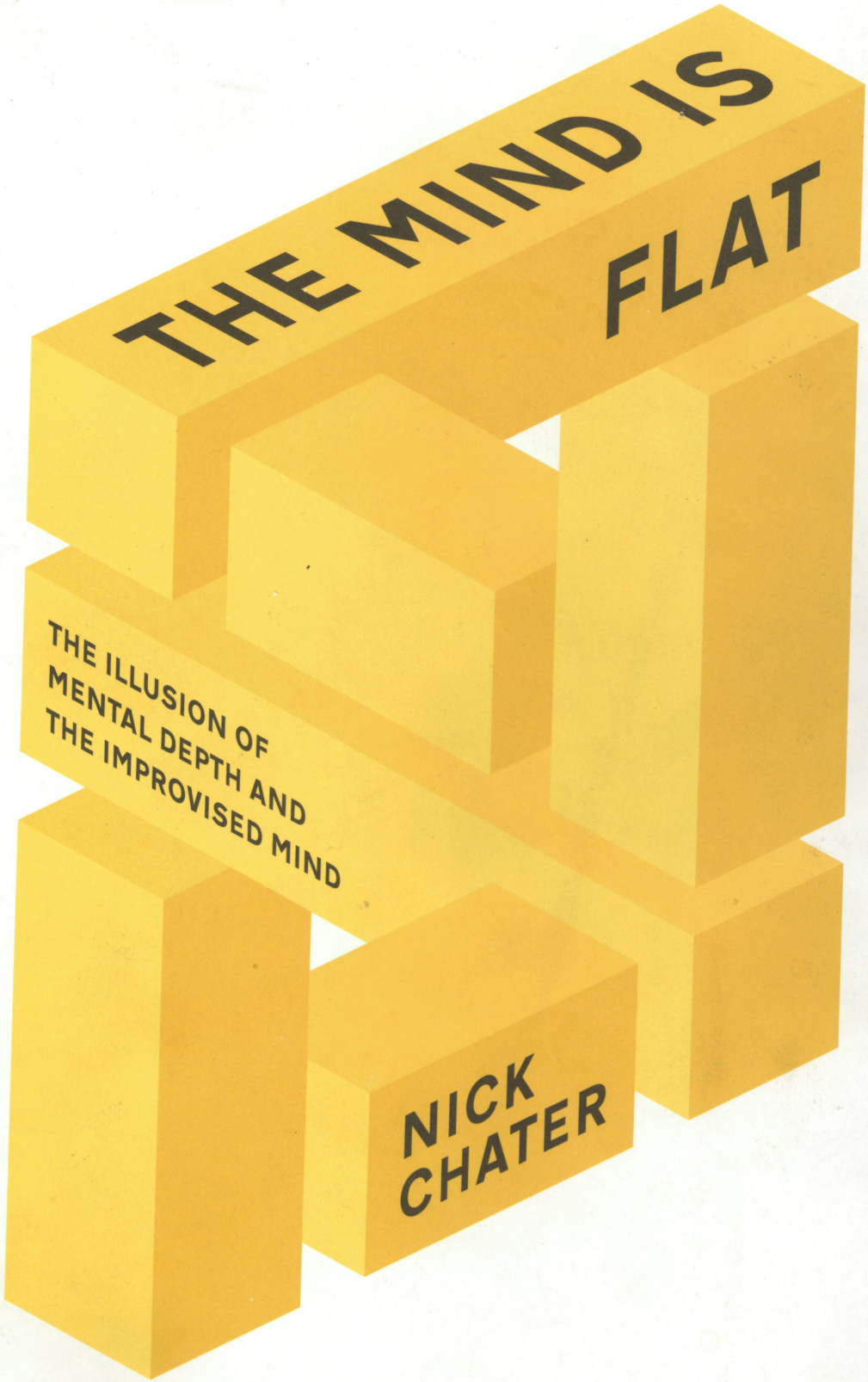


THE ILLUSION OF MENTAL DEPTH AND THE IMPROVISED MIND

NICK CHATER

allen lane

A radical reinterpretation of how your mind works – and why it could change your life

We all like to think we have a hidden inner life. Most of us assume that our beliefs and desires arise from the murky depths of our minds, and if only we could work out how to access this mysterious world we could truly understand ourselves. For more than a century, psychologists and psychiatrists have struggled to discover what lies below the mental surface.

In *The Mind is Flat*, pre-eminent behavioural scientist Nick Chater reveals that this entire enterprise is utterly misguided. Drawing on startling new research in neuroscience, behavioural psychology and perception, he shows that we have no hidden depths to plumb, and unconscious thought is a myth. Instead we generate our ideas, motives and thoughts in the moment. This revelation explains many of the quirks of human behaviour – for example why our supposedly firm political beliefs, personal preferences and even our romantic attractions are routinely proven to be inconsistent and changeable.

As the reader discovers, through mind-bending visual examples and counterintuitive experiments, we are all characters of our own creation, constantly improvising our behaviour based on our past experiences. And, as Chater shows us, recognizing this can be liberating.

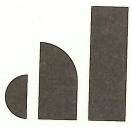
A dazzlingly original and delightfully provocative book, *The Mind is Flat* forces us to reconsider just about everything we thought we knew about ourselves.

£25.00

NICK CHATER

The Mind is Flat

The Illusion of Mental Depth and the Improvised Mind



ALLEN LANE
an imprint of
PENGUIN BOOKS

ALLEN LANE

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

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

Penguin
Random House
UK

First published 2018
001

Copyright © Nick Chater, 2018

The moral right of the author has been asserted

Set in 10.5/14 pt Sabon LT Std
Typeset by Jouve (UK), Milton Keynes
Printed in Great Britain by Clays Ltd, St Ives plc

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

ISBN: 978–0–241–20844–1

www.greenpenguin.co.uk

MIX
Paper from
responsible sources
FSC® C018179
FSC
www.fsc.org

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.

To my mother and fathe
wonderful wife and ch
whose unfailing suppo

To my mother and father, Robert and Dorothy Chater, and to my wonderful wife and children, Louie, Maya and Caitlin Fooks, whose unfailing support and love has made this book possible.

Contents

Prologue: Literary Depth, Mental Shallows 1

PART ONE
The Illusion of Mental Depth

1 The Power of Invention 21
2 The Feeling of Reality 37
3 Anatomy of a Hoax 53
4 The Inconstant Imagination 73
5 Inventing Feelings 91
6 Manufacturing Choice 109

PART TWO
The Improvised Mind

7 The Cycle of Thought 127
8 The Narrow Channel of Consciousness 146
9 The Myth of Unconscious Thought 160
10 The Boundary of Consciousness 175
11 Precedent not Principles 192
12 The Secret of Intelligence 205
Epilogue: Reinventing Ourselves 220

Notes 225
Index 245

Prologue: Literary Depth, Mental Shallows

> . . . when we claim to be just using our powers of inner observation, we are always actually engaging in a sort of impromptu theorizing – and we are remarkably gullible theorizers, precisely because there is so little to 'observe' and so much to pontificate about without fear of contradiction.
>
> *Daniel Dennett*[1]

At the climax of *Anna Karenina*, the heroine throws herself under a train as it moves out of a station on the edge of Moscow. But did she want to die? Various interpretations of this crucial moment in Tolstoy's great masterpiece are possible. Had the ennui of Russian aristocratic life and the fear of losing her lover Vronsky become so intolerable that death seemed the only escape? Or was her final act mere capriciousness, a theatrical gesture of despair, not seriously imagined even moments before the opportunity arose?

We ask such questions. But can they possibly have answers? If Tolstoy says that Anna has dark hair, then Anna has dark hair. But if Tolstoy doesn't tell us why Anna jumped to her death, then Anna's motives are surely a void. We can attempt to fill this void with our own interpretations and we can debate their plausibility. But there is no hidden truth about what Anna really wanted, because, of course, Anna is a fictional character.

Suppose instead that Anna were a historical figure and Tolstoy's masterpiece a journalistic reconstruction of real events. Now the question of Anna's motivation becomes a matter of history, rather than literary interpretation. Yet our method of inquiry remains the

same: the very same text would now be viewed as providing (perhaps unreliable) clues about the mental state of a real person, not a fictional character. Lawyers, journalists and historians, rather than critics and literary scholars, might put forward and debate competing interpretations.

Now imagine that we ask Anna herself. Suppose that Tolstoy's novel was indeed an account of real events, but the great steam engine slammed on its brakes just in time. Anna, apparently mortally injured, is conveyed anonymously to a Moscow hospital. Against the odds, she pulls through and chooses to disappear to escape her past. We catch up with Anna convalescing in a Swiss sanatorium. As likely as not, Anna will be as unsure as anyone else about her true motivations. After all, she too has to engage in a process of interpretation: considering her memories (rather than Tolstoy's manuscript), she attempts to piece together an account of her behaviour. Even if Anna does venture a definitive account of her actions, we may be sceptical that her own interpretation is any more compelling than the interpretations of others. To be sure, she may have 'data' unavailable to an outsider – she may, for example, remember the despairing words 'Vronsky has left me forever' running through her mind as she approached the edge of the fateful platform. However, any such advantage may be more than outweighed by the distorting lens of self-perception. Our interpretations of our own actions seem, among other things, to assign to ourselves greater wisdom and nobility than might be evident to the dispassionate observer. Autobiography always deserves a measure of scepticism.

Might we get closer to Anna's true motivations if we asked her not in hindsight, but at the time? Suppose a journalist, employed by a scandal-hungry Moscow paper, is following Anna's every move, scenting a story. He leaps to save Anna from falling to her death, only to brandish pen and notepad, with the words, 'Now, Ms Karenina, tell me a little about why you decided to launch yourself into oblivion.' This strategy seems unlikely to succeed, to put it mildly. And a discreet query of, 'Ms Karenina, I perceive you to be about to leap to your death. Perhaps you could spare a couple of moments to answer this short questionnaire?' seems equally doomed to fail.

There are two opposing conclusions that one might draw from this

vignette. The first possib
unfathomable 'hidden dep
people reliably to be able t
plete and true account o
behaviour, whether from
during or after the event,

From the 'hidden dep
motivations for human
of asking people directly
need somehow to dive de
to measure directly the
cions and hopes govern
ourselves be only dimly
roscientists have long de
human motivation. Wor
hours of intensive psych
logical recordings and b
over the last century or
clear: to discover the fee
mental 'surface' of cons
den depths. Yet the cor
perpetually elusive. Freu
hidden fears and desi
attempt to draw subtle
heart-rate, skin conduc
flow in the brain. But
ever actually observed.
that is as, or more, mys
penetrate it we will nee
ods of analysis. From th
hidden depths of the h
efforts.

In this book, I will ar
the project of charting
difficult, but fundame
minds contain 'hidden
Anna Karenina's fatefu

vignette. The first possible moral is that our minds have dark and unfathomable 'hidden depths'. From this viewpoint, we cannot expect people reliably to be able to look within themselves and compile a complete and true account of their beliefs and motives. Explanations of behaviour, whether from observers or participants, and whether before, during or after the event, are partial and unreliable at best.

From the 'hidden depths' standpoint, then, uncovering the true motivations for human behaviour needs more than the blunderbuss of asking people directly. It requires delicacy and sophistication. We need somehow to dive deep into the inner workings of the mind, and to measure directly the hidden beliefs, desires, motives, fears, suspicions and hopes governing our actions, forces of which we may ourselves be only dimly aware. Psychologists, psychiatrists and neuroscientists have long debated how best to plumb the deep waters of human motivation. Word associations, the interpretation of dreams, hours of intensive psychotherapy, behavioural experiments, physiological recordings and brain imaging have all been popular options over the last century or so. Whatever the method, the objective is clear: to discover the feelings, motives and beliefs that lurk below the mental 'surface' of conscious awareness – to chart, in short, our hidden depths. Yet the contents of our hidden depths seem to remain perpetually elusive. Freudian psychoanalysts can speculate about our hidden fears and desires; psychologists and neuroscientists can attempt to draw subtle and highly indirect conclusions from actions, heart-rate, skin conductance, pupil dilation and the rate of blood flow in the brain. But no hidden beliefs, desires, hopes or fears are ever actually observed. Perhaps our hidden depths are an 'inner space' that is as, or more, mysterious than the depths of outer space; and to penetrate it we will need more sophisticated instruments and methods of analysis. From this viewpoint, our failure so far to reveal the hidden depths of the human mind should drive us to even greater efforts.

In this book, I will argue for precisely the opposite viewpoint: that the project of charting our hidden depths is not merely technically difficult, but fundamentally misconceived; the very idea that our minds contain 'hidden depths' is utterly wrong. Our reflections on Anna Karenina's fateful act should point us, instead, to a radically

different moral: that the interpretation of the motives of real people is no different from the interpretations of fictional characters. And fictional characters can't, of course, have an inner life, because they have no life at all: there is no more a fact about whether the fictional Anna has a subconscious fear of dogs, doubts about the stability of the Tsarist regime, or prefers Bach to Mozart, than there is a fact about whether she was born on a Tuesday. There is no 'hidden' truth about fictional characters, nothing beneath the 'surface' of the words on the page.

But the picture would be largely the same had Tolstoy's novel been reportage, and Anna a living, breathing member of the nineteenth-century Russian aristocracy. Then, of course, there would be a truth (whether known or not) about whether Anna was born on a Tuesday. But, I argue, there would still be no truths about the real Anna's motivations, any more than for the fictional Anna. No amount of therapy, dream analysis, word association, experiment or brain-scanning can recover a person's 'true motives', not because they are *difficult* to find, but because there is *nothing* to find. It is not hard to plumb our mental depths because they are so deep and so murky, but because there are no mental depths to plumb.

The inner, mental world, and the beliefs, motives and fears it is supposed to contain is, itself, a work of the imagination. We invent interpretations of ourselves and other people in the flow of experience, just as we conjure up interpretations of fictional characters from a flow of written text. Each possibility can be challenged with endless alternatives. Perhaps Anna primarily despaired of her precipitous social fall, or the future of her son, or the meaninglessness of aristocratic life, rather than being tormented by love. There is no ground truth about the right interpretation of the fictional Anna, though some are more compelling and better evidenced in Tolstoy's text than others. But Tolstoy the journalist would have nothing more than interpretations of a 'real' Anna's behaviour; and the real Anna, whether at the time, or looking back months later, could do no more than venture one more interpretation of her own behaviour. We never have the last word in explaining our own actions – our interpretations of our own actions are as partial, jumbled and open to challenge as anyone else's.

The flow of Tolstoy's te
given sketchy reports fro
to tell us about Anna's ch
Seryozha, or how Vronsk
a monk. These scenes wo
ten. As he writes, Tolstoy
her circle; he is not *disco*

Yet the unfolding of a
a novel. We generate our
they are not pre-calculate
memory store just in case
there is no pre-existing
thoughts issue. Thought
instant that they are inve

The very idea of 'looki
take: we talk as if we have
contents of our inner wor
inform us about the exte
not of *perception* but of *i*
pretations and explanati
actions. The inner world

Now in fiction, while
others seem to have real
imagination a vividness t
of our living acquaintanc
beliefs beyond the printed
'in the eye of the beholde
life, save what Tolstoy gi
the lines. As for fictional
behaviour is merely the s
teeming with inner motiv
barely sense is a conjurin
is not that the depths are
is all there is.

As we have already see
psychologists – to suspe

The flow of Tolstoy's text can give us the feeling that we are being given sketchy reports from another world. Tolstoy could have chosen to tell us about Anna's childhood, the impact of her death on her son Seryozha, or how Vronsky (perhaps) retired from society and became a monk. These scenes would spring into being only as they were written. As he writes, Tolstoy is *inventing* more of the lives of Anna and her circle; he is not *discovering* more about them.

Yet the unfolding of a life is not so different from the unfolding of a novel. We generate our beliefs, values and actions in-the-moment; they are not pre-calculated and 'written' in some unimaginably vast memory store just in case they might be needed. And this implies that there is no pre-existing 'inner world of thought' from which our thoughts issue. Thoughts, like fiction, come into existence in the instant that they are invented, and not a moment before.

The very idea of 'looking' into our own minds embodies the mistake: we talk as if we have a faculty of introspection, to scrutinize the contents of our inner world, just as we have faculties of perception, to inform us about the external world.[2] But introspection is a process not of *perception* but of *invention*: the real-time generation of interpretations and explanations to make sense of our own words and actions. The inner world is a mirage.

Now in fiction, while some characters are 'two-dimensional', others seem to have real 'depth'. They can, indeed, assume in our imagination a vividness that may equal, or even exceed, that of some of our living acquaintances. We may attribute to them attitudes and beliefs beyond the printed page. Yet such apparent depth is, of course, 'in the eye of the beholder': there are no facts about Anna Karenina's life, save what Tolstoy gives us; no hidden motives lurking between the lines. As for fictional characters, so for real people. The sense that behaviour is merely the surface of a vast sea, immeasurably deep and teeming with inner motives, beliefs and desires whose power we can barely sense is a conjuring trick played by our own minds. The truth is not that the depths are empty, or even shallow, but that the surface is all there is.

As we have already seen, it is awfully tempting – especially for psychologists – to suspect that while our everyday, common-sense

explanations of ourselves and each other, as guided by beliefs, desires, hopes and fears, may be wrong in detail, they are right in spirit. Anna's leap to her death is, one may think, guided by some beliefs, desires, hopes and fears, even though she may not be able to tell us quite which beliefs, desires, hopes and fears. Her introspection is imperfect or perhaps untrustworthy. But the problem with our everyday view of our minds is far deeper: no one, at any point in human history, has ever been guided by inner beliefs or desires, any more than any human being has been possessed by evil spirits or watched over by a guardian angel. Beliefs, motives and other imagined inhabitants of our 'inner world' are entirely a figment of our imaginations. The stories we tell to justify and explain our own and others' behaviour aren't just wrong in detail – they are a thoroughgoing fabrication from start to finish.

Our flow of conscious thought, including our explanations of our own and each other's behaviour, are creations in the moment, not reports of (or even speculations about) a chain of inner mental events. Our mind is continually interpreting, justifying and making sense of our own behaviour, just as we make sense of the behaviour of the people around us, or characters in fiction. If you cross-examine me, or any other reader, about Anna's motivations (Q: 'Did she think that jumping under the train would mean certain death?' A: 'Yes.' Q: 'Did she believe that Seryozha would be better off without her?' A: 'Possibly – though almost certainly quite wrongly'; and so on), I will generate answers, as quick as a flash. So we clearly have the *ability* to fabricate justifications at will, but these justifications cannot, of course, be conjectures about Anna's mental life, because Anna, being a fictional character, *has* no mental life.

If Anna were real, and had survived, we could cross-examine her, in her Swiss sanatorium, with just the same questions, and she too could reply, quick as a flash. And, for that matter, were you to cross-examine me about some prosaic aspect of my own life (e.g. why I took the train, rather than driving, to London), I can come back with a string of explanations (about carbon emissions, traffic congestion, parking etc.). The sheer inventiveness of our minds implies that the real Anna could be interpreting and justifying her own thoughts and behaviour, in retrospect, using exactly the imaginative powers that we are using when considering her as a fictional character (extended

by the imaginative work
this suggests that this
stream of justifications
ourselves and each othe

In this book, I want t
very idea of mental dep
summate improviser, i
explain those actions, w
inventions are flimsy, fra
a film set, seeming solic
structed from cardboar

An improvising mind
might seem to be a rec
opposite is true: the very
thoughts and behaviour
ter' as well as we are abl
to think and act in the c
possible our prior thoug
each new legal case by
growing body of previou
supposed hidden depths
improvise our present, c

The argument – and th
to clear away what I tak
derstandings of how our
account of the brain as
Part One, we will expl
beliefs, desires, hopes ar
ing fiction, and so effortl
that we take it for realit
we know about our minc
ology textbook. Accordi
the right track, but just
But these modifications
common-sense mind an
just don't seem to fit tog
abandoned, not patched

by the imaginative work of Tolstoy himself in creating her story). And this suggests that this very same inventiveness could underlie the stream of justifications we provide to explain our everyday lives to ourselves and each other.

In this book, I want to convince you that the mind is flat: that the very idea of mental depth is an illusion. The mind is, instead, a consummate improviser, inventing actions, and beliefs and desires to explain those actions, with wonderful fluidity. But these momentary inventions are flimsy, fragmented and self-contradictory; they are like a film set, seeming solid when viewed through the camera, but constructed from cardboard.

An improvising mind, unmoored from stable beliefs and desires, might seem to be a recipe for mental chaos. I shall argue that the opposite is true: the very task of our improvising mind is to make our thoughts and behaviour as coherent as possible – to stay 'in character' as well as we are able. To do so, our brains must strive continually to think and act in the current moment in a way that aligns as well as possible our prior thoughts and actions. We are like judges deciding each new legal case by referring to, and reinterpreting, an ever-growing body of previous cases. So the secret of our minds lies not in supposed hidden depths, but in our remarkable ability to creatively improvise our present, on the theme of our past.

The argument – and this book – has two parts. In the first, I attempt to clear away what I take to be fundamental and widespread misunderstandings of how our mind works. I then turn to present a positive account of the brain as a ceaseless improviser. More concretely, in Part One, we will explore the psychological evidence that talk of beliefs, desires, hopes and fears is pure fiction. Yet it is such convincing fiction, and so effortlessly and fluently invented by our own brains, that we take it for reality. We'll find that almost everything we think we know about our minds is false. This is not the story from the psychology textbook. According to that story, common sense is roughly on the right track, but just needs to be modified, adjusted and filled out. But these modifications and adjustments never seem to work. The common-sense mind and the mind we discover through experiment just don't seem to fit together. The common-sense story needs to be abandoned, not patched up.

Yet while the textbooks don't take a radical line, a growing number of philosophers, psychologists and neuroscientists do.[3] In Part One, I will point to the root cause of the problem with the common-sense view of the mind: that mental depth is an illusion.

We imagine that mirroring the outer world of people, objects, stars and noises there is an inner world of rich sensory experiences (the subjective experience of people, objects, stars and noises), not to mention our emotions, preferences, motives, hopes, fears, memories and beliefs. The possibilities for exploration in this inner world seem vast.[4] Just paying close attention to what we see and hear, and the states of our bodies, seems to reveal that our inner perceptual world is wonderfully rich; and that we just need to step off from direct sensory experience into the realm of the imagination that dreams, meditations and hypnosis seem to provide. Or we can explore the vast archives of our memories, perhaps reliving fragments of childhood or student life; or we can discourse with ourselves endlessly about our beliefs and values.

There are many who suspect that the scale of our inner world is far greater still – that we should add into the mix subliminal perception, which slips into our minds without our noticing; that we have unconscious beliefs, motives, desires and perhaps even unconscious inner agents (for Freud, the id, ego, and superego; for Jung, the collective unconscious). And perhaps there is a self, or many selves, or a soul. Many believe that with the right meditative practice, psychotherapy or even hallucinogenic drug, the doors to the rich inner world of the unconscious might be prised open. And, turning to neuroscience, it is natural to imagine that the contents of our inner world might one day be accessible to brain-scanners – which might 'read off' our beliefs, motives and feelings, whether conscious or not.

But all of this depth, richness and endless scope for exploration is utterly fake. There is no inner world. Our flow of momentary conscious experience is not the sparkling surface of a vast sea of thought – it is *all there is*. And, as we shall see, each momentary experience turns out to be startlingly sketchy – at any moment, we can recognize just one face, or read just one word, or identify just one object. And when, like our imagined Anna, rehabilitating in the Alps, we begin to describe our feelings, or explain our actions, we are only creating stories, one

step at a time, not explori
feelings. The more outré '
induced states, are similarl
of the imagination, not v
ation of dreams, far from
one creative act set atop a

The aim of Part One is
nature of our own minds
been repeated and even a
ology, psychoanalysis, ar
the intuitive picture – of
scious thought is merely
the obvious question is: v
thought and behaviour c

In Part Two, we take u
mental lives must exist pu
improviser, and it base
improvisations: it creates
by drawing not on a hid
motives, but on *memory*
experiences.

The analogy with fictio
words and actions as he w
words and actions as cohe
ter' or her character shou
we interpret the behaviou
aim applies: a good interp
of the present moment, bu
indeed interpretations. Ou
conscious interpretations
by linking the present wi
linking its sentences togeth
world.

Conscious experience
cycle of thought, locking
the sensory world. That is
interpretations of the wo

step at a time, not exploring a pre-existing inner world of thoughts and feelings. The more outré 'inner worlds' of dreams, or mystical or drug-induced states, are similarly nothing more than streams of invention – acts of the imagination, not voyages of inner discovery. And the interpretation of dreams, far from boring deep into our psyche, is no more than one creative act set atop another.

The aim of Part One is to help reinterpret our intuitions about the nature of our own minds, and to undercut misconceptions that have been repeated and even amplified in many areas of philosophy, psychology, psychoanalysis, artificial intelligence and neuroscience. But if the intuitive picture – of a rich and deep 'inner sea' of which our conscious thought is merely the glittering surface – is so utterly wrong, the obvious question is: what possible alternative story about human thought and behaviour could there be?

In Part Two, we take up this question. If the mind is flat, then our mental lives must exist purely at the 'mental surface'. Our brain is an improviser, and it bases its current improvisations on previous improvisations: it creates new momentary thoughts and experiences by drawing not on a hidden inner world of knowledge, beliefs and motives, but on *memory traces of previous momentary thoughts and experiences*.

The analogy with fiction is helpful here too. Tolstoy invents Anna's words and actions as he writes the novel. But he strives to make Anna's words and actions as coherent as possible – she should 'stay in character' or her character should 'develop' as the novel unfolds. And when we interpret the behaviour of other people, and of ourselves, the same aim applies: a good interpretation is one that does not just make sense of the present moment, but links it with our past actions, words, and indeed interpretations. Our brain is an engine that creates momentary conscious interpretations not by drawing on hidden inner depths, but by linking the present with the past, just as writing a novel involves linking its sentences together coherently, rather than creating an entire world.

Conscious experience is therefore the sequence of *outputs* of a cycle of thought, locking onto, and imposing meaning on, aspects of the sensory world. That is, we consciously experience the meaningful interpretations of the world that our brain creates, seeing words,

objects and faces, and hearing voices, tunes or sirens. But we are never conscious either of the *inputs* to each mental step or each step's *internal workings*. So we can report nothing to explain *why* we see an outcrop of rock as a pack of dogs, a fleeting facial expression as condescending or kindly, or why a line of poetry conjures up a vision of mortality or reminds us of childhood. Each cycle of thought delivers a consciously experienced interpretation, but no explanation of where that interpretation comes from.

Throughout this book, my argument will be illustrated through examples from visual perception. Perceptual examples are vivid and concrete; and perception is the area of psychology and neuroscience that is by far the best understood. It makes sense to focus our attention where the evidence is clearest and most easily explained. But I have another reason for focusing on perception, namely that the whole of thought, whether chess-playing, abstract mathematical reasoning, or artistic and literary creation, is really no more than an extension of perception.

We'll see how the cycle of thought works, and point to some key pieces of evidence in support of this story. On closer analysis, the stories we tell about our stable personalities, beliefs and motives can't possibly be right. By contrast, the quirks, variability and capriciousness of human nature make sense when we realize that our brain is an incomparable improviser: an engine for spontaneously finding meaning and choosing actions that make the best sense in the moment. Thus our thoughts and actions are based on a rich tradition of past thoughts and actions, which our brain harnesses and reworks to address the challenges of the moment. Moreover, just as today's thoughts follow yesterday's precedents, so they also set precedents for tomorrow – giving our actions, words and our life a coherent shape. So what makes each of us special is, to a large extent, the uniqueness of our individual history of prior thoughts and experiences. In other words, each of us is a unique tradition, continually in the process of creation.

As the story unfolds, we will see that we are, in a very real sense, characters of our own creation, rather than playthings of unconscious currents within us. And while each new perception, movement and thought builds on our unique mental tradition of past perceptions,

movements and thought
old in a wonderfully fre
become locked into pat
our human intelligence
thoughts from old. And
on rare seeds of genius
fundamental to the bas
dream, talk.

Our freedom has its l
'freely' choose to play
can't spontaneously em
spontaneously reason li
thoughts require buildin
is no shortcut to the thou
on which expertise is l
unique: with those thou
ferent traces of thought
and actions are created.
own way – though some
and poets can 'imperso
students can and do lea
same points arise in our
sometimes bumpy intera
sists not in the ability n
jump, but to reshape our
our current thoughts a
programming our minds

The ideas in this book ar
ing from cognitive, socia
neuroscience. Perhaps th
come from attempts to u
ing machine. Scientists h
biological neurons are
machines since the Secon
style, models of computa
digital computers that h

movements and thoughts, we are also able to build new thoughts from old in a wonderfully free and creative way. Our current thoughts can become locked into patterns of past thoughts, but they need not be: our human intelligence gives us a wonderful ability to create new thoughts from old. And such freedom and creativity does not depend on rare seeds of genius, or the occasional spark of inspiration; it is fundamental to the basic operation of the brain – how we perceive, dream, talk.

Our freedom has its limits, of course. Amateur saxophonists can't 'freely' choose to play like Charlie Parker, new learners of English can't spontaneously emulate Sylvia Plath, and physics students can't spontaneously reason like Albert Einstein. New actions, skills and thoughts require building a rich and deep mental tradition; and there is no shortcut to the thousands of hours needed to lay down the traces on which expertise is based. And for each of us, our tradition is unique: with those thousands of hours each of us will lay down different traces of thoughts and actions, from which our new thoughts and actions are created. So each of us will play, write and think in our own way – though sometimes with remarkable flexibility (musicians and poets can 'impersonate' each other; and generations of physics students can and do learn to reason in Newton's shadow). Yet the same points arise in our everyday lives – our fears and worries, our sometimes bumpy interactions with other people. Our freedom consists not in the ability magically to transform ourselves in a single jump, but to reshape our thoughts and behaviours, one step at a time: our current thoughts and actions are continually, if slowly, reprogramming our minds.

The ideas in this book are drawn from a wide range of sources, ranging from cognitive, social and clinical psychology, to philosophy and neuroscience. Perhaps the most fundamental inspiration for me has come from attempts to understand the brain as a biological computing machine. Scientists have been considering how vast networks of biological neurons are able to serve as powerful computational machines since the Second World War. Such 'connectionist', or brain-style, models of computation work very differently from the familiar digital computers that have revolutionized all our lives – although

connectionist computation, in the form of so-called 'deep neural networks', is currently sweeping all before it in the project of building intelligent learning machines.

Connectionist models work through a process of 'cooperation' between vast numbers of neurons: filling in, and arranging, fragments of information – rather like simultaneously jiggling the pieces of a jigsaw into place. But it is very hard to reconcile this vision of the machinery of the brain with our everyday intuitions that our minds are guided by beliefs and desires. Our common-sense psychology attempts to explain our behaviour by summoning beliefs, motives, hopes and fears, sequenced into a reasoned argument. So an everyday explanation of why I headed down to the shop on a fruitless attempt to buy a newspaper might involve beliefs (that the shop is open; that it sells my favourite paper; that I have money with me) and desires (to read this particular paper; a preference for the print rather than the online edition). And this explanation makes sense because it can be arranged into a credible justification for my actions – the type of justification I might provide to explain why my actions, though unsuccessful, were entirely reasonable. Rather like mathematical proofs, arguments are formulated step-by-step (I'd like a copy of *The Western Bugle*. I could read it online, but spend my life looking at a screen; I bet they sell it at the corner shop; and surely the shop must be open at this time of day. Oh, I'll need some cash, but luckily I just went to the bank, and so on). We introduce assumptions, draw conclusions, add more assumptions, draw further conclusions, and so on. But this step-by-step style of thinking is difficult to reconcile with the simultaneous and cooperative 'jiggling' of multiple constraints that seems to be involved in recognizing a face, a musical style or an object. Indeed, we don't have very convincing 'arguments' for why I think a cartoon is of Winston Churchill, a snippet of music is Motown, or a glimpsed shape in the water is a seal or the Loch Ness monster. In short, common-sense psychology sees our thought and behaviour as rooted in reasoning, but a lot of human intelligence seems to be a matter of finding complex patterns.

Cooperative connectionist computation does not merely appear to clash with the reason-based explanations of common-sense psychology. It is also extremely difficult to reconcile with many scientific

theories of human thoug
gence, cognitive, develop
and behavioural economi
ideas about minds stocke
starting point. Taking acc
threatens, therefore, to hav
tive, consequences. After t
mathematical models of th
mental data, I have come t
own minds, and many of
been built upon that conc

From a broader perspe
The entire history of scien
after another: that the Ea
elements of which it was
that matter can turn into
of chemicals; and that our
isms. The very idea that th
and chemical activity acr
able enough. Indeed, in th
we think we know about t
tive introspections, justif
abandoned wholesale.

There is a rich tradition o
of our minds. Psychologis
Daniel Dennett, have long
tents of our minds or perc
beliefs, motives, hopes and
and actions are no more
humours of Hippocratic m

Some past sceptics abou
different agenda: doubting
computer at all.[5] This has
brain's function is surely to
memory, to figure out the s
In short, the brain faces spe

theories of human thought, from areas as diverse as artificial intelligence, cognitive, developmental and clinical psychology, linguistics and behavioural economics – theories that have taken common-sense ideas about minds stocked with beliefs, desires and the like as their starting point. Taking account of the computational style of the brain threatens, therefore, to have some far-reaching, and potentially destructive, consequences. After thirty years working with computational and mathematical models of the mind, and surveying and gathering experimental data, I have come to accept that our intuitive conceptions of our own minds, and many of our scientific theories of the mind that have been built upon that conception, are fundamentally flawed.

From a broader perspective, perhaps this should not surprise us. The entire history of science has been a tale of one astonishing shock after another: that the Earth goes around the sun; that the chemical elements of which it was composed were expelled from dying stars; that matter can turn into energy; that life is encoded in a double helix of chemicals; and that our distant ancestors were single-celled organisms. The very idea that thought is the product of a whirr of electrical and chemical activity across a hundred billion nerve cells is remarkable enough. Indeed, in this book I will argue that almost everything we think we know about the operation of our own minds – our intuitive introspections, justifications and explanations – needs to be abandoned wholesale.

There is a rich tradition of scepticism about our common-sense view of our minds. Psychologists and philosophers, from B. F. Skinner to Daniel Dennett, have long doubted our abilities to introspect the contents of our minds or perceptions; and many have suspected that the beliefs, motives, hopes and desires with which we justify our thoughts and actions are no more real than the Garden of Eden, the four humours of Hippocratic medicine, or the principles of astrology.

Some past sceptics about those common-sense stories have a very different agenda: doubting, for example, that the brain is any kind of computer at all.[5] This has always seemed a puzzling line to me: the brain's function is surely to integrate information from perception and memory, to figure out the state of the world, and to decide what to do. In short, the brain faces spectacularly complex information-processing

challenges – and 'information-processing' is really just another label for computation. I'm going to take it as uncontroversial that the brain is a biological computing machine.

What is, and should be, controversial is the proposal that our theory of the calculations that the brain carries out match up, in some fairly direct way, with our common-sense psychological explanation in terms of beliefs and desires and the like – a viewpoint that has been widely held ever since early computational models of the mind took hold in the 1950s. It would, indeed, be awfully convenient if it turned out that our mind contained a database containing the very sorts of beliefs, desires, hopes and fears that we talk about in everyday life. Building intelligent machines would be off to a flying start if it happened that we could just *ask* people what they knew – and write this knowledge directly into a computer database. So the thesis that the stories we tell ourselves about how we think are roughly true would be highly convenient for cognitive science and artificial intelligence *if true*.

But common-sense psychology *isn't true*. We'll see that a very different picture emerges from psychological experiments, the 'wiring' of brain circuits, and the processing mechanisms in modern machine-learning and artificial intelligence that are inspired by cooperative, brain-style computation. Our 'computational innards' are not a churning sea of experiences, feelings, beliefs, desires, hopes and fears, whether conscious or unconscious. Our mind spins stories about how we work – driven by motives, beliefs, percepts, moral norms, religious precepts. And they are such compelling stories that we can imagine that they are true, or partially true, or surely at least along the right general lines.

In reality, though, the rich mental world we imagine that we are 'looking in on' moment-by-moment, is actually a story that we are inventing moment-by-moment. A real Anna Karenina would no more have an 'inner mental world' than a fictional Anna Karenina. If this is right, then there is no more hope of finding a real Anna's 'inner feelings', 'deep beliefs' or 'true nature' by probing her brain, than there would be of uncovering the inner life of the fictional Anna by a scientific analysis of the ink and paper of Tolstoy's manuscript.

*

I have struggled long and hard to swallow this troubling truth, for several reasons. The first is that some of the psychological data just seem too extraordinary to believe. For example, the psychological data tell me that my brain can identify just one word at a time, but looking at the text in front of me, I have the overwhelming impression that I am simultaneously, if slightly tenuously, aware of the identities of words across entire paragraphs. The experimental data tell me that I can, roughly, identify *one object at a time*. Yet as I scan the room, I feel that I am taking in almost an entire roomful of sofas, cushions, books, glasses, plants and pieces of paper. The list of counter-intuitive experiments and strange quirks of the brain tells a story so downright peculiar that it is easy to imagine that there must be some mistake or misunderstanding. For our intuitions to be so wrong about the contents and workings of our own minds implies that we are subject to a systematic and all-pervasive illusion. I have now, somewhat reluctantly, come to the conclusion that almost everything we think we know about our own minds is a hoax, played on us by our own brains. We will see later how the hoax is done, and why it is so compelling.

The second reason for my long resistance to adopting the radical perspective outlined in this book is that it clashes, not just with common sense, but with theories of perception, reasoning, categorization, decision-making and more, which have been central to psychology, cognitive science, artificial intelligence, linguistics and behavioural economics. So much of our most sophisticated thinking in these disciplines has involved extending, modifying and elaborating our intuitive conception of our minds – an intuitive conception that is founded on an illusion. Jettisoning such a large portion of the ideas in disciplines I have been closely involved with for so long feels a little like vandalism.

The final reason I have held on so long to our everyday picture of the mind is simply the lack of a credible alternative. Yet in the light of advances in 'brain-style' computation, and machine-learning more broadly, I think that the outline of a credible alternative is beginning to emerge.[6] Increasingly, researchers are finding that the best way to make computers display intelligent behaviour is not to attempt to extract the knowledge and beliefs that people say they rely on; it is far more effective to create intelligent behaviour by designing machines

that can learn from experience. So, for example, when trying to write a computer program to play world-class backgammon, chess or Go, it turns out to be unhelpful to 'program in' knowledge, insights and strategies that top human players claim they are using. It is much better to learn, bottom-up, from the experience of playing vast numbers of games – and, indeed, machine-learning programs are now able to beat the best human players in these and many other games.

This has been a delightful, though slightly unsettling, book to write. As an academic, now in the Behavioural Science Group at Warwick Business School, I have worked mostly on quite specific, though varied, aspects of the mind, from reasoning to decision-making, from perception to language. I have also worked on a wide range of practical behavioural science projects with my colleagues at Decision Technology Ltd. In both the academic and practical worlds, it pays to be circumspect and not to stray too far into unfamiliar territory. This book takes the opposite approach: throwing caution to the wind, I want to give you the most compelling story I can about how the mind works. And in truth, I don't just want to tell this story to you, the reader. I want to tell this story to myself: to zoom out from the patchwork of observations, data and theory that are my daily life and ask 'what does it all mean?' This requires joining up dots, extrapolating exuberantly from the specific to the general, and engaging in doses of outright speculation. Drawing together data and insights from more than a century of psychology, philosophy and neuroscience, the resulting picture is strange, radical and liberating. This viewpoint has, I think, been coalescing in many areas of the cognitive and brain sciences over the past few decades, even as much of the field has continued to proceed with 'business as usual'. But business as usual won't do: when we take seriously what science is telling us about the mind and the brain, we are forced to reconsider just about everything we thought we knew about ourselves. This requires a systematic rethink of large parts of psychology, neuroscience and the social sciences, but it also requires a radical shake-up of how each of us thinks about ourselves and those around us.

I have had a lot of help writing this book. My thinking has been shaped by decades of conversations with Mike Oaksford and Morten Christiansen, and discussions over the years with John Anderson, Gordon Brown, Ulrike Hahn, Geoff Hinton, Richard Holton, George Loewenstein, Jay McClelland, Adam Sanborn, Jerry Seligman, Neil Stewart, Josh Tenenbaum and James Tresilian, and so many other wonderful friends and colleagues. Writing this book has been supported by generous financial support through grants from the ERC (grant 295917-RATIONALITY), the ESRC Network for Integrated Behavioural Science (grant number ES/K002201/1) and the Leverhulme Trust (grant number RP2012-V-022). My colleagues at the Behavioural Science Group at Warwick Business School have provided the perfect intellectual environment for a wide-ranging and downright speculative project such as this; and the interdisciplinary and innovative spirit of the University of Warwick has been an inspiration. This book has benefited hugely from the insights, suggestions and encouragement of my wonderful agent, Catherine Clarke at Felicity Bryan Associates, and of Alexis Kirshbaum and Laura Stickney, my editors at Penguin. I am also grateful to my wife, Louie Fooks, and our daughters Maya and Caitlin Fooks, for their unstinting support throughout this book's uncomfortable and lengthy gestation, for their critical discussions of its arguments, but most importantly for being delightful people with whom to share a life. Finally, I would like to thank my parents, Robert and Dorothy Chater: without their faith in me, and their continuous love and support, I would probably have never persisted long in research, still less written this book.

Nick Chater
Coventry, 2017

11

Precedents not Principles

THE STRANGE CASE OF CHESS

In 1922, the reigning world chess champion, the Cuban José Raúl Capablanca, simultaneously took on 103 opponents in Cleveland, Ohio. After seven hours, he had won 102 games and drawn one. But with so little time to calculate, how could this be possible?

It is all too natural to imagine that Capablanca must have been playing lightning chess with a lightning brain; that he must have been able to race through the forking paths of move and counter-move faster than his bemused opponents. If this was indeed his secret, then it would imply that Capablanca could calculate a hundred times faster than his hundred opponents, just to get a hundred draws (after all, Capablanca has just one hundredth of the time each of them has to think through the next move). Not only that, Capablanca – unlike his opponents – would have to spend a great deal of time moving from board to board (whereas his opponents would be able to focus relentlessly on a single game). Over the entire seven-hour period, he made, on average, about ten moves per minute: this pace is consistent with sidling up to a board, taking a quick look, making a move, and then shifting onto the next game And, of course, if Capablanca's forte was lightning calculation, then he would need not just to match the calculation of his opponents, but to exceed it very considerably – after all, his opposition was comprehensively flattened. So, in short, one might suspect that, if Capablanca's magic trick was lightning-fast calculation, then he must have been able to calculate several hundred times faster than anyone else – a sort of human supercomputer.

Now this story would, indeed, be a pretty good model for

computer chess. Of course, the chess-playing ability of a computer program depends a great deal on how cleverly the program is designed. But a decisive factor in the dramatic rise in the quality of computer chess over the past few decades has been the equally dramatic increase in the raw speed of computer processing. Contemporary computer chess programs really do calculate at lightning speed – evaluating many millions of possible board positions per second. So a super-fast computer chess program could, indeed, simultaneously crush one hundred more sluggish chess programs, if it could calculate, say, 500 or 1,000 times faster. The super-fast computer would simply be able to look further into the forking tree of possible moves and counter-moves that its ponderous opponents could make.

Yet Capablanca did not have the brain of a lightning calculating machine – and he did not need it. A glance at the board would instead call up past games, and past good (and bad) moves – one clue is that experimental tests show that chess grandmasters have a phenomenal ability to remember chess positions from real chess games of the past. After five seconds, a top chess player is able to 'read' the structure of a chess position – searching out which pieces are threatening which others, noting familiar patterns of pieces (e.g. a castled king and rook behind a row of pawns; advanced central pawns defended by knights, and so on), finding, in short, what the position *means*. And this is usually enough, not just to decide on a high-quality next move, but to commit the position, aside from a few incidental details, completely to memory – indeed, a grandmaster may be able to recall that very position many minutes or even hours later. To the inexpert chess player, this astonishing feat suggests that grandmasters supplement fantastic calculating ability with stupendous powers of memory. But the feat is no more remarkable than our ability to commit to memory an enormously long string of letters, simply because they are arranged meaningfully, as in this sentence. Someone unfamiliar with the English language, or even the Roman alphabet, would find this ability equally astounding – precisely because they could impose no meaning upon it (consider your own ability to remember strings of characters in a distant language in an unfamiliar script). Or consider a skilled musician's ability to reproduce a lengthy stream of musical notation, because it can be turned into

a meaningful tune (while, to the rest of us, the pattern of notes on a stave is mere gobbledygook). In each case, memory is the by-product of understanding: what we cannot interpret, we cannot remember.

It seems, then, that grandmasters are special not because of their unusual mental powers, but because, through long experience, they have learned to find meaning in chess positions with particular fluency; and they can do this because they can link the current board position with their memory traces of past board positions, acquired through thousands of hours of chess-playing.

Two further observations reinforce this picture. The first concerns the nature of the memory mistakes that grandmasters make. They recall, more or less unfailingly, the pieces that matter to the progress of the game, but where some peripheral piece plays no active role its precise location need not be encoded exactly. By contrast, non-chess players make mistakes of all and every kind – to them, of course, the board is a mere jumble of pieces, not a subtly interlocking set of threats, counter-threats and defences.

The second observation is that grandmasters are no better than the rest of us at remembering random board positions – their superior memory skills evaporate as soon as they are faced with arbitrary chess positions, because they can impose no meaningful interpretation on these positions in relation to their huge repertoire of past chess experience.[1] In the same way, readers of English will struggle to learn random sequences of letters, just as expert musicians will have no special ability to recall arbitrary forests of musical notation.

The ability to make sense of current chess positions, by linking them with the vast library of past positions, greatly simplifies the problem of choosing the right move (just as familiarity with English makes continuing an English sentence relatively straightforward, and musical expertise makes it possible to write out a plausible next bar or two of a simple tune). In chess, of course, the number of possible continuations explodes as we attempt to see further ahead in the game – but almost all of these moves are wildly implausible and can safely be ignored.

Grandmasters do not reach such impressive levels of performance by out-calculating their bemused opponents and seeing many more moves ahead. Instead, they look only a little further ahead than

amateur chess-players, but their memory bank of past experience and, in particular, their library of meaningful analyses of chess positions, allows them to focus on only the best moves, and to ignore the rest.

Notice, too, that Capablanca had no 'theory' of chess, beyond his vast fund of experience. He wrote a number of celebrated chess books, one promisingly entitled *Chess Fundamentals*, with lists of 'principles' to guide the aspiring player.[2] Yet these principles are, in reality, a series of helpful examples illustrating useful rules of thumb – there is no chess equivalent of Newton's Laws! Tellingly, Capablanca is attempting to distil some of his knowledge not into *principles*, but into particularly helpful *illustrations*.

To learn to play chess is to learn to impose meaning on each board; and each board makes more sense in the light of the meaning imposed on previous boards. Perhaps expertise in any domain, however remarkable, is not based on superior mental calculating power, but on richer and deeper experience: Capablanca could play new chess position because he had a vast library of *precedents* from imposing meaning on prior chess positions – and could use those precedents more creatively and effectively than anyone else. And perhaps that is how skills, learning, memory and knowledge always work. We layer each momentary thought on top of past momentary thoughts, tracing an ever-richer web of connections across our mental surface.

PERCEPTION–MEMORY RESONANCE

The interpretation of each new chess position depends on a vast battery of interpretations of prior chess positions; and, in the same way, the interpretation of each everyday scene depends on a vast hoard of past interpretations of everyday scenes. Indeed, perception works by relating, often in the most flexible and creative fashion, our sensory input with our memory of past experience. We do not interpret every sensory impression afresh, but in terms of the memory traces of past sensory impressions. Consider the rather delightful 'found faces' in Figure 37. In terms of superficial shapes and colours, these items – variously a leather bag, cheese-grater, a block of wood and the detail

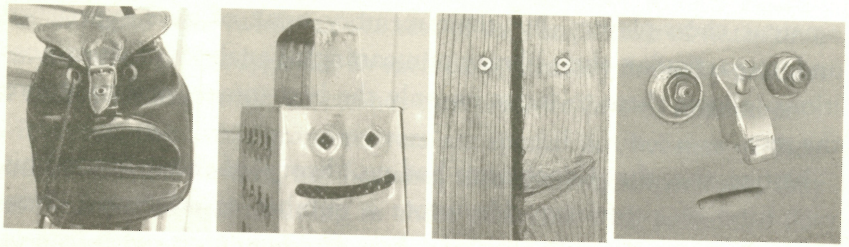


Figure 37. Found faces.[3]

of a wash basin – are about as distant as possible from human form. Yet we see them, more or less immediately, not merely as faces, but as faces with personalities, expressions and even a measure of pathos that seems particularly incongruous given that each of these objects is patently inanimate.

The cycle of thought should, then, create an organization of the sensory input that depends not just on the input itself, but on 'resonance' with memory traces of past inputs (Figure 38) – for example, previous faces that we have encountered. The brain interprets one word, face or pattern at a time, but in doing so it simultaneously explores possible links between the current stimulus and a vast array of memories of interpretations of past stimuli. 'Resonances' between current and past stimuli are not defined by superficial similarity – otherwise the brain could only interpret a cheese-grater in terms of past silver metallic box-like objects, and never as a face. Yet the brain rapidly sees the 'eyes and mouth' pattern in this otherwise incredibly un-face-like object. The 'found faces' of Figure 37 are illustrations of just how flexibly the brain can map past memory traces (of human faces) to impose meaning on the current input.

The resonance between perception and memory must occur 'in parallel', however. Given our sluggish neurons, matching the current perceptual input with each of our vast repertoire of memory traces one at a time would be unfeasibly slow. And, when interpreting a new stimulus, the brain may have little idea which memories it needs to search – indeed, it seems able to draw on its entire stock of memory traces equally easily. Before the interpretation has been made, the

brain can't know which memories will be relevant – so it has to search them all.[4]

Notice that, from this viewpoint, each new perceptual interpretation is based on memories of past interpretations. We *never* see the world 'with fresh eyes'. Each new interpretation is an amalgam and transformation of past interpretations. Consider what happens when you read a word, or 'read' a face, or a chessboard – those perceptual interpretations depend on years of past experience of our language and writing system, of our long history of interactions with other people, and the nature of past experience (if any) with the game of chess. As usual, of course, we have awareness only of the *output* of thought – the result of our current interpretation; all other mental processes leading to that output are never conscious. Thus we have no awareness of the memories activated, or how they were transformed and combined to interpret the current stimulus. Listening to our own native language, we 'hear' the speech sounds, words and pauses as if they are plainly observable aspects of the speech signal. But listening to an entirely unfamiliar language, we are confronted with a baffling, uncategorized and apparently chaotic flow of sound. The difference is that, for our own language, we can map new streams of sounds to a vast range of previously interpreted speech sounds, words, phrases and more. We can interpret new speech input in terms of a vast repertoire of memory traces from our past interpretations of past speech input. As with learning any other skill, we break down the code of our language piece by piece, over many months and years; and, as with any other skill, we are aware only of the results of our current competence, oblivious to the myriad of memory traces on which that competence depends.[5]

What are memory traces like? What information do they contain? The most natural answer is that they are nothing more than the remnants of past interpretations of past perceptual inputs. As far as we know, these remnants are not later reorganized, filtered, corrected or generally tidied up; and there is no internal librarian to carefully file and index each memory trace into a coherent archive. The remnants of each individual episode of perceptual processing lie, as it were, where they fall; the brain is immediately busy with the next cycle of thought, and the next.

The brain is therefore not a theorist trying to distil deep abstract principles from experience – it is, instead, focused on coping with the present, as far as possible, *by relating the present to amalgams and transformations of the past*. According to this viewpoint, then, memory traces are fragments of past processing – so that it is past *interpretations* that are stored in memory, rather than raw, disorganized sensory input. So, for example, if we see a cheese-grater as a face in one situation, then that interpretation will be stored in memory. When we next encounter a similar cheese-grater, we are more likely to see that grater, too, as a smiley face – we remember the interpretation.

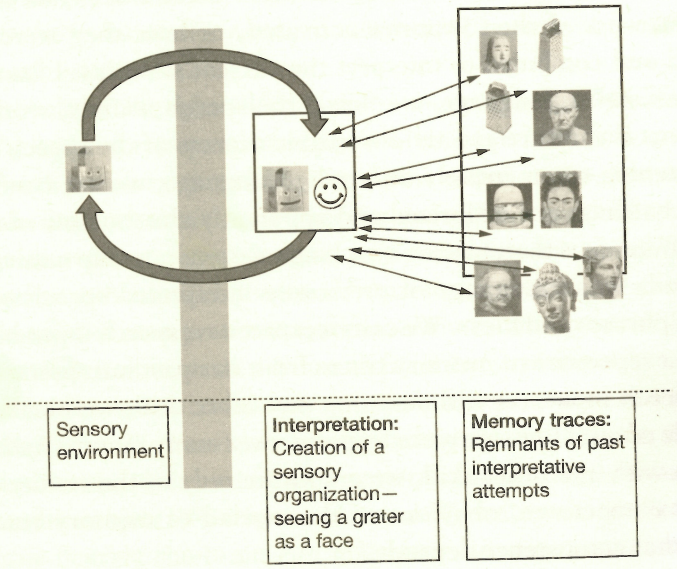


Figure 38. Perception–memory resonance. *Left*: Our senses provide us with highly ambiguous information (here, the face-like image of a cheese-grater in Figure 37). *Middle*: we interpret the cheese-grater as a rather foolish grinning face (this interpretation is depicted by the addition of a smiley face). *Right*: but, of course, the interpretation of the image as a cheese-grater, or as a face, depends on transformations of previous experiences of cheese-graters and faces (a selection of images of past cheese-graters and faces are shown) – interpretation involves the transformation of past memory traces.

Conversely, uninterpreted aspects of the sensory world will be forgotten: a piece of handwriting that is too difficult for us to read, a fragment of a language we don't understand, or the outlines of a distant figure in the trees that we fail to notice, will not, by this account, be filed away in our memory to be subject to future analysis, or to shape later perceptions. They will be lost for ever. (In passing, note that this is reassuring news for those worried about subliminal messages that might be planted by advertisers or nefarious forces wishing to control our minds by stealth.)

Perception and memory are therefore intricately entwined. Recognizing a friend, a word or a tune requires not merely linking together different aspects of the perceptual input, but connecting these fragments to stored memories of faces, words and melodies. So, for example, a recognized face typically does not merely feel vaguely familiar: we can also access information about the corresponding person. The string of letters making up a word is typically also linked to its meaning, its sound, and much more; and recognizing a tune may conjure up the associated lyrics, singer, era in which we first heard it, and more. So the interpretation of the information flowing through our senses depends on a huge body of remembered information, but this information is, of course, nothing more than the memory of past interpretations of previous sensory information. Today's memories are yesterday's perceptual interpretations.

Successful perception, then, requires almost instant conjuring up, and deploying, of the 'right' memory traces to make sense of the current sensory input, just when we need them. This is remarkable in the light of the sheer number of such traces, accumulated over a lifetime. It is more remarkable still, given that the right memories can be very indirectly related to the perceptual fragments: that images ostensibly of handbags, cheese-graters, blocks of wood or hand basins can trigger memories of faces (see Figure 37); that a few scratches of ink on a sheet of paper can conjure the human figure in a landscape (Figure 39b); that the rearrangement of the same few geometric shapes can remind us of a rocket, a kneeling person, a rabbit and many other images (Figure 39a). And the same spectacular flexibility of thought arises of course throughout the metaphors that pervade our language and our thoughts – memories of one thing are fluidly and



(a) (b)

Figure 39. The interaction of memory and perception. (a) Tangrams are seven simple geometric shapes made from dissecting a square (*top left*), but can be made to resemble an almost infinite variety of people, animals and objects.[6] (b) Picasso's famous sketch of Don Quixote, Sancho Panza and a background of distant windmills is remarkably underspecified, yet sufficient to conjure up these literary characters, the burning sun and the bare Spanish landscape.

naturally linked to our memories of another thing. So, for example, we can 'see' our boss variously as a conductor, a general, a robot or a shark.

A helpful way to view how perceptual or memory information is analysed is to regard the given information (the clues from memory and perception) as providing parts of a pattern, with the brain's task being to 'fill in' the gaps in the pattern. But this image underplays the spectacular flexibility of the brain in transferring knowledge from one domain to some apparently entirely unrelated topic. So, for example, taking some fragments of visual information about a handbag, the brain may indeed fill in various other details (inferring, perhaps, that the sealed container – the handbag – has an opening only at the top, or that it comes from a specific period or country, or that it has a certain monetary value), but equally the brain may well interpret it as an alarming, if slightly comical, snarl. The brain

operates by the routine exercise of an astonishingly exuberant imagination.

Very much the same story applies when we consider consulting not the contents of the visual world (handbags, cheese-graters, actual human faces), but our own memories. We can have the feeling of immediate and simultaneous access to a vast repository of general knowledge, autobiography, likes and dislikes, moral and religious convictions, and more. But in truth we impose meaning on one set of memory traces at a time. And when we do lock onto those memory traces, we impose meaning on them with the same flexibility and urgency that applies in perception. The memories themselves are not thoughts: they are not, for example, beliefs, choices or preferences. And we cannot merely 'read off' their contents to know what we think, what we like, or what sort of person we are. Instead, they consist of mere fragments of past thoughts, to be reused, reconstituted and transformed by the cycle of thought.

PEOPLE AS TRADITIONS

What makes me, me? What makes you, you? Tempted, as ever, to peer into the darkness of our mental depths, we search for the attributes of character: bravery, stoicism, anxiousness, kindness or cruelty. What are we really like, deep down? It seems hard to be sure of our own nature – each of us is, of course, a mixture of momentary thoughts and feelings, at one moment brave, at another timorous; sometimes stoical, at other times anxious. But it is a seductive idea that our confusions and contradictions are merely the stuff of our turbulent surface, the vagaries of the moment pushing us here and there. Deep down, perhaps too deep for us to fathom, lies our inner self, replete with the virtues and vices of our true character. But mental depth is, we have seen, an illusion; there is no inner core, virtuous or venal.

Yet if the mind is an engine of precedent, continually reshaping past thoughts and actions to deal with the present, then each of us is not just a bundle of character traits, but a rich store of distinctive past

experience: we are like corals layered, polyp by polyp, into infinitely diverse forms. What makes each of us unique is our individual and particular history – our own specific trail of precedents in thought and action. We are each unique, in short, because of the endless variety of our layered history of thoughts and actions.

This viewpoint might seem to imply that we are nothing more than 'creatures of habit'. Not at all. As we shall see in the next and final chapter, it is our remarkable ability to make imaginative leaps, both large and small, that breaks us free of blind repetition. We can project our experience of everyday human faces onto the 'found faces' in handbags and sinks (remember Figure 37). The subtle and measured application of past precedents, not merely the sheer number of chess-playing hours, surely distinguishes the brilliance of great champions like José Raúl Capablanca, Bobby Fischer or Magnus Carlsen from the everyday run of Masters and Grandmasters. Or consider our ability to improvise dance, create songs, paint and draw, spin stories, or invent imaginary worlds – not from scratch, but rather through the reinterpretation and reconfiguration of the elements of the world we know.

Each of us is a tradition, guided and shaped by our past. Like traditions in music, art, literature, language or the law, we are capable of refinement, adjustment, reinterpretation and whole-sale reinvention. Our mental present is built out of our mental past, but our imagination need not be trapped in cells we constructed long ago; we continually build and rebuild ourselves. Re-routing our minds is always slow and difficult. But where we can purposefully shift our present, there is hope of reshaping our future.

PRECEDENTS NOT PRINCIPLES

Each turn of the cycle of thought leaves a trace, and that trace may shape future turns in the cycle. Thoughts are like water droplets finding their way from high ground to the sea, following the channels in the landscape, whether gullies, streams or river valleys. And, in its passing, each droplet cut those channels just a little more deeply. The landscape, then, is partly a history of past water flow, as well as a

guide for how water will flow in the future. In the same way, our mental life follows channels carved by our previous thoughts, and traces of our present thoughts and actions will shape how we think and act in the future.

Droplets of water make their way downhill by the steepest available path, each heedless of the incalculable number of droplets that have created the landscape through which it runs, and heedless, too, of how its own imperceptible corrosive force may alter, ever so slightly, the path of the next droplet, and the next. Similarly, each cycle of thought lays down memory traces that may smooth, or obstruct, future cycles. Each momentary interpretation draws on and adapts our previous attempts to understand the world. So each cycle of thought maybe viewed as creating, over time, mental channels along which our thoughts most easily flow; and, indeed, each cycle of thought attempts, as far as possible, to make our interpretation of current sensory information cohere with our interpretations of prior sensory information.

Over a lifetime, the flow of thought shapes, and is shaped into, complex patterns: our habits of mind, our mental repertoire. These past patterns of thought, and their traces in memory, underpin our remarkable mental abilities, shape how we behave and make each of us unique. So, in a sense, we do, after all, possess some inner mental landscape. But this is not an inner copy of the outer world or, for that matter, a library of beliefs, motives, hopes or fears; it is, instead, a record of the impact of past cycles of thought – rather than, as it were, any mysterious subterranean geological forces.

The brain operates by precedents not principles. Each new cycle of thought makes sense of the information to which we are currently attending, by reworking and transforming remnants of past related thoughts. And the result of each cycle of thought becomes, itself, raw material for future thoughts.

So the failure of early artificial intelligence to discover the principles that underlie our knowledge of the physical and social worlds; the failure of linguistics to uncover the grammatical principles that generate language; and the failure of philosophy to articulate the principles underlying the true significance of truth, goodness and the nature of mind – all have a common source. The system of precedents

underpinning human intelligence can be contradictory, highly flexible and open-ended – especially in relation to cases for which no nearby precedent has yet been set. But this very open-endedness is precisely what is required when dealing with a world that is still far too complex for us fully to understand.[7]

12

The Secret of Intelligence

In the 1950s, Canadian psychologist Craig Mooney created a remarkable array of black and white images (Figure 40). The faces were created for a rather narrow purpose, as a test of the development of face perception in childhood.[1] Yet thinking about these faces, and how our brain understands them, provides an opportunity to encapsulate some of the key themes we have explored in this book – and provides a starting point for investigating the 'secret' of intelligence.



Figure 40. Craig Mooney's black and white faces. Many of these images are, on first inspection, pretty baffling. But look more closely, and we see not just faces, but glimpses of full, rounded, psychologically rich, individual people.[2]

Initially, the display is rather mystifying, but spend a minute or two looking at these strange black and white patterns and the experience becomes more interesting. Somewhat miraculously, the initially inchoate patterns start to make sense. To my eye at least, each face creates a vivid sense of a particular person caught in a particular mood, with a specific expression, gender, age, personality, and perhaps even a historical period. Some of these simple arrangements of black and white patches seem to have as much life, pathos and a sense of human drama as the most detailed photographic image. They are also rather beautiful.

Some faces jump out at us almost right away; others remain enigmatic until, mysteriously, the baffling bumps, curves and blobs resolve into a portrait of a human being. You may not be able to see all the patterns as faces – indeed, there are still a few of these images that I can't make sense of, having looked at them countless times over the years. But if you see even a few, you are still far ahead of any computer vision technology yet invented.

On reflection, finding faces in these incredibly simplified and stylized patterns is an astonishing feat: these strange 'woodcut' patterns are so utterly unlike the colourful, three-dimensional, moving people we see around us. Where are the eyes, nose and mouth that we might be expecting our brain to pick up on? These features appear to be wholly absent, at least until we crack the code revealing the entire face. The components of the image are just jumbled bumps, curves and blobs, until, of course, the delightful moment that a human face breaks through to us from the darkness. Once the 'insight' occurs, our interpretation of that image is permanently or near-permanently recorded in memory – a jumble has become a face, and will now be a face for ever.[3]

To make life more difficult still, try turning the page upside down. A few of the faces you initially identified will still make sense. Many, though, will return to seeming no more than abstract patterns. Give yourself time, though; look back at the face the right way up. Gradually, more and more images will snap into coherence.

There are several interesting conclusions to draw from looking at Mooney's faces, both the right way up and upside down. Seeing the images upside down gives us a sense of how truly chaotic they

are – that there are, indeed, no eyes, noses, mouths or ears that can be identified in isolation. When the faces are the right way up, some of these features can, sometimes sharply and sometimes only dimly, be discerned – we see the parts only when we are able to grasp the whole. The holistic way we perceive Mooney faces (i.e. the interdependence of perceiving a whole pattern and its component parts) reflects the general operation of the brain. Reading scrawled handwriting, we recognize the letters in tandem with understanding the whole word. We recognize speech sounds, word-breaks and so on at the same time as we decode the message that a person is attempting to convey (and recall that speech in an unfamiliar language is little more than a blur of sound). And we interpret a small wooden object being slid from a white square to an adjacent black square as advancing a pawn, threatening the opponent's knight, or leading to checkmate in three moves, by interpreting a physical aspect of the world in terms of the rules of chess. Further, as we saw in Chapter 5, Lev Kuleshov demonstrated how the same facial expression can be interpreted as subtly conveying sorrow, hunger or lust; and Schachter and Singer showed that the way we interpret our own bodily states (rushes of adrenaline, racing heart) as anger or elation depends on our interpretation of the social interaction we are engaged in (e.g. whether the person we are interacting with is being annoying or amusing).

Note, too, that our sense of the entire face is remarkably rich, even if those individual features are difficult to discern; I suspect you feel, as I do, that you might be able to pick out some of these people if they appeared in a police line-up. The leap from a stark, black-and-white image to a three-dimensional real person is, of course, vast – the fact that we can make such remarkable perceptual leaps is a reminder of the astonishing flexibility and, one might even say, creativity of the perceptual system.

The human brain is wonderfully attuned to faces – so that we can detect and reconstruct them on the flimsiest of evidence, as the Mooney faces attest. And, indeed, we can see faces when the evidence is flimsier still. Recall the 'found faces' we saw in Figure 37. The snarling handbag at the extreme left looks, to my eye, at once angry and supercilious; the cheese-grater is youthful, eager to please, and

perhaps a little anxious; the block of wood has a relaxed but slightly drunken look; while the taps and sink on the extreme right seems preoccupied with its own thoughts, and riddled with anxiety. Yet what is really remarkable about this array of faces is that we can see them at all. Each image is incredibly distant from any real human (or indeed animal) face we have ever encountered. Yet they not only seem face-like, but, like the Mooney faces, they convey emotional expressions and even, to a degree, personalities. The imaginative leap required to distort and stretch these everyday objects onto the templates of a prototypical human face is very great; our brain makes such leaps so rapidly and so naturally that we think of it, if we consider it at all, as entirely unremarkable.

These imaginative jumps are, I believe, at the very core of human intelligence. The ability to select, recombine and modify past precedents to deal with present experience is what allows us to be able to cope with an open-ended world we scarcely understand. The cycle of thought does not merely refer passively to past precedents – we imaginatively create the present using the raw materials of the past.

THE UBIQUITY OF METAPHOR

The wild inventiveness of the human brain is perhaps best exemplified by the ubiquity and centrality of metaphor in our thoughts. Beyond seeing cheese-graters as grinning stupidly and hand basins looking diffident, we are forever seeing one thing in terms of something else completely. We portray each other as bursting with (or just full of) emotion, letting off steam, feeling down or depressed, or on top of the world; being weighed down or buoyed up, feeling flat or bouncy. We can be mixed up, messed up, straightened out, put straight; thoughts can be crowding in on us, we can be teeming with ideas, or our mind can be blank, empty, a void; we can be lighthearted or our thoughts can turn dark or even black. Our ideas can be sharp, penetrating, incisive, sparky, sparkling, fizzing, bright, brilliant, illuminating, or simply dull and blunted; and our words can be cutting, barbed or pointed, or perhaps smooth, silky, silvery or even oily. Or consider our physical condition, which can be in shape, out

of shape, full of – or drained of – energy, broken down or patched up, out of gas, or having a second wind. And, of course, the very idea of *seeing* one thing as another is itself a metaphor, using visual perception as a proxy for thought. Our language is thoroughly soaked in metaphor.

Metaphors also thoroughly permeate our thoughts. Consider the illusion of mental depth that we discussed in Part One. Once we begin to see how thoughts are 'hidden' below a mental 'surface', it seems only natural to try to 'uncover' them, 'bring them to the surface'; to suspect that some people's thoughts run deep, while others are shallow. And the very idea that the mind is flat is, of course, just one more metaphor, albeit, I hope, a helpful antidote to our standard repertoire. And, needless to say, *antidote* and *repertoire* are metaphors too![4]

Just like faces, metaphors are everywhere, when we stop to notice them. And, indeed, it is hard to imagine how we could talk, or think, without them. If you have your doubts, try rewriting a few sentences of this book with the metaphors taken out, stripped away, removed, excised, deleted, cut out, missing, absent or eliminated, or whatever would be the appropriate, completely literal, phrasing, if one exists.

Some metaphors can of course become fossilized in the language (to deploy yet another metaphor), divorced (help, another one!) from any original meaning, and sometimes continue to be used where their origin is entirely forgotten. Presumably, *riding* a bicycle was once a metaphorical extension of riding a horse, but in a world in which bicycles are plentiful and horses are few, any such link has been lost. Indeed, our entire language (including this very sentence) is a graveyard of dead or half-dead metaphors.

Found faces and metaphors have a lot in common. Like found faces, metaphors embody three characteristics. First, a metaphor requires non-obvious, lateral thinking (another metaphor, of course), to link together two apparently unrelated domains (cheese-graters versus faces; real objects being hidden, buried, brought to the surface, and thoughts being 'hidden', 'buried', or 'brought to the surface'). Second, metaphor, by its very nature, requires the *transformation* of past experience into present experience: between, say, resolving outstanding issues in a project and 'tying up loose ends' of string or rope.

We can only see an unthreatening shark as a teddy bear if we have some experience or knowledge of teddy bears. And we can only see a person as a shark if we have some experience of, or at least some prior knowledge of, sharks. And, third, metaphors are just as apt to *mislead* as they are to inform. Just as a cheese-grater doesn't really feel eager to please and a block of wood can't get slightly inebriated, our inventive metaphorical leaps can be wildly off-target. For example, if the intricate mechanism of Nature is like the workings of a watch, as the eighteenth-century clergyman and theologian William Paley famously suggested, then it is easy to leap to the conclusion that it must have a designer, and one far more skilled and intelligent than any watchmaker. Evolutionary biology, of course, tells a very different story.

Battles of ideas are often fought over which metaphor is appropriate: is light made up of particles or waves? Are humans 'risen' apes or fallen gods? Is nature a harmonious society or a brutal war of all against all? Such metaphors are not marginal to thought, but its very essence. Our continual search for meaning is the struggle to find patterns in our present experience, in the light of the past. And so we see one thing in terms of another: a wash basin as a face; a mind as a container, a sea, or an inner world. Metaphors, too, are also employed to impose *meaning* upon one aspect of the world, by drawing on an understanding of another. We have lots of everyday experience with water waves (whether in puddles, ponds or the sea); and this helps us understand what it might mean for sound, light or even gravitation to be a type of wave (interfering, refracting, diffracting and so on). Or we can take our intuitions about water flow to help us understand the flow of heat or electricity.

Or, to consider a final, more whimsical, example, consider the game of charades. The title of a book, song or film is whispered in your ear; and you must communicate its identity to other members of your team with a sequence of improvised gestures and movements. One moment we may be transformed into a ravening dog for *The Hound of the Baskervilles*; the next we may desperately attempt to indicate the form of a rose flower, closing into mere bud, or, failing that, to stagger about with an imaginary stick, in an attempt to convey *Citizen Kane*. What is astonishing about charades is that we can

play it at all: that a person can take on enough of the movement and gnashing teeth of a dog to convey the famous Hound; that a mime with the hands can imply the stalks, thorns, flower and even bud of a rose; and that we can walk *as if* supported by a stick, when the stick is patently absent. And we can do this naturally and spontaneously, and without years of dedicated preparation or rehearsal.

The team-mates attempting to decode our movements face a challenge somewhat akin to seeing found faces: they need to find some aspects of the communicator's demeanour and actions which suggest, however remotely, the target idea (and they will, of course, have additional clues from what they know about the communicator, their shared cultural knowledge, and so on). The communicator has to go in a reverse direction, creating something akin to the 'found hound', 'found rose' or 'found stick' sensory input for her team-mates – and to do so on the spot.

The search for interpretation is centre-stage again: we are able to create and decode inventive *transformations of past experience* – all the way from the book, to a hound, to a hound engaged in a violent raging attack (we don't attempt to portray the hound sleeping peacefully, or drinking from a bowl of water) – then map the imagined hound's body movements onto our own pantomime actions. And all of this requires the ingenious transformation of a wealth of knowledge – of the book and its central theme, and some crude grasp of the motifs of a canine attack, and how this can map onto the human body (so that, for example, arms become front legs, and fingers become claws). Such coding and decoding is all too imperfect, as players of charades will be aware. A mime intended to portray the flailing claws and gnashing teeth of the famous hound can so easily be read as *Jurassic Park*.

IMAGINATION AND INTELLIGENCE

The whimsical imagination which creates faces from handbags or from apparently meaningless black and white patterns, which plays with metaphors and invents stories, and which creates and perceives songs and art, might seem charming but impractical. Such

exuberant mental leaps may appear primarily to be the province of the arts. After all, the worlds of painting and sculpture are replete with incredibly sketchy, incomplete and distorted representations of the human face, or indeed of many other things; and literature is full of metaphors, one character or story representing some other character or story, and relies on our brains turning a text, a stage, or the movements of a TV or film screen into the semblance of an entire world.

Why has our species developed such rich imaginative powers? Survival and reproduction surely don't depend on seeing faces in inanimate objects, viewing moods as spatial locations (up or down) or personalities as light sources (brilliant, sparkling, dazzling, bright on the one hand, or dull and dim on the other), or generating a continual stream of stories, whether credible or completely fantastical, to explain the world around us. Why, indeed, hasn't the iron logic of evolution by natural selection eliminated such playfulness in favour of a relentless focus on, say, commando-style survival skills? And how could the infinite imaginative landscape of romance and love arise in a species shaped, one might think, to have a one-dimensional focus on sex for reproduction?

Putting evolution to one side, and simply considering the demands of modern life, one might also suspect that capricious flights of fancy, while perhaps entertaining and enriching, must be secondary to the core practical business of the mind: understanding the world around us, making decisions and planning actions, and sending precise instructions and observations to others. Years of formal education, and the experience of working, as most of us do, somewhere in the middle of a complex web of bureaucratic processes, might reinforce such suspicions. Isn't the essence of thought discipline and control, rather than whimsical and wild connections? It is tempting to see imagination as no more than a mental 'bauble': conspicuous but entirely dispensable.

But we have seen that the opposite is true: imaginative leaps are essential for perceiving the world and making sense of each other. Such leaps allow us to project our past experiences onto an open-ended and continually surprising world. Any discipline of thought (learning to programme computers, play in an orchestra, prove

mathematical theorems) restricts the direction in which our imaginations take us; and the ability to work within these restrictions is learned painfully and effortfully. Indeed, I suggest such discipline is difficult precisely because it requires taming and trammelling our sometimes unruly imagination. Our natural 'mode' of thinking is wildly flexible – we only think of discipline and control as the essence of thought because these require our conscious and careful attention. The sheer ubiquity of our imaginative flexibility renders it invisible.

Leaps of the imagination are, on reflection, evident even when solving what we might think of as 'logical' puzzles, which are central to intelligence. Spend a few minutes on the perhaps familiar IQ puzzles below:

(1) *Space* is to *ruler* as *time* is to:

A: Metronome
B: Chronometer
C: Clock
D: Stopwatch

(2) *Sound* is to *echo* as *light* is to

A: Shadow
B: Reflection
C: Refraction
D: Mirror

(3) *Replicate* is to *duplicate* as *divide* is to

A: Split
B: Segment
C: Fractionate
D: Halve

IQ questions of this type are tests of mental elasticity – but also precision. Considering (1), for example, we need to map between time and space. A ruler measures the distance between two points in space. So what measures the distance between two points in time? Aha! – a stopwatch (a clock – which is a good second choice – measures the

time, not the difference between times). But finding this mapping is not easy (and not entirely uncontroversial, perhaps).

Or consider (2): when *sound* bounces back off a surface (e.g. the walls of a ravine or the roof of a cave) we sometimes hear a copy of the original sound: an *echo*. When *light* from an object bounces off a surface (e.g. a mirror, or the surface of a still pond), we sometimes see a copy of that object: a *reflection*. This line of thinking suggests that B is the right answer.

Finally, in (3), when we make one or many copies of a gene, a computer file or a sheet of music (perhaps many times), we *replicate* it. When we create precisely one copy, we *duplicate* the original: we end up with precisely two versions of our gene, file or sheet of music. When we divide a number, we break it into several equal-sized pieces; when we divide it to create precisely two equal sized pieces, we halve it. This line of thinking suggests that D is the right answer.

Notice that solving these puzzles is very different from solving Sudoku or squaring a large number. In open-ended IQ problems, imagination is required even to make the problem meaningful at all. What, after all, is the relationship between the words *space* and *ruler*? A ruler takes up space? A ruler measures spaces? Space has no ruler – i.e. it is not ruled by anyone? Rulers typically have lots of space to live in? These fairly daft possibilities didn't even cross your mind, I suspect. One clue that they would be on the wrong track comes from the fact that we need to find a link between *space* and *time*. So presumably *space* and *time* must be considered quite abstractly. And then we might think of a ruler as a measure of space, but then several of the available options (*chronometer*, *clock*, *stopwatch*) are measures of time – so this doesn't give us a unique answer. A bit more thought might lead to the idea that both rulers and stopwatches measure intervals (in space or time), whereas clocks just give an absolute value. At least this interpretation picks out a unique answer, *stopwatch*, to be paired with *time*. Of course there are endless other ways of picking out a unique answer – though some seem more natural than others.

Suppose, for example, that we reason: a ruler can be used to divide space into equal intervals – and a metronome can be used to divide time into equal intervals – and hence we plump for option A. This

seems a bit of a stretch (another metaphor), because rulers can do many other things than divide space into equal intervals (in particular, a ruler has a measuring scale, which has no analogue in the metronome).

Or, even less convincingly, we notice that the first letters of *space* and *ruler* are consecutive letters in the alphabet: *r* is the letter before *s*. Now time begins with a *t*, so perhaps the related word should begin with the letter of the alphabet before *t*, namely *s*. And indeed, there is a unique option beginning with an *s*, namely *stopwatch*. So we get the same 'right' answer – but using rather tenuous and forced reasoning. After all, the fact that the words are all about time, space and measurement is purely coincidental.

So, unlike Sudoku or arithmetical calculations, for which a solution is precisely defined (and there are, indeed, methods for systematically finding such solutions), these IQ test problems are actually remarkably open-ended. They are capturing our ability to search in the vast space of possible metaphorical links between words (or the ideas they bring to mind), and judging our ability to find natural, sensible mappings rather than perverse and tortured links.

But what, precisely, *counts* as the right answer in analogy problems like these? It is not entirely clear what the right criterion is, but surely the best answer is something like the consensus answer – not the one most of us think of, perhaps, but the one that most of us think is the best, when it is pointed out.[5]

This is not so different from the problem of interpreting faces in found objects. These ambiguous images *could* be given all sorts of interpretations, at a stretch. The handbag in Figure 37 looks like a snarling face to most of us, but could also vaguely remind one of an owl, a rubbish bin, the mouth of a fish, or a beetle. But all interpretations are not equal – again, there is a clear consensus; indeed, once we've seen our 'found faces' *as faces*, it is difficult to see them any other way. Imagination – our ability to construct rich interpretations in complex, open-ended problems – seems to be just what many IQ tests are measuring. Thus the secret of intelligence is imaginative interpretation, rather than 'cold logic'.

But intelligence requires us to harness our imagination in a disciplined way – intelligence requires more than merely blurting out the

first interpretation or metaphor that comes to mind. It turns out to be highly productive to think of gases as consisting of clouds of tiny frictionless billiard balls continually bumping into each other in three dimensions – indeed, this is the standard model in physics. This insight proves to be crucial in understanding how the microscopic properties of trillions of molecules can lead to 'macroscopic' observations of pressure, temperature, volume and so on. Suppose, for example, sticking to two dimensions for a moment, we imagine expanding our billiard table dramatically – the billiard balls will continue to race about and bounce into each other as before, but now they will be much sparser – and hence far fewer balls will thump into the sides of the billiard table at any moment. And the reduced pummelling of the sides of the billiard table corresponds to lower pressure. Or suppose that we have two adjacent billiard tables, one in which a rather sparse set of balls is hurtling around at tremendous speed; and another where a much denser cloud of balls is moving about, on average, much more slowly. If we remove the partition between the two tables, then the density, and the average speeds, of the balls will gradually even out. The speed of the balls corresponds to temperature in the billiard-ball model of gas – so this means that the temperature in the two initially separate 'gases' will gradually become equal as they are joined together; and their density (i.e. the crowdedness of the billiard balls) and pressure (pummelling of the sides of the billiard tables) will reduce to an equilibrium as well.

The billiard-ball model of gases turned out to be incredibly useful, and now serves as the basis for capturing all sorts of aspects of known gas behaviour. It produces a beautiful link between the microscopic behaviour of, not billiard balls, but tiny gas molecules, whizzing about roughly according to Newton's laws of motion (the very same laws we use for projectiles and planets), and the way gases work. Developing and utilizing an analogy like the billiard-ball model of gases requires great care and subtlety. The goal is to create a rigorous, mathematically precise model – not merely a fanciful metaphor – and to test it in experiments to see where it works, and where it fails. Such carefully developed analogies are the foundation of many areas of science.

But to do this, I suggest, doesn't require switching to some entirely

different style of thought – some imagined cold logic with which the products of our imagination can be soberly assessed. Instead, we need to direct successive thoughts to search for alternative interpretations, check that those interpretations make sense, and so on, by 'locking onto' different aspects of the information in the problem. Figuring out unexpected predictions from our analogy, deciding how to map them into known mathematics, and establishing which key experiments we need to conduct, all require ingenuity and inspiration, not mere handle-cranking. Both intelligence and analogy-making, even in science, are driven by our wonderfully elastic imagination, properly harnessed and directed.

THE DISTANT PROSPECT OF INTELLIGENT MACHINES

If our spectacular mental elasticity – our ability to imaginatively interpret complex, open-ended information into rich and varied patterns – is the secret of *human* intelligence, what does this imply for the possibility of *artificial* intelligence (which we began to examine in Chapter 1)?

My suspicion is that the implications are far-reaching. As we saw, the early attempts to extract and codify human 'reasoning' and knowledge into a computer database failed comprehensively. The hoped-for hidden inner principles from which our thoughts and behaviour supposedly flow turned out to be illusory. Instead, human intelligence is based on precedents – and the ability to stretch, mash together and re-engineer such precedents to deal with an open-ended and novel world. The secret of intelligence is the astonishing flexibility and cleverness by which the old is re-engineered to deal with the new. Yet the secret of how this is done has yet to be cracked.

The spectacular progress in computational intelligence over the last half-century has not been achieved by replicating the elasticity of the human imagination, which allows us to see, for instance, in a game of charades, a middle-aged man beating his chest and swatting the air with his fists as a representation of the film *King Kong*, or read rich emotion and humanity into Mooney's black and white patches,

or see the world through an endlessly shifting blizzard of metaphors. Computational intelligence has instead taken a very different tack: focusing on problems, like chess or arithmetic, that require no free interpretation at all, but which can be reduced to vast sequences of calculations, performed at lightning speed. In addition it has proved to be invaluable for things like speech recognition, machine translation and general knowledge tests, hoovering up solutions to almost unimaginably vast quantities of past problems to enable the machine to solve new problems, which are only a little different.[6]

Yet what is astonishing about human intelligence, and perhaps biological intelligence more broadly, is its spectacular flexibility. To us, it is unremarkable that our brains can map J. M. W. Turner's smudges of paint, or Debussy's sinuous orchestration in *La Mer* into a turbulent or calm seascape; that we can interpret the movements of cartoon characters, shadow puppets or ballet dancers as playing out human dramas; that we make instant mental leaps from properties of tensile materials to states of mind (being under strain, stretched, over/under-stretched, pulled in too many directions, spread too thinly, taut, tense, likely to snap, at breaking point, (in)flexible, rigid, stiff, fragile, shattered); that our brains are able to discern links between ripples on water, sound, light, radio, vibrations, the wriggling of a rope and perhaps even the force of gravity itself.

But my suspicion is that it is our mental elasticity that is one of the keys to what makes human intelligence so remarkable and so distinctive.[7] The creative and often wildly metaphorical interpretations that we impose on the world are far removed from anything we have yet been able to replicate by machine.

To those who, like me, are fascinated by the possibilities of artificial intelligence, the moral is that we should expect further automation of those mental activities that can be solved by 'brute force' rather than mental elasticity – the routine, the repetitive, the well defined. This is part of a trend that started with the development of stone tools in the Olduvai Gorge in present-day Tanzania, more than two and a half million years ago and accelerated spectacularly in the Industrial Revolution: people and technology can achieve far more than people alone. It is a continual source of amazement to us that tasks that seemed to require the full power of human ingenuity can be solved,

often far more efficiently, by processes of standardization and mechanization: the flexibility and dexterity of human hand-weaving could, in many cases, be replaced by the precision achieved by the hand-loom, and then the steam-driven Jacquard loom, controlled by punched cards around 1800, and on to the phenomenal productivity of the computerized power looms of today. At each step, the environment is made more precise and more standardized; and more of the task can be handed over to machines.

In the same way, the rise of digitization and big data can create an ever more frictionless and more precisely defined world in which computers can operate far better than we can. But the secret of human intelligence is the ability to find patterns in the least structured, most unexpected, hugely variable of streams of information – to lock onto a handbag and see a snarling face; to lock onto a set of black-and-white patches and discern a distinctive, emotion-laden, human being; to find mappings and metaphors through the complexity and chaos of the physical and psychological worlds. All this is far beyond the reach of modern artificial intelligence.

It is our wild imposition of meaning upon the world by the appropriation and transformation of past experience that is the essence of human thought, from which our more sober reflections can, with difficulty, be constructed. Watching the mind at play is our best guide to its natural mode of operation. This reveals the search for interpretation that drives us – the effort to find meaning that is channelled by, rather than replaced with, step-by-step conscious deliberation.

To those who fear the march of the machines, this should be of some comfort. If imagination and metaphor is the secret of our intelligence, then that secret may, perhaps, be safely locked away in the human brain for centuries and perhaps for ever.

Epilogue: Reinventing Ourselves

We have all been victims of a hoax, perpetrated on us by our own brains. Our brains are spectacular engines of improvisation that can, in the moment, generate a colour, an object, a memory, a belief or a preference, spin a story, or reel off a justification. And it is such a compelling storyteller that we are fooled into thinking that it is not inventing our thoughts 'in the moment' at all, but fishing them from some deep inner sea of pre-formed colours, objects, memories, beliefs or preferences, of which our conscious thoughts are merely the shimmering surface. But our mental depths are a confabulation – a fiction created in the moment by our own brain. There are no pre-formed beliefs, desires, preferences, attitudes, even memories, hidden in the deep recesses of the mind; indeed, the mind has no deep recesses in which anything can hide. The mind is flat: the surface is all there is.

Our brains are, then, relentless and compelling improvisers, creating the mind, moment by moment. But, as with any improvisation – in dance, music or storytelling – each fresh thought is not created out of nothing. Each fresh improvisation is built from the fragments of past improvisations – so each of us is a unique history, together with a wonderfully creative machine for redeploying that history to create new perceptions, thoughts, emotions and stories. The layering of that history makes some patterns of thought natural for us, others awkward or uncomfortable. But while drawing on our past, we are, none the less, continually reinventing ourselves; and by directing that reinvention, we can shape who we are, and who we will become.

So we are not driven by hidden, inexorable forces from a dark and subterranean mental world. Instead, our thoughts and actions are transformations of past thoughts and actions, and we often have

considerable latitude, and a certain judicial discretion, regarding which precedents we consider, which transformations we allow. As today's thought or action is tomorrow's precedent, we are, quite literally, reshaping and reinventing ourselves thought by thought.

This is not a familiar story. It is not an intuitive story. It is a story that shakes our faith in everything we think we know about how our minds work – including what we 'see' and 'feel', the nature of memory, decision-making and personality. It is a story in which we cannot find ourselves, because there are no selves to find. So yes, it is a story of a hoax, a conspiracy, an illusion, which enfolds us so completely that we cannot see behind its veil – or even notice that there is a veil we might wish to lift. But it is also a story of how more than a century of scientific investigation of the mind has, step by step, allowed the hoax to be uncovered. Once the spell of the inner worlds, true selves, mental depths and unconscious mental forces is broken, we can see ourselves in a clearer light: we are astonishingly inventive ad hoc reasoners, creative metaphor-machines, continually welding together scattered scraps of information into momentarily coherent wholes. We are very different from the image we create for ourselves, and much more remarkable.

This is all very well, you may say. But surely we need beliefs and motives to explain why our thoughts and behaviour make sense, rather than being a completely incoherent jumble. Surely there are crucial inner facts about us, large and small, that set the course of our actions: the things we value, the ideals we believe in, the passions that move us. But if the mind is flat, despite the stories we tell about ourselves and each other, beliefs and motives cannot, in reality, be driving our behaviour – because there are no inner beliefs, and motives are a projection rather than a reality.

On the other hand, layers of precedents – the successive adaptation and transformation of previous thoughts and actions to create new thoughts and actions – can provide a very different, and more convincing, explanation for the orderly (and, on occasions, the disorderly) nature of thought. Moreover, moving from individual minds to society as a whole, our culture can be viewed as a shared canon of precedents – things we do, want, say or think – which create order in society as well as within each individual. By laying down new precedents, we

incrementally and collectively create our culture. But our new precedents are based on old, shared precedents, so that our culture also creates us. Considered in isolation, our 'selves' turn out to be partial, fragmentary and alarmingly fragile; we are only the most lightly sketched of literary creations. Yet collectively we can construct lives, organizations and societies which can be remarkably stable and coherent.

The idea of continuous reinvention is particularly vertigo-inducing when we realize that, once the hoax is uncovered, the very idea of an objective, and external, yardstick against which to judge our behaviour as individuals, and as a society, is not just impractical, but also wholly unsustainable. There is, after all, no solid foundation upon which we can build. New thoughts, values and actions can only be justified or criticized from within a tradition of past precedents. Of course, which precedents should be applied, and which should dominate, may be matters of dispute, just as they are in the law. This does not mean that anything goes – but it does mean that the construction of our lives and our society is an inherently open-ended, creative process, and that the standards by which we judge our decisions and actions are part of that same creative process. In short, life is a game in which we play, invent the rules and keep the score ourselves.

This perspective might appear to be a recipe for a relativist nightmare: that any point of view is equally valid or equally dubious. But the reverse is true. If there are no ultimate foundations on which the good life, or the good society, is based, then the challenge in our own lives and societies is to explore and resolve conflicts in thought, whether within ourselves or between individuals. While holding onto old precedents can be seen as a conservative political impulse, the aim of bringing distinct traditions into contact, and ultimately perhaps into alignment, is the impulse behind the liberal political tradition. For example, the principle of freedom of speech is designed to allow public debate to link together momentary beliefs across time, individuals and communities; and such debate may be codified by mathematical and scientific methods. Free markets, money, trade and the modern economic system connect together our preferences by means of the exchange of goods, services and money. Democratic politics and the rule of law can resolve potential clashes between our actions (where a decision at one moment may encroach on a decision

at some future moment, by the same person or another). In a liberal society, then, we do not merely dream our own individual dream or write our own particular story; rather, we continually strive to link our stories together into a single cohesive whole.

Yet, even in a liberal society in which preferences, beliefs and actions can be connected, precedent-based thinking appears inherently conservative. How, then, are perceptual reorganizations, sudden insights, religious conversions and conceptual and political revolutions possible? One answer is that memory is fragile; so that we can often end up starting again from scratch, and coming up with a different answer. We forget the old 'story' and create a new one.

But there is another possibility: that changing *part* of the story can lead to a cascade of far-reaching consequences. For all its focus on the authority of the past, we should remember that successions of precedents can, step by step, lead to remarkable metamorphoses. Legal and political systems can transform themselves over the generations, though at each step guided by the interpretation and reinterpretation of what went before. A mathematician may use past precedents in methods of reasoning in order to show that an entire theory is inconsistent – and may even conclude that a great swathe of apparent precedents (previous 'results' from that theory) must be abandoned. Individuals may gradually begin, or cease, to trust a cult leader, or a religious or political text; they may take up, or abandon, a cause, a project, a relationship – or in a million other ways transform their lives So what counts as a precedent can itself shift – and shift again. We hope to stumble towards better and better 'stories' – but we can only create new stories by starting with the stories we have.

We should, moreover, never assume that achieving greater coherence necessarily implies cultural or intellectual progress. And we must be ever-vigilant against allowing ourselves or our societies to ossify into coherent, but crushing systems of precedent. But we should also remember that we are not hemmed in by occult psychic forces within us: any 'prisons' of thought are of our own invention, and can be dismantled just as they have been constructed. If the mind is flat – if we imagine our minds, our lives and our culture – we have the power to imagine an inspiring future, and to make it real.