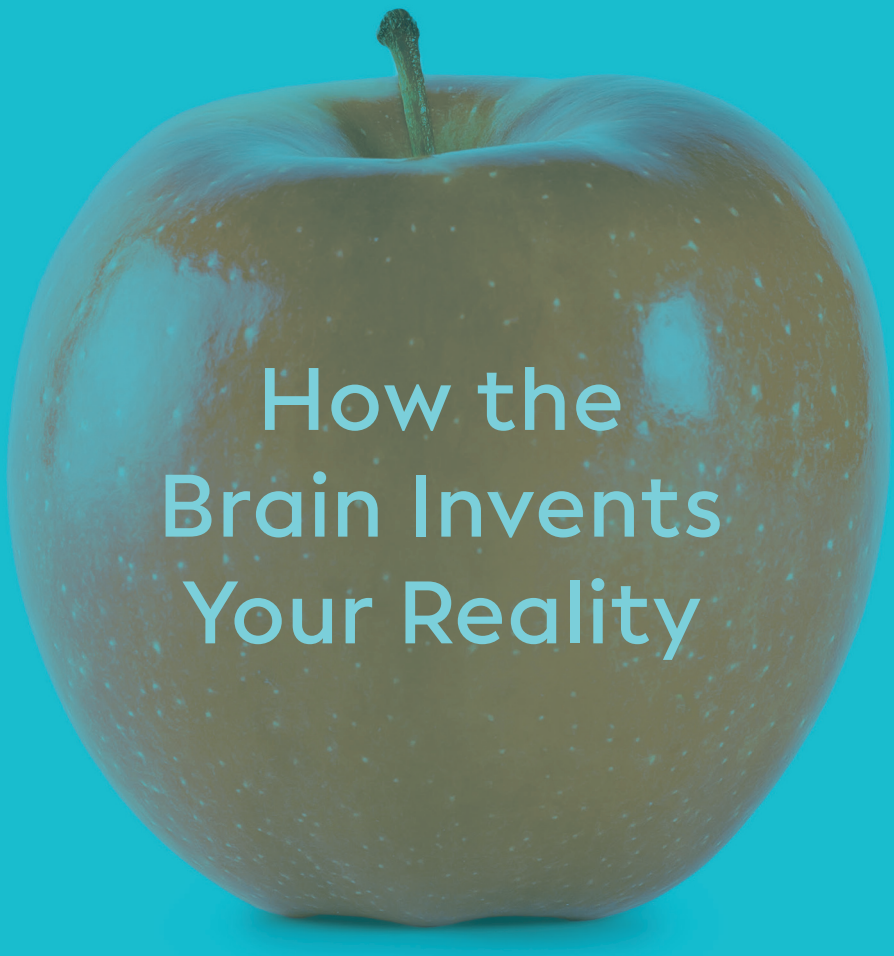
'Fascinating . . . will profoundly change the way you consider your own mind' **Lewis Dartnell, bestselling author of *Origins***

A Trick of the Mind



How the Brain Invents Your Reality

Daniel Yon

Cornerstone Press

A Trick of the Mind

Copyrighted Material

Copyrighted Material

A Trick of the Mind

DANIEL YON

Cornerstone Press

Copyrighted Material

CORNERSTONE PRESS

UK | USA | Canada | Ireland | Australia
India | New Zealand | South Africa

Cornerstone Press is part of the Penguin Random House group of companies
whose addresses can be found at global.penguinrandomhouse.com

Penguin Random House UK,
One Embassy Gardens, 8 Viaduct Gardens, London SW11 7BW

penguin.co.uk
global.penguinrandomhouse.com



First published 2025
001

Copyright © Daniel Yon, 2025

The moral right of the author has been asserted

Penguin Random House values and supports copyright. Copyright fuels creativity, encourages diverse voices, promotes freedom of expression and supports a vibrant culture. Thank you for purchasing an authorised edition of this book and for respecting intellectual property laws by not reproducing, scanning or distributing any part of it by any means without permission. You are supporting authors and enabling Penguin Random House to continue to publish books for everyone. No part of this book may be used or reproduced in any manner for the purpose of training artificial intelligence technologies or systems. In accordance with Article 4(3) of the DSM Directive 2019/790, Penguin Random House expressly reserves this work from the text and data mining exception.

Typeset in 12/14.75pt Dante MT Std by Jouve (UK), Milton Keynes

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

The authorised representative in the EEA is Penguin Random House Ireland,
Morrison Chambers, 32 Nassau Street, Dublin D02 YH68

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

ISBN: 978–1–529–90006–4 (hardback)
ISBN: 978–1–529–90007–1 (trade paperback)

Penguin Random House is committed to a sustainable future
for our business, our readers and our planet. This book is made from
Forest Stewardship Council® certified paper.



Copyrighted Material

For Rosa

Copyrighted Material

Copyrighted Material

Contents

Introduction 1

PART ONE: THE MATERIAL WORLD

1. Measuring Reality 11
2. Cause and Effect 45

Interlude I: Nullius in Verba 73

PART TWO: THE MENTAL WORLD

3. Other Minds 79
4. Knowing Our Own Minds 119

Interlude II: The Library of Babel 151

PART THREE: THE WORLD OF IDEAS

5. Wonderlust 157
6. Originality 176
7. Paradigm Shifts 198

Conclusion 234

Acknowledgements 239

Notes 241

Index 271

Copyrighted Material

Copyrighted Material

Introduction

It's like being inside a noisy coffin.

If you come to take part in one of our experiments, we'll greet you warmly at reception, before leading you deeper into the bowels of the building where the machine lives. Then we'll ask you a few questions. We'll check that you aren't wearing any watches, belts, rings, or earrings. That you don't have a pacemaker in your chest, or any metal rods propping up your limbs. No shrapnel from any explosions, no recent fillings, no wires or braces. No tattoos from the more laissez-faire countries where some artists mix their ink with a little too much lead. We'll check that you aren't pregnant, and that your contraceptive implant – if you have one – is the kind that won't heat up in a strong magnetic field. And we'll check one more time that you don't mind being left in tight, enclosed spaces.

Once every sliver of metal has been shorn, we'll take you inside.

The first thing you'll hear is the dull thrumming beat of the liquid-helium pumps that keep the innards of the machine implausibly cold. As we walk towards the source of that distant war drum, the scanner itself comes into view, a great white block with a cylindrical hole bored through its middle. That hole is where we'll put *you*. We'll lay you flat, padding your head with foam so it stays fixed in place. Then we'll arrange a cage over your face, and tape a panic alarm to your chest that you can squeeze if you want the whole thing to stop. Once that's done, we'll feed you slowly, head-first, into the machine's open mouth – and head into the console room next door.

Then the experiment really begins. The scanner whirs into life, issuing a strange cacophony of crackles and whistles, a robot's idea of a symphony. These sounds are in fact the noises made by the scanner coils as they shoot magnetic pulses into your head, and catch

Copyrighted Material

the echoes to generate a picture of what's happening inside. As you lie there, we'll flash images to your eyes, sounds to your ears. Ask you to think. To decide. To feel. And every time you see or hear or think or feel something, the crossfire of magnetic pulses running through your head will render an image of the brain inside, the one that's conjuring up the contents of your mind in real time.

Sitting in the console room, watching as our scanner renders an image of the brain inside your head, can be a strangely humbling experience. Every thought you've ever had, every choice you've ever made, everything you've ever felt – in short, your *mind* – is only made possible by these rivulets of salt and protein and fat that fold up inside your skull.

As a cognitive neuroscientist, it's my job to pull back the curtain on nature's strangest trick, to try to understand how, in a cold dark universe, arranging matter in our heads in this particular way makes a mind *happen*. And to explain how the peaks and troughs of the folded flesh in our skulls make up the edges and contours of the minds they conjure up.

Looking at *your* brain is trippy enough, but things get even weirder when I think that the same thing is happening inside *my* head too. As I sit inside that console room and use the scanner to peek inside your skull, a brain regards a brain, a mind a mind. All the ideas I can muster to make sense of what's happening inside your head depend on the processes and patterns unfolding in mine. Though science is full of many strange wonders, there's something particularly peculiar about my patch, where the object and the instrument of inquiry are precisely the same thing. Perhaps making minds happen is nature's strangest trick, because minds are the only way that matter can contemplate itself. And maybe the strangeness of that thought explains why – even though I've had my own brain scanned dozens of times – I've never had the courage to look at the pictures.

Our relationships with our brains are complicated. That's because our minds seem to have a double life, simultaneously a source of pride and a source of shame. On the one hand, the brain is the site

Copyrighted Material

of some heady wonders: this is the organ that discovered penicillin, invented democracy, law, literature, and art; that tamed the earth, the seas, and the skies; that put itself on the moon. But on the other hand, the human mind is a fragile thing. It's these very same brains that leave so many of us ensnared by superstition, prejudice, and bias; in the thrall of fringe political movements or bizarre conspiracy theories; or – just as insidiously – in personal prisons, with feelings, thoughts, and experiences that make our minds miserable, threatening, or unsettling places to be.

Scientists like me scramble for an idea, a theory, a story to make sense of the brain and this apparent duality. The brain's apparent double life has led lots of us to think that our theories of the brain need two prongs too, one capturing what makes our minds so impressive, the other explaining how the weaknesses and frailties arise**.** These theories tend to render a picture of the mind as a biological cut-and-shut, some kind of biological supercomputer designed by evolution, welded together with a bundle of irrational instincts running on legacy hardware that nature selected for something else. In this dichotomous picture, whether our minds succeed or err just depends on which of our mind's two halves has the upper hand.

Parcelling up the mind like this makes for a seductive story. The part of us that's nourished by fairy tales likes the idea of good and bad, black and white. But this picture of the brain that feels so intuitive to us on the inside doesn't match up with the picture of the brain that scientists can see from the outside. When I'm sat in that console room, blowing up your brain on my monitor, your grey matter doesn't neatly divide into black and white. I don't see the edges of two separate minds grafted together, or the dark parts of the cortex jostling with your brain's better angels.

What we actually see are communities of entangled circuits, trading information near and far, bundles of cells burning metabolic fuel, collaborating as a whole to make your mind happen. However the triumphs and tribulations of our very human minds arise, they do so via this one, single unfolding process. But if the

Copyrighted Material

brain doesn't have a double life, what's the best way to describe its single occupation?

In recent years, a new idea has begun to transform how scientists like me think. The science has gone full circle. In scrambling for an idea to make sense of the brain, scientists have begun to think that – actually – the brain might be just like them. The brain is just a single thing – a *scientist*. And this is intended as both an insult and a compliment.

The scientist's task is to try to understand reality one step removed. Nature doesn't give up her secrets easily. A physicist can't talk to subatomic particles to find out how they work, and a biologist can't interrogate a single cell until it spills its organelles. Whatever purchase scientists get on reality, they arrive at it through their *theories*. Yes, we run experiments, take measurements, observe, but to spin the straw of data into the gold of science, we need to formulate a theory to make sense of what our evidence means.

Science might be the best method humanity has ever devised for making sense of the world around us, but that doesn't make it perfect. After all, the history of science is a history of failure. Countless clever minds have laboured under false paradigms, and saw their worlds through the lens of false hypotheses. Careful astronomers, taking conscientious measurements, were still fooled into thinking the sun revolves around the Earth.

Thinking through hypotheses provides a powerful way for scientists to make sense of the mysterious reality we inhabit. But even our best theories can turn out to be wrong, and we've no idea that we've been looking at the world through the wrong lens until a new paradigm comes along to sweep it away.

It turns out that this isn't just a nice metaphor. Neuroscience itself – including experiments in my own laboratory and studies by many others – is beginning to reveal that something like science really does unfold inside our brains. Just as we scientists take measurements from the world around us and concoct theories to explain them, so too does your brain sample the world around itself and conjure up theories to make sense of what its measurements

Copyrighted Material

mean. These theories then become your brain's paradigm – the filter through which everything else is understood.

If thinking is like science, we can begin to see how our minds end up with the same virtues and vices that the scientist does. The remarkable achievements of our minds become a bit more intelligible. Our brain's capacity to spot the patterns and regularities in the sea of data it swims in allows us to construct rich models of what makes our worlds and ourselves tick. Through these models, the organs sealed within our skulls make contact with reality outside.

But seeing the world in this way is not without risks. When our brains have formed the wrong theories about the world or about ourselves, we become prone to misperceive and misunderstand. Our grip on reality becomes tenuous.

Of course, reality isn't just one thing. The philosopher Karl Popper thought that we actually live in three worlds at once. The first world was the world of *matter* – the world of tissues, atoms, and molecules. The second was the world of *minds* – the world of other people and their mental states. And then there was the world of *ideas* – the products that these minds produce, like languages, mathematics, religion, myth, paradigms, and concepts that are bigger than any one human being, but no less real than the matter or the minds that make them possible.

Being 'in touch' with reality means being in touch with these three worlds simultaneously. And this book will reveal how the scientist that's your brain puts you in and out of touch with all three.

We begin in Part One with the material world, the raw physical stuff that surrounds us. Just like scientists, our bodies and brains are equipped with probes and sensors to measure the physical world, but also like scientists, our measurements of reality are meaningless without a theory to make sense of them. We'll see how even the mere act of seeing, hearing, or doing requires your brain – behind the scenes – to construct its own unconscious theory of the extracranial world. The theories the brain entertains allow us to perceive what's really there – and to hallucinate what isn't.

In Part Two we step up a level, and move into that mental

Copyrighted Material

world – the world of people, of minds, and of hidden mental states. To make sense of reality, scientists often have to go beneath the surface, theorising about forces like evolution or gravity, which animate and shape the world around us, but which can't be directly observed. In the same way, our brains must make sense of other minds – the beliefs, intentions, and desires of others – even though these thoughts and feelings can never be directly observed. We shall see how, just like scientists, our brains get beneath the surface by generating hypotheses about the things they cannot see – that is, what's happening in other people's heads. But this same hypothesis-building instinct also gets turned inwards. We build up an introspective theory of our own minds and who we are, which can give us an accurate – or inaccurate – picture of ourselves.

Finally, in Part Three we turn to that last plane of reality, the world of ideas. Just as scientists can think about their own theories, your brain can model its own models too. Here we'll see how our brains get in touch with that world of thought and possibility. We'll explore how embodied brutish beasts like us end up with a drive as profound and as useless as curiosity. We'll also see how it is possible for our brain's hypotheses-forming processes – recycling data from the past into theories about the present – make it possible for us to generate ideas that are genuinely new.

In the end, we'll unpack how our brain decides when its theories need to change. Just as scientists must be alert to when paradigms begin to shift, so too must our brain track when the tide is turning, and when old ideas need to be replaced with new ones. But while we want our brain's paradigms to change when the ground around us begins to shift, prematurely discarding theories about our surroundings and ourselves can leave our mind vulnerable, anxious, and at sea in an unstable and uncertain world.

The idea that your brain is a scientist is transforming the way that neuroscientists think. And this idea can transform how you think too. Meeting the scientist inside your skull will start to make *you* think like a scientist. You'll begin to see your mind and brain from the outside-in. From this new vantage point, things seem

Copyrighted Material

rather different. Some of the seemingly simplest things your mind achieves turn out to be the strangest. And at the same time, experiences and beliefs that seem odd to begin with can start to make sense when you realise your mind is your brain's best theory that you just happen to live inside.

So take a step outside yourself, and take a closer look at you.

Copyrighted Material

Copyrighted Material

PART ONE

The Material World

Copyrighted Material

Copyrighted Material

1.

Measuring Reality

Hearing gods and seeing devils

During a party at an English country house in the 1930s, a couple of Bright Young Things were playing table tennis when a clumsy misstep crushed the ball. As the group upturned the cupboards in Southgate House to find a replacement, they instead discovered a cache of small leatherbound books. Among them was a missing medieval manuscript, thought by some to be the first autobiography ever penned in the English language – *The Book of Margery Kempe.*

Margery Kempe was born in England in the fourteenth century, and her book describes the trials and tribulations of her life as a Christian mystic. Initially, Kempe seems to have lived a rather normal middle-class medieval life. She was the daughter of a merchant in Bishop's Lynn, and around the age of twenty married the 'honourable burgess' John Kempe. However, while pregnant with their first child, she suffered from sickness and fever. After a difficult birth, she began to fear for her life and called for a priest, describing unusual experiences of hearing the Devil in her head, telling her she would be damned for failing to confess her sins.

The priest came to hear Kempe's confession, but this didn't provide the absolution she'd hoped. For months she was tortured by visions of demons, with flames burning in their mouths, clawing at her, crying to her, and shouting threats. She was so tormented that she tried to end her own life – biting her hand until it bled and clawing her nails into her own skin, until eventually she had to be restrained.

But then Kempe was rescued. The demons were exorcised and

Copyrighted Material

the mental torture subsided. In her visions she was instead visited by Jesus Christ himself, beautiful, clad in purple silk, sitting at the bottom of her bed, telling her that she was not forsaken.

This episode had a profound effect on Kempe. She decided that she had to dedicate her life to God, interpreting these visions as a sign that she should live a chaste and holy life. Indeed, she suggested to her husband that these visitations from Christ may be a sign that they should abstain from 'the lusts of their bodies' to avoid displeasing God. Though John Kempe seemed generally encouraging of his wife's spiritual transformation, he was less keen on the prospect of celibacy. He suggested that he should wait for a sign from God too – just in case – before they took that plunge.

Margery seemed to win the argument, though, much to John's discontent. In a later chapter, she recounted a conversation with him one midsummer's eve:

> 'Margery, if a man came with a sword and wanted to chop off my head unless I had sexual intercourse with you as I used to before, tell me the truth from your conscience . . . [would you] allow my head to be chopped off, or else allow me to have sex with you as I previously did?'
>
> 'Alas, sir . . . why are you raising this matter? Haven't we been chaste these eight weeks?'
>
> 'Because I want to know your heart's truth.'
>
> 'Truthfully I'd rather see you slain.'

Perhaps you shouldn't ask questions if you can't handle the answers.

The rest of Kempe's book describes her journeys throughout England and Christendom, pursuing the spiritual mission these mystical experiences inspired. Indeed, throughout her life she described continuing to perceive the supernatural, for instance hearing blissful music sent from Heaven, or hearing God speak again at opportune moments on her journeys.

Kempe's unusual experiences persuaded her that she was on the right track, and persuaded others in her milieu that she was truly

Copyrighted Material

touched by the divine. But more recent readers of Kempe's book have been less convinced.

Scholars in the twentieth century supposed instead that Kempe's experiences reflected a clear case of historical madness. Although accurate retrospective diagnoses are practically impossible, some writers have claimed that Margery's experiences bear all the hallmarks of psychotic illness – perhaps postpartum psychosis shortly after childbirth. In this way of thinking, Margery was plainly not a conduit for the voice of God. She was just *hallucinating*.

In the traditional way of thinking, hallucinations represent a bright line between a sound mind and a sick one. There's a reason why we think seeing visions or hearing voices are signs of mental illness. In hallucinations, we lose our grip on reality, a stark contrast with our typical perceptual experiences, which we think put us directly in touch with the world around us.

But is this way of thinking *right*? Do our perceptions always put us directly in touch with the reality outside our heads? Or is the boundary between hallucination and perception fuzzier than it initially seems? How much are any of our brains really in touch with what's happening in the world outside our heads?

A brain in a vat?

Precisely this kind of worry is still exercising people centuries after Kempe. For instance, in 1973, Gilbert Harman pondered the very real possibility that the external world he could see, taste, and touch might not actually exist.

Harman thought it was possible that – in actuality – he could just be a brain floating in a vat. He'd reasoned that all our brain does is pick up signals from the outside world and turn them into signals that travel through circuits of neurons inside our head. If this were true, then perhaps a mad scientist could have excised Harman's brain from his skull, connected it up to a series of electrodes, and stimulated it in a particular way to *fool him* into thinking this

Copyrighted Material

familiar external world was really there. This illusion would seem just like real life, but really he'd be floating in a jar on the shelf of some malevolent genius, none the wiser.

If a stranger on the bus disclosed this theory to you in passing, you might be worried about their state of mind. You may at least get up to change your seat. But because Harman was a philosopher, writing about this kind of thing was enough to earn him a tenured post as a professor at Princeton.*

Now, you may not be a brain floating in a vat. But there is an important kernel of truth behind Harman's worry. There is a very real sense in which our access to the world around us is indirect, not because your brain is suspended in jelly on a shelf in some strange lab, but because your brain is stuck somewhere else: *inside your head*.

For your entire life, your brain sits firmly sealed inside your skull, cut off from the rest of the material world. And yet it has to somehow conjure up a picture of what's going on outside.

Luckily for you, this brain of yours isn't completely disconnected from the outside world. It's equipped with a host of tools and instruments that help it to sense and measure the signals emanating from its surroundings. Eyes that can pick up those electromagnetic waves we call light. Your ears, your skin. The porous surfaces that line the top of your tongue and inside your nose – your taste and smell. All

---

* It's not actually too clear who should get the credit for the 'brain in a vat' thought experiment. The philosopher Hilary Putnam offers a very vivid description of the idea in a piece entitled 'Brains in a Vat', which appears in his 1981 book *Reason, Truth and History*. Here, though, I have given Harman the credit because he mentions a similar idea in his 1973 book *Thought*. It seems plausible to me that the idea of a meddlesome scientist interfering with our brains could have been circulating around philosophical circles for a while in these years without anybody putting it down in print. If we are being purists, we could say this idea really originated with René Descartes' 1641 'evil demon' thought experiment, in which he postulated that we do not know if we live in the real world, or in an elaborate illusion constructed by a cunning, all-powerful demon who is intent on deceiving us. For better or worse, demons have largely fallen out of favour as explanatory devices in modern psychology.

Copyrighted Material

of these intricate organs taking measurements of vibrations in the air, of pressure on our bodies, of gravity, and of the chemical gradients diffusing into and around us.

All these measurements are sent along to the laboratory of that skull-bound scientist, who sits ready to begin the process of interpretation and analysis we call *perception*.

Like a scientist, your brain is in touch with physical reality, but is one step removed – access comes only through the veil of its measurements. And the problem with seeing the world through measurements is that measurements can be ambiguous.

Casting shadows on our senses

Take vision, for example. Imagine looking at something familiar, like the face of your dearest friend. When you glance at them, you construct a picture of what they look like from the light bouncing off their features – the line of their jaw, the arch of their eyebrow, maybe a stray wrinkle or two. The light radiating off of them is all the input your eyes receive, the raw material your sensing organs gather up to sketch this picture of what they look like.

But even in this innocuous example, there is a thorny problem. Your friend's face is a three-dimensional object, but the surface of the retina in your eye – the bit that captures incoming light – is two-dimensional. It's flat. This means that when you stare adoringly at your friend, you don't directly see them: you see the two-dimensional shadow that their face casts upon the surface of your eyes. And these shadows are fundamentally ambiguous.

Think about when you play shadow puppets with a small child – contorting your hands in front of light to create the illusion of a howling wolf or a fluttering bird, projected on a wall. When shadow puppetry works well, it works because of *confusability*. A master puppeteer can make it look like there is the shadow of a wild animal creeping across the wall, even though in reality the shadow is being cast by a clever arrangement of fingers and thumbs.

Copyrighted Material

The same kind of confusability afflicts your brain too. You can never see the true sources behind the signals you are sampling, like the fingers over the lampshade. All you can see are the shadows. And your brain has to do its best, from those shadows alone, to work out what the sources truly were.

The problem with seeing through shadows like this is that it's impossible to get from the impoverished image reaching the eye back to the source that created it with any ironclad certainty. Vision is what engineers call an 'ill-posed inverse problem'. There are infinitely many physical objects that can all cast exactly the same shadow on our senses. This many-to-one mapping means that we can't know from just the shadow alone what the real object is actually like.

For instance, if your friend's face doubled in size, but they stood twice as far away from you, this would lead to exactly the same image landing on your eyes as their normal face would. Likewise, if their head had suddenly shrunk but they stepped a little bit closer, or for any other combination of distances, sizes, and angles you can think of.

But while, strictly speaking, the shadow of your friend's face landing on your eyes is completely ambiguous, our brains do seem able to solve this inverse problem. You don't have the sense, while you look at your friend over dinner, that their head is rapidly changing in shape and size, as your brain tries on various interpretations of this ambiguous visual image. When you look at your friend, you see the same face every time, not one of the infinitely many distorted possibilities that would all cast the same shadow on your senses. How, then, does your brain pick out the right interpretation from the limitless contenders? How does it make sense of the measurements it's taking?

Seeing like a scientist

Scientists – and people thinking about science – have known for a long time that raw measurements can be impossible to decode on

Copyrighted Material

their own. To see the signal through the noise, we need to learn how to perceive what the measurements mean.

Thomas Kuhn argued that scientists are only ever able to make sense of what their measurements mean because they have been inculcated into a paradigm that tells them what they mean. Learning this correspondence between the signals and their sources is a key part of how one sees like a scientist in the first place. Or, as Kuhn put it:

> Looking at a contour map, the student sees lines on paper, the cartographer a picture of a terrain. Looking at a bubble-chamber photograph, the student sees confused and broken lines, the physicist a record of familiar subnuclear events. Only after a number of such transformations of vision does the student become an inhabitant of the scientist's world, seeing what the scientist sees and responding as the scientist does.

Your brain deals with the confused and broken lines of its own measurements in exactly the same way. Your brain can come to see the sources that lie behind the sensory shadows by forming its own paradigm and concocting its own theories about what it thinks the world is like.

Unconscious inference

The idea that perception works like this has been around since at least the nineteenth century, and is often credited to the German polymath Hermann von Helmholtz. Helmholtz was the kind of renaissance man that makes modern scientists feel a bit inadequate. He made significant contributions to thinking about electromagnetism, fluid dynamics, and the eventual heat death of the universe. And around all this, he found time to pen a volume called *A Treatise on Physiological Optics*, which continues to have an enduring impact on how scientists think about the perceiving brain.

Copyrighted Material

Helmholtz realised that the measurements of the outside world we get from our senses are hopelessly degraded and ambiguous, saying of the eye, 'If an optician wanted to sell me an instrument which had all these defects, I should think myself quite justified in blaming his carelessness in the strongest terms.'

He thought our brains must have a solution to these defects – a solution that he called unconscious inference. His idea was that our visual system overcomes the ambiguities in the raw bottom-up visual image by adding some top-down knowledge into the mix, tacit assumptions the visual system makes about which configurations of objects are actually likely to be out there in the world around us. These bundles of assumptions become a kind of unconscious theory maintained in the brain about how the visual world works, which allows the visual system to make educated guesses about what's happening in the world around us.

In this way of thinking, the reason why, when you look at your friend, you don't see their head as being twice as big or twice as small is because your visual system already holds some assumptions about what your friend's face will look like. And these hypotheses – buried in parts of your mind beneath the level of conscious awareness – guide how your perceptual systems interpret the ambiguous patterns of sensory data that land upon your senses.

Helmholtz's ideas resurfaced in the 1970s, particularly in the work of the British psychologist Richard Gregory. Gregory drew an explicit analogy between the way that scientists generate hypotheses to understand the puzzling signals generated by their instruments, and the way that our perceptual systems generate hypotheses to make sense of our sensations. The key similarity that Gregory saw was that hypotheses in both science and perception allow us to fill in the gaps in the incomplete data that we have sampled.

Just like Helmholtz, Gregory was eager to stress that the hypotheses entertained by your perceptual system don't have to be conscious propositions – explicit thoughts, like sentences in your head, describing what you think the world is like. He speculated that there might one day be an alternative, non-propositional way

Copyrighted Material

of describing how our brains build hypotheses, rooted in concepts from mathematics and computer science. And it turns out he was right.

The Bayesian brain

Nowadays, psychologists and neuroscientists have turned to mathematical ideas to try to understand how our brains form hypotheses and compute their inferences. One transformational idea in modern neuroscience is that our brains are Bayesian.

This moniker honours the distant grandfather of probability theory, Thomas Bayes. The Reverend Bayes was an eighteenth-century scholar with a set of interests that might seem unusual for a priest. He was obsessed with understanding games of chance like coin tosses and dice rolls and evaluating the odds for different kinds of outcomes. Little surprise, then, that he developed a set of mathematical precepts that allow us to quantitatively derive the probability that certain events will or won't happen.

Bayes is famous for the eponymous Bayes theorem, which tells us that when we make inferences about the world around us, we shouldn't just rely on incoming evidence alone. Instead, every piece of data we encounter should be evaluated against our background knowledge about what is likely to be true in the first place – our beliefs about *prior probabilities*.

Our standard way of thinking about rationality can make this seem very counterintuitive: surely thinking clearly means focusing on the evidence at hand rather than leaning on beliefs you already hold? But a quick moment of reflection can reveal the virtue of a probabilistic way of thinking.

Imagine you find yourself staring at the stars late one night. Suddenly you see what looks like a flying saucer dart through the sky – there one moment, gone the next. What should you believe about what you have just witnessed? Based on the raw data alone, it looks like you may have just had a close encounter with

Copyrighted Material

extraterrestrial life. However, the raw data isn't all you have to go on. You might know, for example, that a new satellite was due to go into orbit this very evening and that it might appear in your slice of the starscape tonight. Alternatively, you might remember that the annoying kid next door got an aerial drone for Christmas and likes to take it for a spin after dark. These background possibilities make it less likely that you are really seeing an alien traverse the night sky. From a Bayesian perspective, your inferences *should* be guided by what is most likely. There's no need to call NASA just yet.

Bayes and the Bayesians that succeeded him were not, for the most part, directly interested in the human mind. The laws of probability theory are normative rather than descriptive: they tell us how we *ought* to think, rather than necessarily setting out to describe how our minds actually work. But one of the most tantalising ideas in neuroscience today is that the brain actually *is* structured in a way that causes it to implement or approximate the kinds of Bayesian inferences that mathematicians extol, interpreting all the raw incoming data through a hypothesis it holds about how the world is likely to be.

One of the principal champions of this idea is Karl Friston. Friston's model of the brain points out an underappreciated feature of our neural circuitry: information doesn't just flow 'forwards' in one direction through our heads – from simple sensory analysis up to more abstract thought. It also flows 'backwards' – from higher brain regions back down to lower ones.

This kind of architecture leaves us with brains that can behave like scientists do. Broad hypotheses about the world, held in higher levels of the brain, can be projected back down to lower levels of the brain. These back-projected hypotheses – conveying our prior theories and assumptions – can then go on to shape how we interpret the uncertain and ambiguous data arriving at our senses. Our perceptions become Bayesian inferences – confluences of incoming evidence and prior beliefs. We see the confused and broken lines of our measurements through the lens of the theory our brain has already constructed.

Copyrighted Material