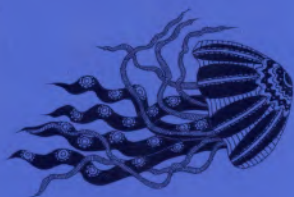


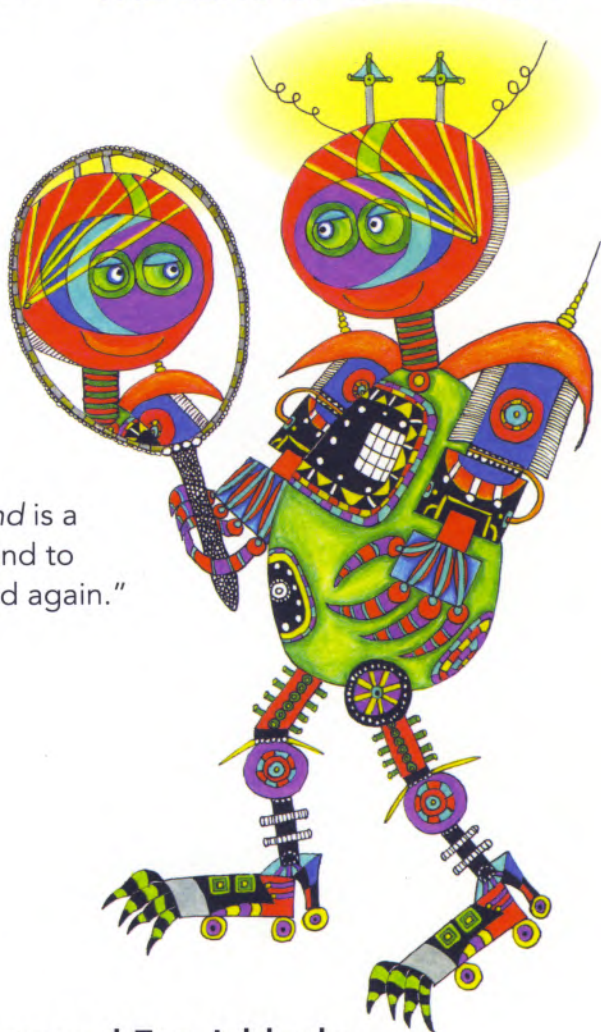


PICTURING THE MIND

CONSCIOUSNESS THROUGH
THE LENS OF EVOLUTION



*"Picturing the Mind is a
work to treasure and to
return to again and again."*
—Anil Seth



Simona Ginsburg and Eva Jablonka
illustrated by Anna Zeligowski

PICTURING THE MIND CONSCIOUSNESS THROUGH THE LENS OF EVOLUTION

Simona Ginsburg and Eva Jablonka
illustrated by Anna Zeligowski



What is consciousness, and who (or what) is conscious—humans, nonhumans, non-living beings? How did consciousness evolve? *Picturing the Mind* pursues these questions through a series of “vistas”—short, engaging texts by Simona Ginsburg and Eva Jablonka, accompanied by Anna Zeligowski’s lively illustrations. Taking an evolutionary perspective, Ginsburg and Jablonka suggest that consciousness can take many forms and is found not only in humans but even in such animals as octopuses (who seem to express emotions by changing color) and bees (who socialize with other bees). They identify the possible evolutionary marker of the transition from nonconscious to conscious animals, and they speculate intriguingly about aliens and artificial intelligence.

Simona Ginsburg is Associate Professor at the Open University of Israel, where she developed and headed the MA Program in Biological Thought. **Eva Jablonka** is Professor at the Cohn Institute for the History and Philosophy of Science and Ideas at Tel Aviv University. Ginsburg and Jablonka are coauthors of *The Evolution of the Sensitive Soul: Learning and the Origins of Consciousness* (MIT Press). **Anna Zeligowski** is an artist and physician. Her illustrations have appeared in numerous books and articles on scientific subjects.

“Two of the central voices of evolutionary consciousness science present a remarkable work about the mind and its embodiments. Clear exposition of deep concepts, many new ideas, and incredible artwork will move readers on many levels. Most highly recommended.”

—**Michael Levin**, Distinguished Professor and Vannevar Bush Chair, Department of Biology, and Director of the Allen Discovery Center, Tufts University

“A unique and beautiful book, full of insights and delights. This rich mixture of science, art, philosophy, and poetry provokes many questions, and offers some answers, about the deep mystery of consciousness. *Picturing the Mind* is a work to treasure and to return to again and again.”

—**Anil Seth**, author of *Being You: A New Science of Consciousness*

The MIT Press

Massachusetts Institute of Technology
Cambridge, Massachusetts 02142
<http://mitpress.mit.edu>

US \$39.95 / \$53.95 CAN

ISBN 978-0-262-04675-6



9 780262 046756

5 3 9 9 5



OCRLab

 <http://asenic.ru/ocrlab>

PICTURING THE MIND

CONSCIOUSNESS THROUGH THE LENS OF EVOLUTION

SIMONA GINSBURG AND EVA JABLONKA

ILLUSTRATED BY ANNA ZELIGOWSKI

The MIT Press
Cambridge, Massachusetts
London, England

© 2022 Massachusetts Institute of Technology

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Stone Serif and Stone Sans by New Best-set Typesetters Ltd. Printed and bound in the United States of America.

The Ogden Nash poem “The Octopus” is from the book *Candy and Dandy*, 1994, Welbeck Publishing, London. © 1940 by Ogden Nash, renewed. Reprinted by permission of Curtis Brown Ltd.

Library of Congress Cataloging-in-Publication Data

Names: Ginsburg, Simona, author. | Jablonka, Eva, author. | Zeligowski, Anna, illustrator.

Title: Picturing the mind : consciousness through the lens of evolution / Simona Ginsburg and Eva Jablonka ; illustrated by Anna Zeligowski.

Description: Cambridge : The MIT Press, 2022. | Includes index.

Identifiers: LCCN 2021013277 | ISBN 9780262046756 (hardcover)

Subjects: LCSH: Intellect. | Consciousness. | Consciousness—Physiological aspects. | Evolution (Biology)

Classification: LCC BF431 .G5346 2022 | DDC 153.9/3—dc23

LC record available at <https://lcn.loc.gov/2021013277>

10 9 8 7 6 5 4 3 2 1

CONTENTS

PREFACE ix

VISTA 1 CONSCIOUSNESS: METAPHORS AND CONCEPTIONS 1

- 1.1 What Is the Mind? 4
- 1.2 Streams, Waves, Birds 6
- 1.3 The Butterfly 8
- 1.4 Dualism 10
- 1.5 Panpsychism 12
- 1.6 Physicalism 14
- 1.7 Aristotelian Naturalism 16
- 1.8 What Is It Like to Be a Bat? 20
- 1.9 Ways of Knowing 22
- 1.10 Self: The Hole in The Doughnut 24
- 1.11 Are You a Brain in a Vat? 26
- 1.12 Philosophical Zombies? 30

VISTA 2 WHO IS CONSCIOUS? 33

- 2.1 A View from Biology 34
- 2.2 Are Bacteria Conscious? 36
- 2.3 Conscious Slime Molds? 38
- 2.4 Are Plants Conscious? 40
- 2.5 Are Sponges Conscious? 42
- 2.6 What's So Special about a Nervous System? 44
- 2.7 Enchanting Medusae, Graceful Sea Anemones 46
- 2.8 Brainy Worms 50
- 2.9 A Fuzzy Consensus 54
- 2.10 Are Fish Conscious? 56

- 2.11 Are Reptiles Conscious? 60
- 2.12 Birds and Mammals 64
- 2.13 Some Arthropods? 68
- 2.14 What about Cephalopods? 72
- 2.15 Are Only Humans Conscious? 74
- 2.16 What Are the Social Dimensions of Consciousness? 76

VISTA 3 HOW DID CONSCIOUSNESS EVOLVE? 79

- 3.1 Why Take an Evolutionary Approach to Consciousness? 82
- 3.2 Evolutionary Theory 84
- 3.3 Evolutionary Transitions 88
- 3.4 An Evolutionary-Transition Approach to Consciousness 92
- 3.5 The Transition to Life 94
- 3.6 The Transition to Consciousness: Compiling a List 96
- 3.7 Learning and Consciousness: An Unlikely Relation? 100
- 3.8 Unlimited Associative Learning (UAL) Is the Evolutionary Transition Marker of Consciousness 104
- 3.9 Feelings and Conceptions 108
- 3.10 The Functions and Goals of Consciousness 112
- 3.11 Which Organisms Manifest UAL? Who Is Conscious? 114
- 3.12 The Big Bang of Animal Evolution 118
- 3.13 The Origins of Suffering 122
- 3.14 The Evolution of Imagination 124

VISTA 4 THE PREEMINENCE OF MAN? THE RATIONAL SOUL 127

- 4.1 Human Nature 130
- 4.2 What Sets Humans Apart? 134
- 4.3 Why Take an Evolutionary Approach to Human Nature? 138
- 4.4 Before *Sapiens* 140
- 4.5 The Blushing *Erectus* 144
- 4.6 The Extended Mind: Hands, Tools, Memorials 148
- 4.7 The Symbolic Dimension 152
- 4.8 Language and Imagination 156
- 4.9 Thoughts and Feelings 158
- 4.10 On Some Peculiarities of the Symbolic Species 162
- 4.11 The Symbolic Explosion 166
- 4.12 A Split Soul? 170

VISTA 5 VISIONS, FUTURES, FANTASIES 173

- 5.1 Stretching the Limits 176
- 5.2 Savant Minds 178
- 5.3 Perceptual Memory and the Art of Forgetting 182
- 5.4 Consciousness Upheavals: Psychedelic Drugs 186

5.5	Mental Pains: Images of the Unconscious	190
5.6	“Higher” Forms of Consciousness?	194
5.7	How Did Our Ancestors Perceive the World?	198
5.8	Telling Artifacts	202
5.9	Conscious Robots?	206
5.10	Virtual and Cyborg Realities	210
5.11	Extended Ethics	216
5.12	Solaris: The Limits	220

ACKNOWLEDGMENTS	223
-----------------	-----

NOTES	225
-------	-----

INDEX	243
-------	-----

To Marion. For everything.

PREFACE

Being alive yet unable to experience—unable to feel, move, smell, see, hear, taste, or touch, like people in a deep coma—seems to most of us a life not worth living. It makes us acutely aware that our cherished and intimate capacity for subjective experiencing, which is also referred to as *consciousness* or *sentience* (we use these terms interchangeably), cannot be taken for granted. What is it that makes an entity, a living body, conscious? What happens when the state of mind we call consciousness is lost?

Philosophers call the puzzling relation between the mind and the body the mind–body problem. Once the brain was found to be essential for conscious activity, the mind–body problem was reframed in terms of the relation between the brain and the mind. Here is the vivid way that T. H. Huxley expressed the difficulty of understanding this relation in 1866:

How it is that anything so remarkable as a state of consciousness comes about as the result of irritating nervous tissue, is just as unaccountable as the appearance of the Djin when Aladdin rubbed his lamp.

We use evolutionary biology as our Ariadne’s thread to lead us through the maze that is the mind–body problem. Evolutionary biology leads us to the questions that run through the book: Which entities are conscious? How did consciousness evolve? Which varieties of consciousness do we recognize? Is human consciousness special, and if it is, in what ways? Can we envisage alien or artificial forms of consciousness? We present some of the views of writers, poets, philosophers, psychologists, and biologists, but since their ideas breed even more questions, the book is focused more on questions than on answers.

We use visual and verbal images as well as explanatory texts to present these questions and some tentative answers. The richness and inherent ambiguity of images and metaphors allow a wide scope for imagination and interpretation. They are vistas through which we look at the many aspects of consciousness. Each page of text, which presents a specific topic, faces a picture, a visual metaphor that engages with the text and with the reader's imagination, and opens up additional perspectives. Every topic is both autonomous and integrated with the preceding and subsequent text, so the book can be read either "in bits" or as a continuous picture-text. Sources for quotations and further reading about each topic are presented in notes at the end of the book.

The book has five parts, or vistas, each discussing twelve to sixteen topics. The first vista introduces some of the ways in which people have imagined consciousness and conceived of the nature of the mind. We start with metaphors that capture aspects of consciousness, go on to describe general approaches such as dualism and physicalism, and end with puzzles and thought experiments that illuminate its seemingly paradoxical nature.

Our own naturalistic evolutionary approach focuses on living organisms. In the second vista we ask: Are all living beings, including bacteria, conscious? Are plants conscious? Are only humans conscious? We leave these questions open, but in the third vista we present an evolutionary approach that provides a way of addressing them, an approach focusing on the transition from non-conscious to conscious animals. In this vista we suggest that the evolution of consciousness was entailed by the evolution of an open-ended form of learning that can be found in some brainy animals. Some of the texts that develop this suggestion may be difficult for nonbiologists, but they can be safely skipped without losing track of the overall message. Human consciousness—its uniqueness, origins, and wondrous and monstrous consequences—is described in the fourth vista, while the fifth vista presents some of the varieties of human consciousness that allow us to go beyond the boundaries of the familiar and explore new forms of future and alien consciousness through visions and fantasies.

The book is intended for all people who are fascinated by consciousness, from curious high school students to retired professors. We hope that the combination of art, philosophy, and science will engage people's imagination and reveal the many ways of exploring the landscapes of the mind.

VISTA 1 CONSCIOUSNESS: METAPHORS AND CONCEPTIONS

When I turn my gaze skyward I see the flattened dome of sky and the sun's brilliant disc and a hundred other visible things underneath it. What are the steps which bring this about? A pencil of light from the sun enters the eye and is focused there on the retina. It gives rise to a change, which in turn travels to the nerve-layer at the top of the brain. The whole chain of these events, from the sun to the top of my brain, is physical. Each step is an electrical reaction. But now there succeeds a change wholly unlike any which led up to it, and wholly inexplicable by us. A visual scene presents itself to the mind; I *see* the dome of sky and the sun in it, and a hundred other visual things beside.

. . . It is a far cry from an electrical reaction in the brain to suddenly seeing the world around one, with all its distances, its colours and chiaroscuro.

—Sir Charles Scott Sherrington

1.1 WHAT IS THE MIND?

Ikkyū Sōjun, a fifteenth-century Japanese Zen Buddhist monk, heretic, vagabond, free spirit, and poet, offers one of the most philosophically powerful images of the mind:

What is the mind?
It is the sound of the breeze
That passes through the pines
In the Indian-ink picture

The poem depicts the mind as a high-level representation: the words of the poem, themselves representations, ask the reader to identify the mind with the imagined sound of a breeze that passes through imagined pines in an imagined Indian-ink picture. To these levels of representation, the artist added yet more levels, representing the pine tree as a calligram made up of the words of the poem in English and Hebrew.

1.2 STREAMS, WAVES, BIRDS

William James, the nineteenth-century philosopher and psychologist, was a master of metaphors.

In 1884 he described consciousness or subjective experiencing as a flow of water, a stream consisting of waxing and waning waves, with different rates and paces of activity, with echoes of the past and a dawning sense of the future. Experiencing, James thought, is an activity that is always personal, always changing yet continuous.

Every definite image in the mind is steeped and dyed in the free water that flows round it. With it goes the sense of its relations, near and remote, the dying echo of whence it came to us, the dawning sense of whither it is to lead.

He described the stream of consciousness as higher at its end than at its beginning “because the final way of feeling the content is fuller and richer than the initial way” (James, *The Principles of Psychology*, 1890). He then offered yet another metaphor:

As we take . . . a general view of the wonderful stream of our consciousness, what strikes us first is this different pace of its parts. Like a bird’s life, it seems to be made of an alternation of flights and perchings.

Birds symbolize departed souls, or, in the Christian tradition, saved souls. James offered a metaphor of moving streams, waves, and flying-perching birds, which represent the material activity that constitutes the mind.



The stream of consciousness: A turbulent stream, an alternation of flying and perching birds.

1.3 THE BUTTERFLY

One metaphor for the mind is the butterfly. Aristotle gave the butterfly the name *psyche*, the Greek word for “soul.” In many cultures, butterflies—beautiful, flying products of metamorphosis, creatures of air and sky—are symbols of the human soul, of freedom from flesh and toil, of spiritual transformation.

It is probably not surprising that Zhuang Zhou, the traditional author of the *Zhuangzi*, written in the fourth century BCE, which is considered one of the greatest literary works in all of Chinese history, uses the image of the butterfly to wonder about the distinctions and transformations of consciousness:

Once, Zhuang Zhou dreamed that he was a butterfly, flitting and fluttering like other butterflies, enjoying himself very much. He was completely unaware that he was Zhuang Zhou.

Suddenly he woke up, and knew, vividly and certainly, that he was Zhuang Zhou. He was not clear, however, whether he was Zhuang Zhou who just dreamed he was a butterfly or a butterfly that is dreaming that he is Zhuang Zhou. Nevertheless, there must be some difference between Zhuang Zhou and that butterfly. This is the so-called distinction between things.

What are the “things” that are being distinguished? Are they physical? Are they spiritual?



... a man dreaming that he is a butterfly dreaming that he is a man dreaming that he is a butterfly ...

1.4 DUALISM

Dualists assume that the mental (mind) is not identical with the physical (body). Some maintain that the physical and the mental are two kinds of “stuff” that pervade all substances, while others think that only rational human thinking requires additional (mental) stuff. René Descartes is best known for expounding the rationalistic dualistic position, which, for him started from the intellectual activity of doubting:

I will suppose, then, that everything I see is fictitious. I will believe that my memory tells me nothing but lies. I have no senses. Body, shape, extension, movement and place are illusions. So what remains true? Perhaps just the one fact that nothing is certain!

Descartes concluded that the only certain thing that he cannot doubt is doubting, for even doubting that he doubts would still be doubting. Doubting is an act of thinking, and therefore thinking proves the doubter’s existence. Hence: *Cogito ergo sum*—I think, therefore I am.

He drew from this train of reasoning a far-reaching conclusion: he assumed that the attributes of a thinking substance, a mind, are incorporeal, nonspatial, and indivisible, and therefore must be immortal. Only reasoning humans have this immortal thinking soul. It is located, he believed, in the pineal gland in the brain, and it interacts with the body. Animals, which lack a rational soul, are mortal and machine-like, although such machines may express and feel emotions.

Many dualists, past and present, extend the assumption that there are mental and physical “stuffs” or aspects of consciousness in all substances.

For those who follow Descartes, the biggest problem is: How, if one believes that the mind and body are qualitatively different kinds of stuffs or properties, do the mind and body interact? How is it that a state of mind (e.g., sadness) causes a physical state? How, conversely, can physical changes in the body (e.g., those caused by the consumption of a drug) cause changes in consciousness? And what is the “self”? Is it an attribute of the body, or the mind?

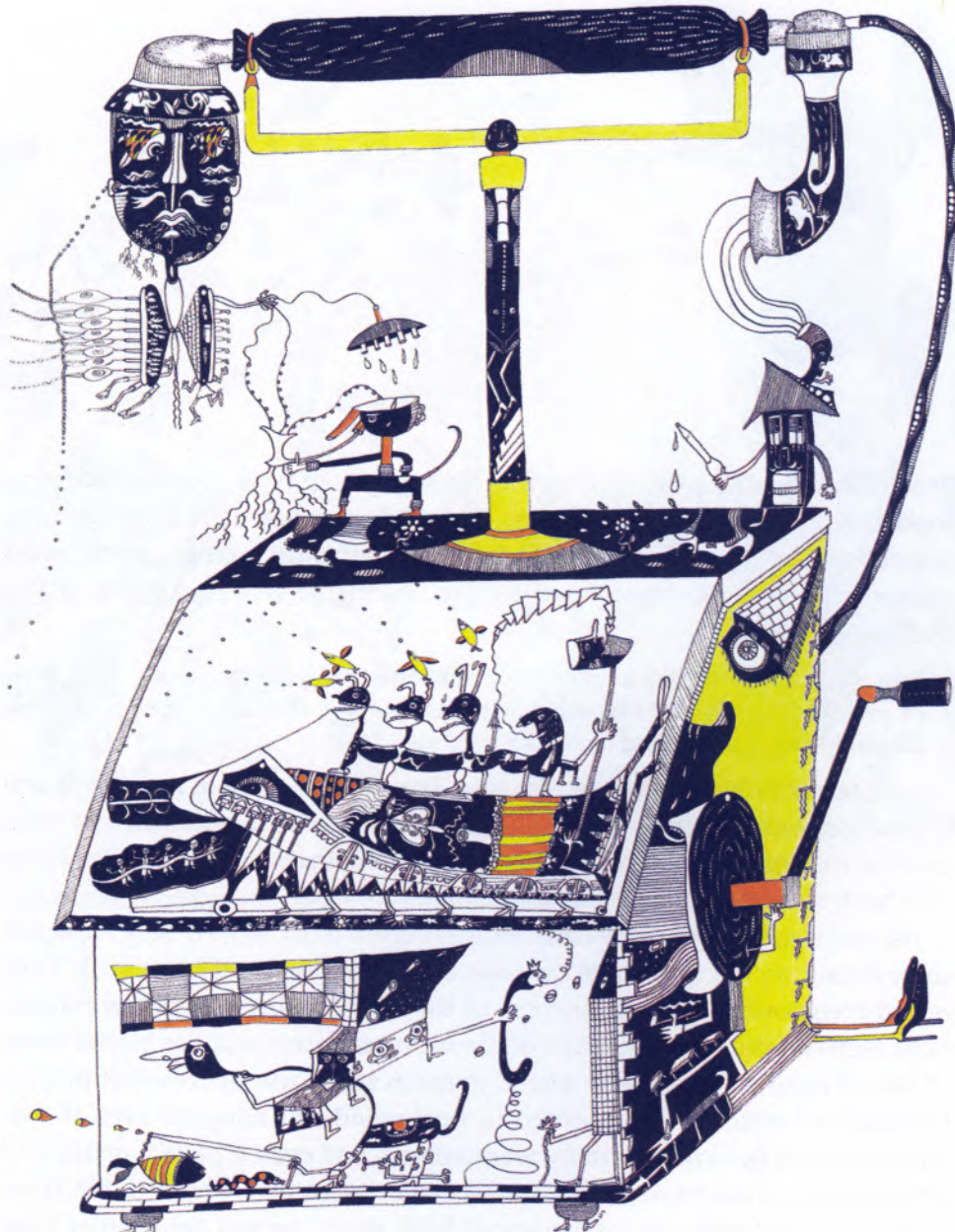


Rays of spirit emanating from his pineal gland render the little robot self-aware.

1.5 PANPSYCHISM

Panpsychists claim that sentience (the ability to feel and perceive) is a fundamental and pervasive feature of all things. They believe that there is one substance that is both fundamentally physical and fundamentally mental. They assume that there are intrinsic properties of things that cannot be understood in terms of physics. There is, to use a Buddhist term, a “suchness” or “thatness” to things, which is fundamental to their being and is immaterial. This is what Buddha’s famous wordless “Flower Sermon” tried to convey. Smiling, not saying a word, Buddha held a white flower in front of his disciples. In this wordless sermon he communicated to them the ineffable nature of the flower, and by extension, of all things.

Panpsychism takes several forms. While some panpsychists argue that everything has a degree of subjective experiencing (including your cup of tea), others think that there are two types of experiencing: micro and macro. The micro-experiencing of elementary particles like quarks is different from the macro-experiencing of a toothache. Many micro-experiences interact to form macro-experiences following as yet unknown principles. But how, one wonders, do the little micro-experiences of, say, the elementary particles in your rotting tooth, combine to form the macro-experiencing of a toothache? And what does it mean to be unconscious (e.g., in a deep coma), according to panpsychists?



The inner world of an ancient telephone.

1.6 PHYSICALISM

Physicalists identify mental properties with physical, typically neural or neurophysiological, states and processes. They assume that there is no separate mental stuff or aspect, and regard the claim that consciousness pervades all matter as scientifically void. Francis Crick articulated a clear example of this position in 1994:

You, your joys and your sorrows, your memories and your ambitions, your sense of personal identity and free will, are in fact no more than the behavior of a vast assembly of nerve cells and their associated molecules.

This statement hides many complexities. The vast assembly of nerve cells and their associated molecules is, as Crick himself assumed, an *organized* assembly, so what kind of organization is it? Can isolated brains be conscious? Are nerve cells both necessary and sufficient for consciousness?

Neuroscientists who study consciousness often differ in how they approach these questions, but they share the same broad philosophical framework. They regard consciousness as the outcome of the self-organizing, dynamic interactions between the low-level parts of the nervous system and the higher levels of neural organization. These neural dynamics give rise to coherent images, feelings, and actions. The question is, how could this happen? Even if consciousness can be experimentally manipulated, and even if perfect neural and other bodily correlates of consciousness can be identified, why is it that these correlates entail subjective experience? Why does one feel pain rather than just adaptively respond to a damaging stimulus? In other words, why are we not zombies? Why do we feel that our mental activities are qualitatively different from our physical activities?



The endless ramifications of 1,500 grams of matter.

1.7 ARISTOTELIAN NATURALISM

Aristotelian naturalists are physicalists who regard consciousness as a system-wide activity that is *constituted* by biological-cognitive processes. Just as liver cells are the building blocks of the liver and its activities, so nerve cells and biological-cognitive processes are the building blocks of conscious activities. The dynamic *organization* of the system is all-important. As they see it, only material beings with a particular type of organization can be considered conscious. According to this naturalist view, no single part (e.g., the brain), however necessary its function, is sufficient for consciousness. It is the organism as a whole that is conscious.

Aristotle, who emphasized the goal-directed nature of living organisms, can be seen as the first philosopher to offer a systematic naturalistic view of sentience. He called the goal-directed, active organization of living beings the “soul,” and distinguished between three levels of the soul.

The first, most basic level characterizes the simplest living beings. The intrinsic goal of such beings is self-maintenance and reproduction. Plants were Aristotle’s example of beings with only this basic level of soul, which he called the “nutritive soul” or the “reproductive soul.” The second level, which builds on the first, characterizes sentient animals that can have perceptions and feelings. Living beings with such capacities have a “sensitive soul,” and their goal is to satisfy their felt desires, passions, and wants. The third level, the “rational soul,” is built on the first two. Human values such as “goodness,” “truth,” and “beauty” are the goals of the rational soul, which humans strive to satisfy. Aristotle wrote:



Aristotle's three levels of the soul.

Plants have none but the first, the nutritive, while another order of living things has this plus the sensory. If any order of living things has the sensory, it must also have the appetitive; for appetite is the genus of which desire, passion, and wish are the species. . . . Lastly, certain living beings—a small minority—possess calculation and thought, for (among mortal beings) those which possess calculation have all the other powers above mentioned.

Aristotelian naturalism raises new problems for biologists and philosophers. If consciousness is an attribute of animals, which animals are conscious? And since the sensitive soul depends on the nutritive-reproductive soul, can there be consciousness without life? For example, can nonliving robots be conscious? Finally, can we find, within the Aristotelian hierarchy of souls, principles that are common to all souls, to all three goal-directed modes of being—living, “sensitive,” and rational organisms?

1.8 WHAT IS IT LIKE TO BE A BAT?

How can we fathom the experiences of others, especially of organisms that are very different from us? If we assume that bats have private subjective experiences, can we ever access them? In an influential paper titled “What Is It Like to Be a Bat?,” the philosopher Thomas Nagel argued that it is impossible to scientifically explain subjective experience. His example was a bat that senses the world through echolocation, a sense we lack. Nagel argued that we cannot share in the subjectivity of a bat: we cannot feel its feelings or experience its perceptions *as* a bat.

Is Nagel right? Let us look at a remarkable man, Daniel Kish. Kish, who was born in 1966, lost his eyesight to retinoblastoma when he was thirteen months old. Yet he navigates his way through city streets, rides mountain bikes, plays basketball, can find a small ball on a large golf course, and teaches other people who are blind—mainly youngsters—to do the same. He does all this because he echolocates, just as some bats do, by making clicking noises that return as echoes and inform him about the shape, position, size, texture, and motion of objects. Studies of Kish and another echolocator have shown that part of the brain, the visual cortex, is activated when the echoes are picked up and as the echolocators report what they perceive on the basis of the echoes. Their experiences are of a rich, unified space. According to Kish: “It’s not computational. There’s a real palpable experience of the image as a spatial representation—here are walls, here are the corners, here is the presence of objects.” Elsewhere he added, “The sense of imagery is very rich for an experienced user. One can get a sense of beauty or starkness or whatever—from sound as well as echoes.”

One can rightly claim that there is a big difference between the world experienced by human echolocators and that experienced by bats. But since we can acquire some of the bat's sensory experience, have we not progressed somewhat in our understanding of what it is like to be a bat?

The picture of Kish riding his mountain bike captures an aspect of his world. While we take our vision as a passive reflection of the world out there, Kish's echolocation is a self-consciously active process through which he encounters responsive objects.



Navigating by sounds and echoes.

1.9 WAYS OF KNOWING

What is the relation between the private, acute pain that you feel subjectively after you have slipped a disk and the public, textbook description of such pain?

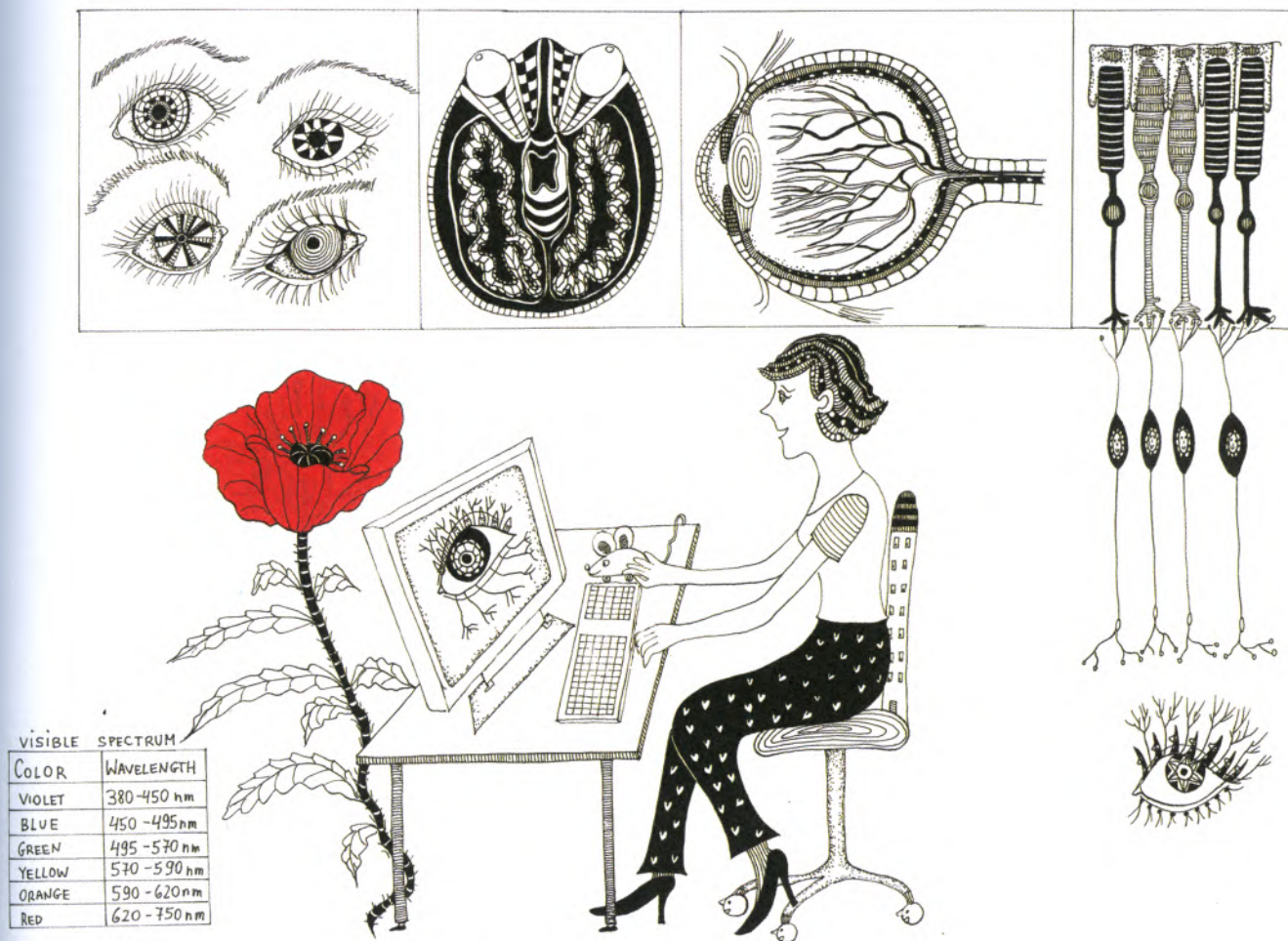
We can generalize and abstract from private to public knowledge, but can public knowledge be transformed into private knowledge? Great novels, which are forms of public knowledge, make us experience things that are distant from us in content, time, and space. A woman can feel what it is like to be a young tormented prince in a Shakespeare play; a man can feel what it is like to be an imperial lady-in-waiting in the emperor's court in a tenth-century Japanese novel. But how far can such transformations go?

We know that we can subjectively share in the feelings of others. In biblical Hebrew, the word “know” was used to describe not just abstract knowledge, but also the act of sexual love, as in “and Adam knew his wife Eve” (Genesis 4:1). In sexually shared love, a lover gets to subjectively “know” the experiences of the loved one.

What about scientific knowledge? In a famous thought experiment, Frank Jackson described Mary, a brilliant neuroscientist, who has *perfect* knowledge of the physics and neurobiology of color vision. However, although she has a perfectly normal capacity to see color, she has been confined from birth to a black-and-white world, so she has never been exposed to any color. What, asks Jackson, will happen “when Mary is released from her black and white room or is given a color television monitor? Will she *learn* anything or not?”

Our common sense tells us that Mary will acquire new knowledge—she will have a *direct* subjective perception of, for example, red. But is this assumption valid? It is based on the supposition that her perfect theoretical and technical

knowledge could not enable her to simulate the effects of absorbing wavelength between 635–700 nm in her own brain. But why should we assume that? The problem may be just a failure of imagination—we may simply fail to imagine what perfect knowledge and potent technology can achieve. There is no *a priori* reason to assume that with perfect knowledge and technology, Mary, enclosed within her black-and-white world, would not be able to implement color in her own brain and directly experience the redness of a poppy.



Constructing an experience of red.

1.10 SELF: THE HOLE IN THE DOUGHNUT

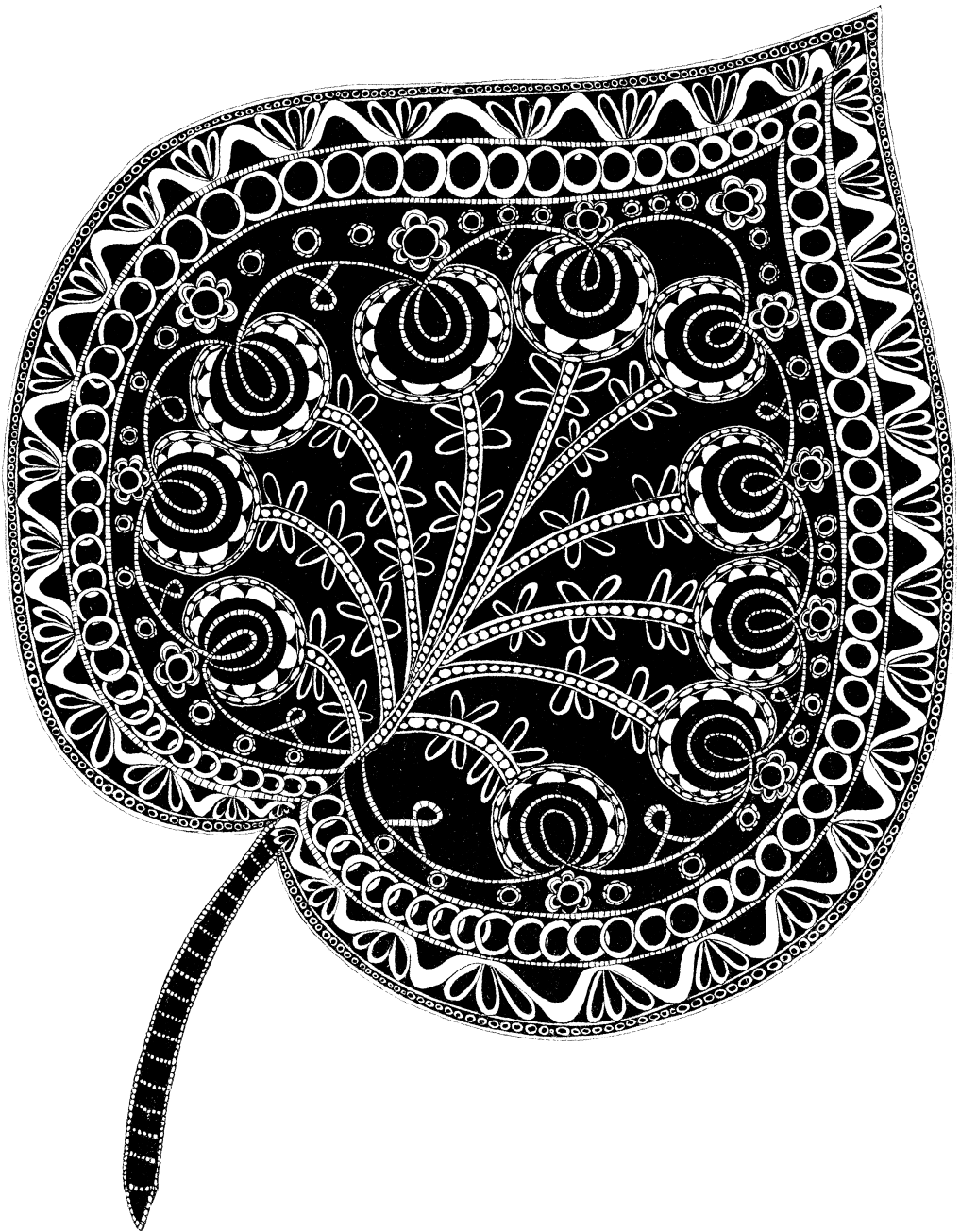
What is the self? The philosopher David Hume thought that his own self, like that of other people, was just a patchy flux of perceptions:

I may venture to affirm of the rest of mankind, that they are nothing but a bundle or collection of different perceptions, which succeed each other with an inconceivable rapidity, and are in a perpetual flux and movement.

Most natural philosophers agree with Hume. However, there must be some stable reference point in this flux that can account for our feeling of personal continuity and our feeling that we *own* our experiences, that they are *ours*. Bjoern Merker, a present-day cognitive scientist, has suggested that there is an “ego-center,” a point of view from which we perceive, evaluate, and act upon an image of our body that is embedded within an image of the world. Like the center of a vortex, our ego-center is constructed by the processes that constitute our world, body, and action models.

We think that these words of the poet Osip Mandelstam capture the paradoxically simultaneous existence and nonexistence of the self:

What I prize in the doughnut is the hole. . . . Real work is Brussels lace, the main thing in it is what holds the pattern up: air, punctures, truancy.



The holes hold the pattern together.

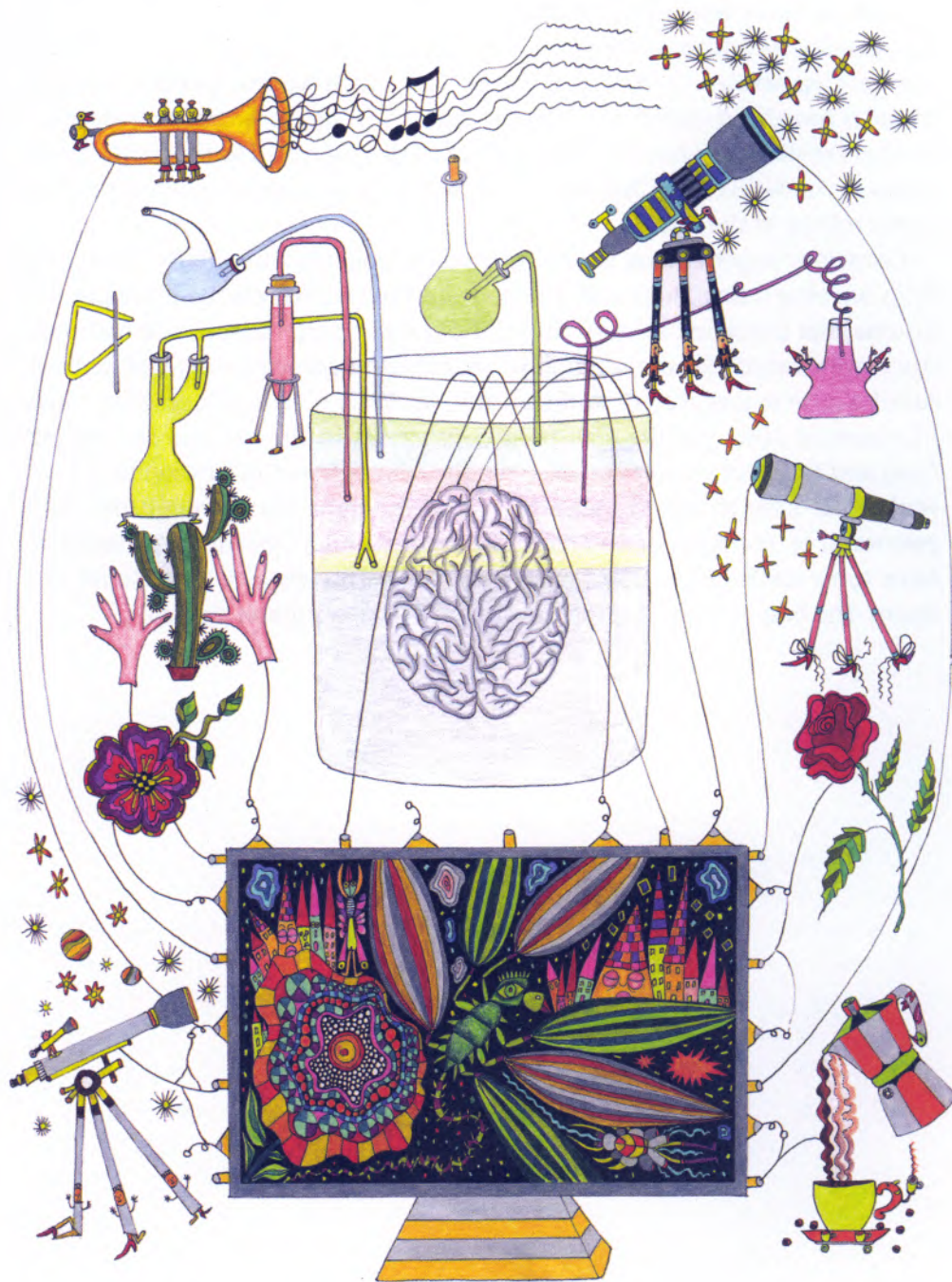
1.11 ARE YOU A BRAIN IN A VAT?

A modern rendition of the butterfly thought experiment described earlier in 1.3 is *the-brain-in-a-vat* thought experiment: how do you know that your conviction that you are now having dinner with your partner, eating spaghetti with mushrooms and leeks and drinking a glass of Primitivo red wine while discussing the science of consciousness, is not a Matrix-like illusion created by a supercomputer to which your brain, floating in a vat with nutrients to keep it alive, is hooked up?

The brain-in-a-vat thought experiment, like all good thought experiments, makes us scrutinize the assumptions on which it is based. The underlying assumption in this case is that the brain, rather than the enbrained human, is conscious. The body is merely a replaceable container keeping the brain alive and providing it with inputs that could be produced by an external source, such as a supercomputer.

Is this assumption valid? What do we know about the brain of a conscious individual? The brain gets chemical and electrical inputs from within the body and from the outside world, processes them through interactions between different brain areas, and sends the output signals back to the body in an ongoing reciprocal dialogue. Newly constructed representations update and evaluate previously remembered representations, motivating the individual to act or to refrain from acting.

Let us assume that we can keep an isolated brain alive for a long time. This may be possible: although at present a pig's brain can be kept alive for only thirty-six hours, we can assume that in the future it will be possible to prolong the life of the brain. But is the brain's life a *conscious* life? In the case of the



A brain in a vat.

pig's brain, there was no global activity and no EEG response, because blocking agents were added to the vat as a safeguard to ensure that the brain would not develop sentience, even if it could. If no such agents had been added, would it have been conscious? If the isolated brain produced patterns of activity that are similar to those it produced while the pig rolled in the mud shortly before it was killed, does this mean that the envatted brain is experiencing the joyful rolling in the mud?

Conscious experiencing seems to involve global brain activity stemming from ongoing interactions with the body and the world. We therefore have to assume that simulated sensory stimuli from the supercomputer will maintain the brain's conscious, ever-changing yet coherent patterns of global activity. And because brains receive and transmit not only electrical signals but also biochemical ones, the life-maintaining processes in the vat and the stimuli from and to the computer may also require the transport of biochemical molecules that have to pass through the vat at the right time through the right pathways, to the right brain areas. The vat-computer system then begins to have some similarity to a responsive body. So who, then, is conscious? The brain? The brain-computer system? The brain-vat-computer system?

1.12 PHILOSOPHICAL ZOMBIES?

The zombie, a fictional being created through the reanimation of a corpse, is the subject of horror fantasies. Zombies lack free will, have abnormal mentality (if any), and harm truly living sentient humans. Philosophers borrowed this popular notion and ran with it.

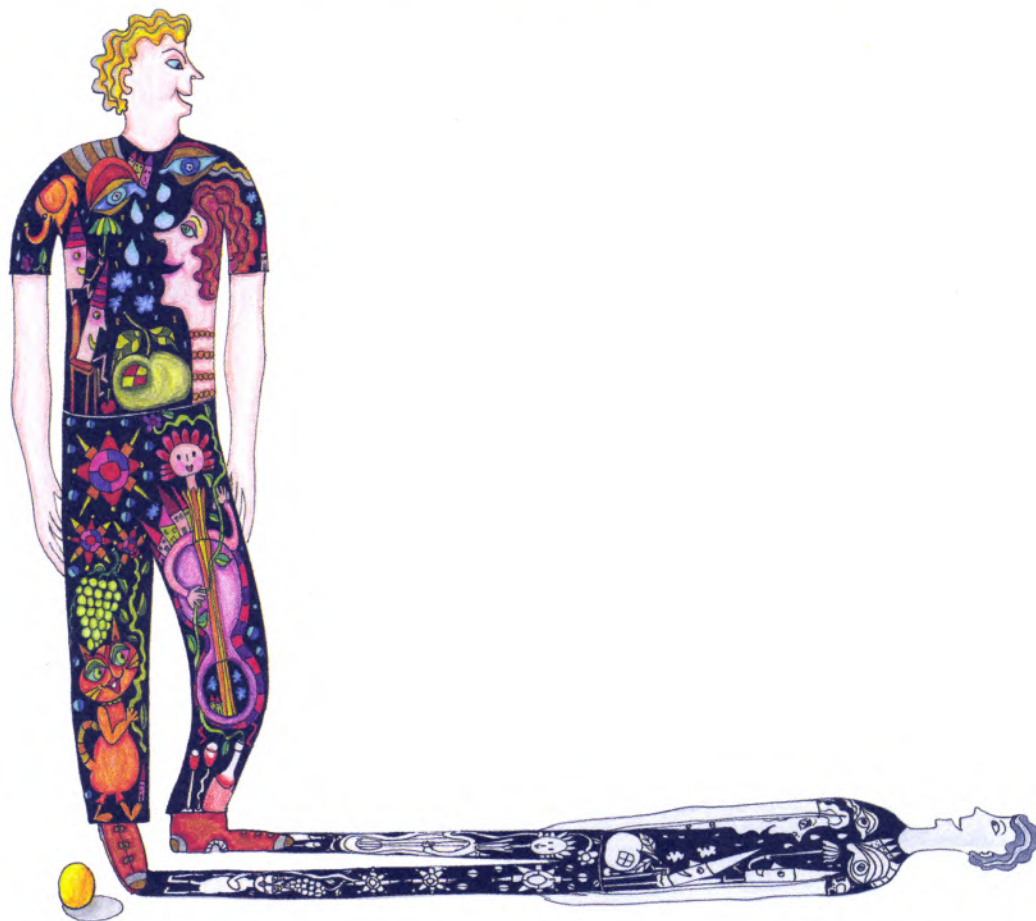
A philosophical zombie (p-zombie) is physically indistinguishable from a living sentient human being, yet it lacks subjective experiencing. If you kicked a p-zombie really hard, it would feel no pain—but it would behave exactly as if it did. Some philosophers think that they can conceive of p-zombies that are atom-for-atom identical to a human but nevertheless lack subjective experiencing. If p-zombies are conceivable, these philosophers argue, this proves that the physical can be separated from the mental, as dualists claim.

What is meant by “conceivable” is, however, a mystery. The p-zombie argument is, in fact, simply a restatement of the dualistic position. It is based on a willingness to accept the proposition that physical identity does not entail mental identity. If one does not accept that this assumption is conceivable, then the zombie argument is an elementary logical contradiction.

An interesting question, which does not sink into a logical contradiction, is whether we should consider as conscious an entity such as a silicon-made, humanly designed robot that behaves in a way that is indistinguishable in our eyes from that of a conscious human being. Does being conscious depend on the way the robot was built, the material it is made of, and the particular processes guiding its behavior?

An in-principle answer to these questions depends on one’s theory of consciousness. If one is a dualist, the complexity of behavior is irrelevant, whereas

for a panpsychist all things are conscious if their conscious elementary particles combine in the right way. It is only for physicalists' and naturalists' theories of consciousness that the question has any significance. Questions about the distribution and evolution of consciousness, which we pursue in the next two vistas, are also of significance only to physicalists-naturalists.

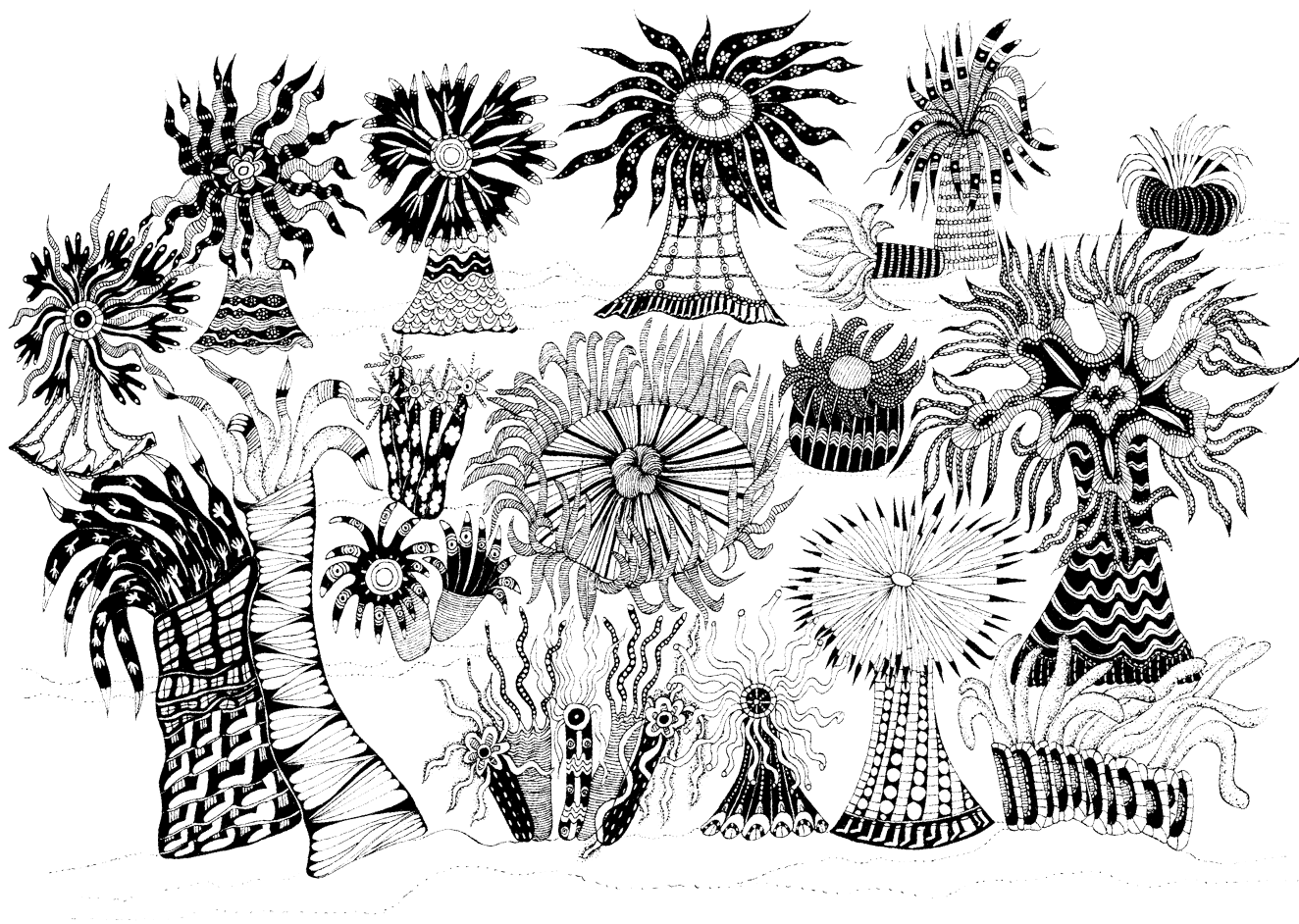


The young philosopher's zombie-shadow

There is a story which tells how some visitors once wished to meet Heracleitus, and when they entered and saw him in the kitchen, warming himself at the stove, they hesitated, but Heracleitus said, "Come in, don't be afraid, there are gods even here." In like manner, we ought not to hesitate nor to be abashed, but boldly to enter upon our researches concerning animals of every sort and kind, knowing that in not one of them is Nature or Beauty lacking.

—Aristotle

VISTA 2 WHO IS CONSCIOUS?



Dancing, fanciful sea anemones: "In not one of them is Nature or Beauty lacking."

2.1 A VIEW FROM BIOLOGY

Which living organisms are conscious? Opinions differ, and the likely consciousness of the whole spectrum of organisms, from bacteria to humans, has been debated. Our own opinion, which we will justify in vista 3, is that the evolution of learning and the evolution of consciousness are intimately linked, even entangled. When we were writing an academic book about the evolution of consciousness, we discussed the origins of consciousness with many people. One encounter left a lasting impression.

The three of us—Eva, Simona, and Anna—were invited to a garden party at an Oxford college. As we indulged in strawberries and champagne, we became cheerful, and our un-British laughter attracted an elderly man with a white shock of hair, who, we later learned, was a famous philosopher. He joined our table and asked who we are and what we do.

Eva said, “We are writing a book about the evolution of consciousness in animals. We are trying to understand how and when consciousness first originated and how it evolved.”

“Consciousness in animals!” the philosopher snorted. “What nonsense! Consciousness requires awareness of percepts and beliefs. Only a system capable of representing primary, and therefore unconscious, mental states can be said to have subjective experiences. Language is probably needed. Nonhuman animals are not aware of what they see or hear or feel.”

Anna, a dog-lover, was upset. She asked: “Have you ever had a dog? Or a cat?”

The philosopher smiled. “Yes, I have a dog, Pythagoras,” he said.

Simona asked: “And is it happy to see you when you come back home?”

The philosopher exhaled with exasperation. "Reacting is not feeling. You will agree, I hope, that a headless chicken running around the yard is not conscious of anything. But even if I cannot deny that dogs do feel in some obscure sense, rabbits do not."

We are not telling this story to ridicule the philosopher's view. We tell it because it brought home to us the need for biological literacy when discussing the mind.



What does the dog think?

2.2 ARE BACTERIA CONSCIOUS?

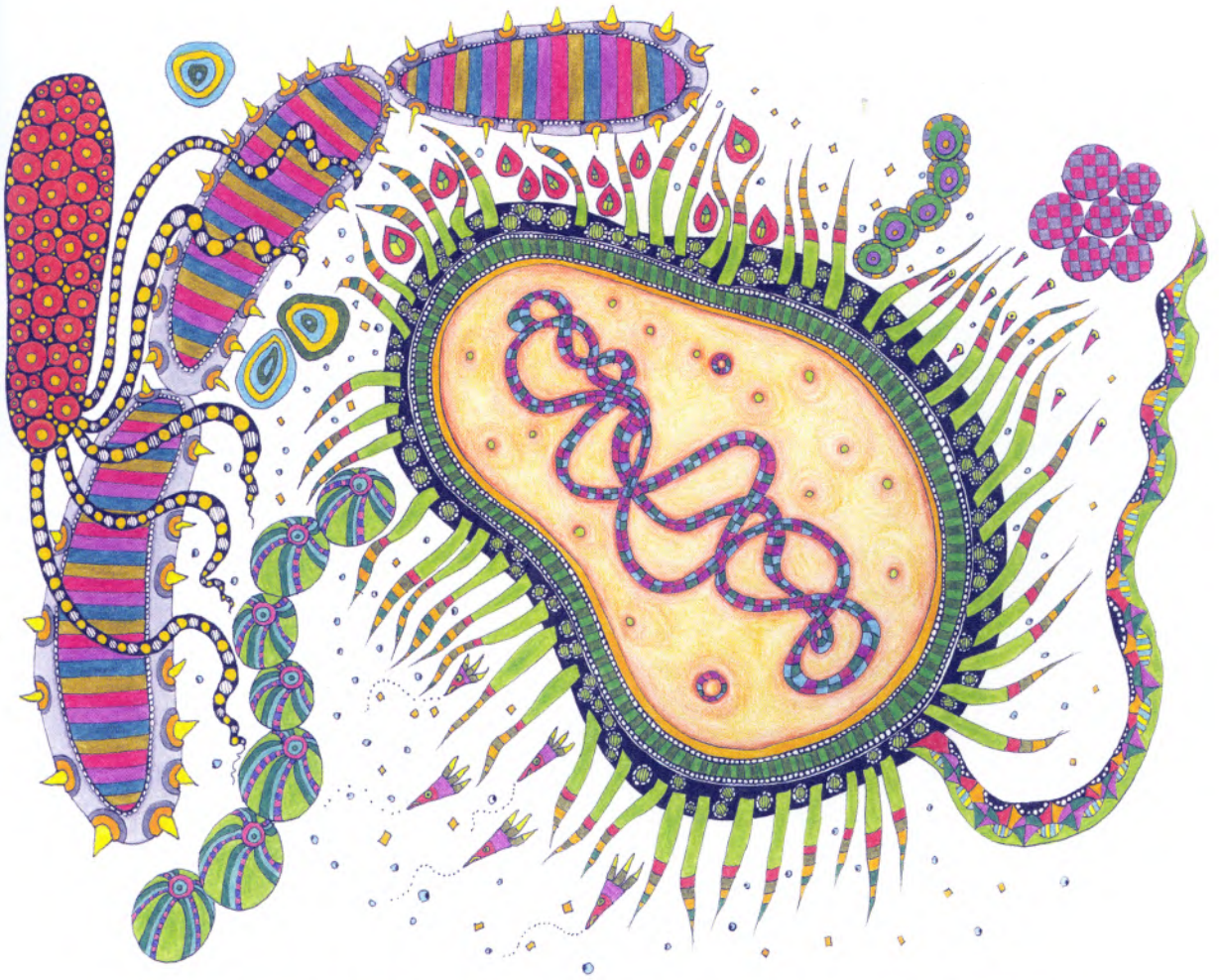
Are bacteria conscious? Some thinkers—biopsychists—believe they are. They believe that sentience is inherent in all living things, including bacteria, but *not* in cups of tea or elementary particles. The nineteenth-century father of experimental psychology, Wilhelm Wundt, was a biopsychist. He argued that since all life is made up of the same irritable protoplasmic material, and since it is impossible to draw a clear line between involuntary and voluntary movement (voluntary movement was his criterion for consciousness):

We cannot resist the conjecture that psychical life and the capacity of giving expression to it are universally represented in contractile substance. From the standpoint of observation, then, we must regard it as a highly probable hypothesis that the beginnings of the mental life date from as far back as the beginnings of life at large.

Arthur Reber, a twenty-first-century psychologist, shares Wundt's view: "Consciousness, subjectivity, phenomenal experience or, if you prefer, sentience, is an inherent feature of living organic form(s)."

The belief that life entails sentience is inspired by the incredible complexity and adaptability of cells. There is an "inwardness" to a living cell—an internal molecular commotion, a dynamic organizational intricacy, ceaseless spontaneous and semidirected streams of interactions that occur within a porous membrane that separates the living entity from its external milieu yet enables it to interact with it. Regulated coupling between sensors and effectors, epigenetic (developmental) memory, and mechanisms that evaluate the adaptiveness of responses enable cells to maintain and regenerate themselves. Only feelings and perceptions, claim biopsychists, can allow such a complex cell to survive and reproduce.

Biopsychism raises problems that are similar to those raised by panpsychism. How do the little micro-sentiences of single cells combine to construct the macro-sentience of an awake human? If life entails sentience, what does it mean for an organism to be unconscious, in a deep coma, for example?



"Might that watery slurry contain an ember of emotion, a prototype of sentience?"
(Dennis Bray, *Wetware*).

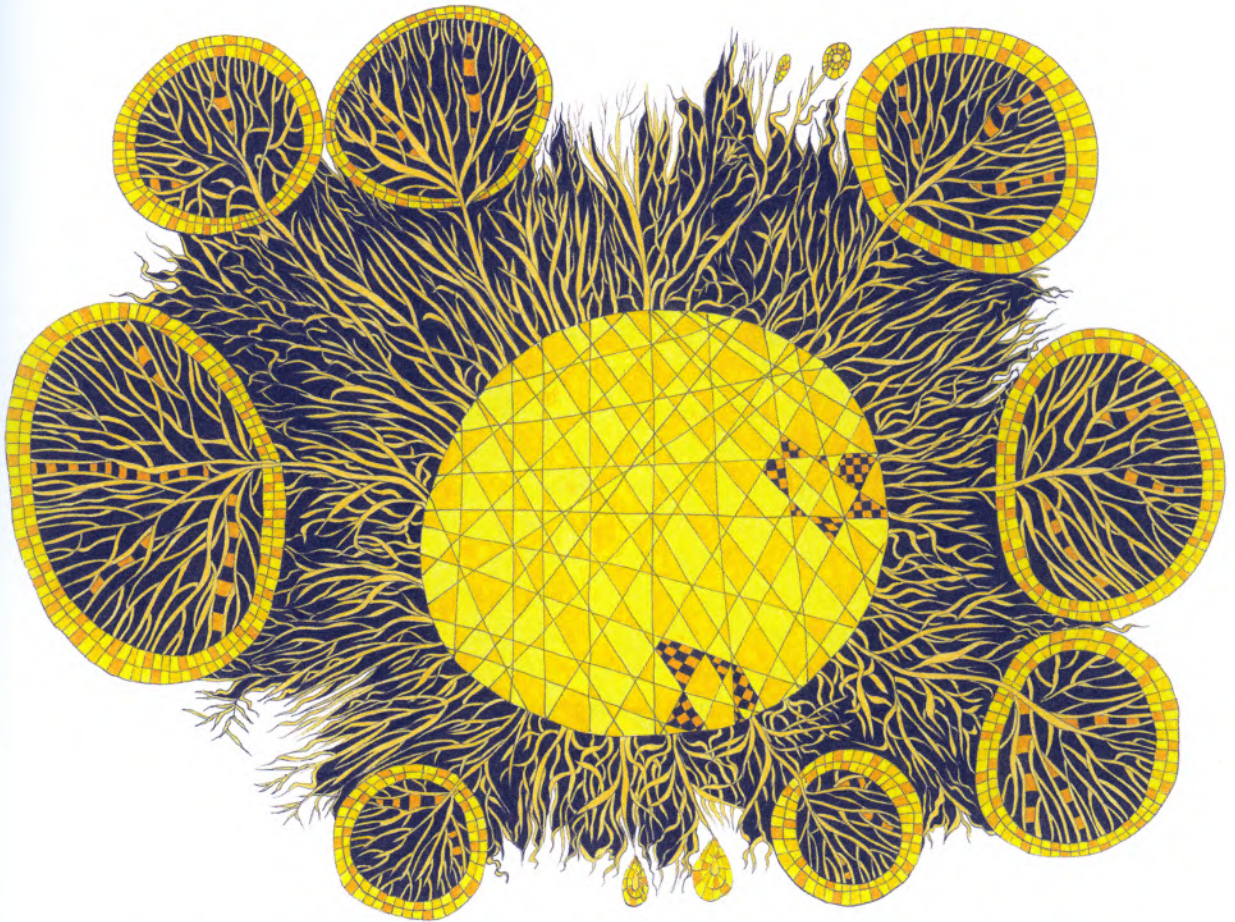
2.3 CONSCIOUS SLIME MOLDS?

Some people think that to be sentient organisms must be intelligent enough to solve challenging problems. So do slime molds, which are kinds of amoebae, fit the bill? *Physarum polycephalum* is a bright-yellow slimy creature with a vast number of nuclei within one cell; it can reach several meters in size and weigh up to 20 kilograms (about 44 pounds). It moves around and alters its shape through changes in the streaming of its cytoplasm. The cognitive feats it can perform have amazed the world and led some of the scientists who study it to say that “it hints at the cellular origins of primitive intelligence.”

What is meant by intelligence is not always clear, but usually the ability to learn and make complex decisions are considered reliable markers of intelligence. Slime molds can do both. They can learn to ignore an irrelevant, repeated stimulus to which they initially respond. This type of learning, known as habituation, does not involve fatigue of sensory or motor systems. One slime mold, for example, learned to ignore low concentrations of quinine and caffeine, but could respond perfectly well to a new irritant.

A slime mold can also make strategic decisions. When allowed to grow in a maze with a favored food (oatmeal) put at two spots in the maze, it retracts from all the arms of the maze except those that form the shortest route between the two spots of food. And this is just the beginning: when oat-flakes were arranged to represent Tokyo and thirty-six surrounding towns, the creature created a network similar to the existing train system in Japan “with comparable efficiency, fault tolerance, and cost.” Such a design feat, drawing transport networks that minimize the distance between many different locations, is difficult even for trained human engineers.

How these amoebae accomplish these feats is at present unknown. But does the answer to the “how” questions matter? Is the ability to learn through habituation and to make complex “decisions” a sign of sentience?



Marvels of graceful elegance and beauty, as the biologist Thomas MacBride described them: the slime mold's inner, as yet unknown, order enables it to make sophisticated decisions, form fruiting bodies, and thrive.

2.4 ARE PLANTS CONSCIOUS?

Not many people would argue in favor of slime-mold sentience, but a surprising number passionately maintain that plants are conscious. Plants constitute 99 percent of the biomass on Earth, and we depend on them for our air, food, shelter, and medication. We admire their beauty, get intoxicated by their scent, and are awed by the majestic size and age of some of them. The adaptations of plants are legendary, and their sensory capacities surpass those of many animals. They can sense different light wavelengths, vibrations, airborne and soluble molecules, gravity and touch. When injured, some plants produce toxic substances in the site of damage, send information about the damage to the rest of the plant, communicate it to their neighbors, and transmit its memory-traces to their offspring. In *Overstory*, a novel about trees, Richard Powers describes the communication among trees and the forests' knowledge:

A forest knows things. They wire themselves up underground. There are brains down there, ones our own brains aren't shaped to see. Root plasticity, solving problems and making decisions. Fungal synapses. What else do you want to call it? Link enough trees together, and a forest grows *aware*.

So are trees, or forests, sentient? Are plants, as a botanist friend has told us, "very very very slow animals"? They use electrical signaling and have reflex-like responses based on them. The Venus flytrap snaps and closes its leaves on insects and spiders that touch the tiny hairs on its surface. So do plants have something akin to neurobiology? This is a topic of debate among plant physiologists and ecologists. If it is true that the physiology of plants has parallels with the neurophysiology that constitutes the sentience of animals, and if plants' cognitive capacities equal those of sentient animals, then the claim that plants are sentient will gain support. But is this the case?

Animals and plants took different evolutionary routes. The transport systems of sessile plants and the nervous system of moving animals have different functions, structures, and capacities. The learning abilities of plants are very limited. Maybe we lack a concept to describe the “inwardness” or “overall sensation” of plants, but is sentience the right concept?



“Plants extend their cognition beyond their bodies, forcing us to reconsider where a plant ends and its environment begins” (Parise, Gagliano, and Souza, “Extended Cognition in Plants”).

2.5 ARE SPONGES CONSCIOUS?

Are animals the only sentient beings? Humans are undeniably conscious and undeniably animals, so is animality sufficient for consciousness? Are the animals at the base of the animal tree—the placozoans and the sponges, which lack a nervous system and muscles—sentient?

Placozoans are tiny amoeba-like creatures made up of several thousand cells. We know a lot about their DNA, little about their physiology, and nothing about their learning capacities. It is not clear if they once had a nervous system and lost it, or if “nervelessness” is their ancestral state. These animals are far less complex than the most modest weed in your garden.

The second and much more colorful and species-rich nerveless group of animals are the sponges. Sponges, too, hardly move as adults, although their tiny larvae swim with the help of cilia, as do some unicellular organisms. As with placozoans, it is not clear whether sponges originally had nerves and muscles and later lost them, or if the absence of nerves is their ancestral state. Sponges have reflex-like reactions and electrical signaling, but this is not a hallmark of animality since plants also have both. And although some of the genes that are expressed in neurons are also expressed in some sponge cells, we know that the genetic tool kit of organisms is often reused in new ways during evolution. We cannot yet determine if nerveless sponges were the ancestors of all neural animals.

We know nothing about sponge memory, learning, or intelligence. No argument except a biopsychic axiom has ever been advanced to suggest that they are sentient. Their vitality, however, is undeniable.



Are immobile, colorful creatures like these the ancestors of all animals?

2.6 WHAT'S SO SPECIAL ABOUT A NERVOUS SYSTEM?

Is a nervous system required for sentience? After all, nonneural organisms process and transmit information—in the biochemical networks within a single cell, in the streaming cytoplasm of slime molds, and through the transport system of plants. What, then, is so special about a nervous system?

The nervous system has five special properties:

- **Connectivity:** a nerve cell has a special morphology, with projections leading to and from the cell body. Through these projections, each nerve cell can be connected to many others. Neurons can form many modifiable neural circuits and networks that can be embedded within larger networks.
- **A common language:** the discrete neural electrical signal, the action potential, is the common communication currency within the nervous system. All types of sensory stimuli—photons, chemicals, heat, sound waves, pressure—are translated into this common communication currency.
- **Contact specificity:** contact-wiring between cells and cell-projections enables neurons to communicate with great precision. Contact wiring enables the formation of three dimensional spatial maps of the relations between objects and events.
- **High speed:** the speed of electrical transmission makes it possible to react rapidly to changing conditions.
- **A new type of memory:** memory is stored not only in the epigenetic, biochemical circuits within the cell and the bioelectrical fields within the body, but also at connections between cells. The synapse, the “point” of contact between neurons, can be strengthened or weakened by activity, and traces

of past activity can persist as altered synaptic connections. Neural animals thus have a multitiered neuronal memory: in addition to memory stored *within* neurons, they have between-neuron synaptic memory.

Neural networks can map and update the changing patterns of stimuli coming from the world and the body. Importantly, they coordinate adaptive motor activity—enable animals to move around, harmonize the movements of their internal organs, and form models of intended actions. It is noteworthy that almost all animals that have nerves also have muscles. Did neurons and muscles co-evolve? Is an ability to neurally map world, body, and action necessary for sentience? Are all neural animals sentient?



Complex, generative, flexible, natural information processors.

2.7 ENCHANTING MEDUSAE, GRACEFUL SEA ANEMONES

Which animals were the first to have a nervous system? What difference did this innovation make to their cognitive capacities? Did it grant them sentience?

The simplest neural animals do not have brains. They have nerve nets spread throughout their bodies. Among these creatures are the exquisite cnidarians, some of the most beautiful animals on the planet. They include the sea anemones, corals, and jellyfish. The extraordinary beauty of medusae—radial, soft-bodied animals such as jellyfish with specialized poison cells in snake-like tentacles springing from their rounded body—is the stuff of myths. In Greek mythology, Medusa is a Gorgon, a monster with snakes growing out of her head, whose gaze turns men to stone. In Ovid's version of the myth, Medusa was raped by Poseidon, god of the sea, cursed for *his* transgression and turned into a monster. She was beheaded by the hero Perseus, who gave her head to Athena, goddess of wisdom and civilization, to place on her shield. Current metaphors depict Medusa as a symbol of female power, beauty, and rage, a protector of women.

The varied sensors and muscle sheets of medusae enable them to prey on other animals, escape predators, form swarms, and rapidly change directions and goals, something that no nerveless multicellular organism can do. Maybe the most impressive among medusae is the deadly box jellyfish. Not only do these jellyfish have multiple sense receptors like other medusae, they also have a complex *sense organ*: four clusters of six eyes, carried on four dangling lobes, gaze up at the world. One species of box medusa can navigate around the mangrove swamps in which it dwells, using terrestrial landmarks seen through the water's surface.



The mythological Medusa may turn whoever gazes at her to stone, while many living medusae may cause shock and death to the victims they sting.

Can medusae learn? We know that they can learn to ignore repetitive, weak, reflex-eliciting stimuli (habituation) and can become more responsive when a strong, reflex-eliciting stimulus is sensed (sensitization). But can they also associate stimuli? Can they learn that a neutral stimulus—one that is normally unimportant to them and to which they are indifferent—predicts an associated salient one? Alas, no experiments testing for such associative learning in medusae have been reported. A single case of associative learning was observed in a sea anemone. Though sessile, these graceful animals can move their tentacles rapidly, dramatically change their shape and learn just as medusae do. Is such limited learning indicative of consciousness?

2.8 BRAINY WORMS

Does sentience require a brain? The simplest animals with a brain are the creepy-crawly worms, an ill-defined assembly of limbless, elongated animals with bilateral symmetry, which includes annelids, flat worms, nematodes, and others. Some are very tiny, but all have a condensed cluster of neural networks at one end of their body, where most of their sensors are concentrated: they have a center of communication, a brain.

Worms illustrate the marvels of behavioral adaptation, memory, and the many feats of learning that a simple brain can support. “What can’t a worm learn?” wondered Catharine Rankin who has been studying *Caenorhabditis elegans*, a one-millimeter-long, transparent, free-living nematode worm. Not much, was her conclusion. Within the limitation of their senses (they have no eyes, but can smell, taste, and respond to oxygen, temperature, and pressure), these worms can learn to approach or avoid tastes, odors, or temperatures that predict the presence or absence of food; can distinguish between simultaneously presented rewarding and punishing cues; can integrate cues from different senses; can explore mazes; can respond to the context in which they received rewards and punishment; and can make decisions.

Another type of worm, the flat worm *Dugesia japonica*, one of the simplest brainy worms, can learn about its world and actions using its primitive eyes, and is able to discern chemical gradients, vibrations, and electric and magnetic fields. Champions at regeneration, when a flat worm is cut into pieces, the pieces can regrow both head and tail. If a worm is decapitated, will the newly regenerated head remember what the old brain had learned? Yes it will, affirm the results of old and new experiments. Memory is distributed: it inheres in



"The ability to build a specific anatomical structure from different starting conditions, and stop when precisely the right pattern is finished, is one of the big, unsolved problems of developmental biology and regenerative medicine today" (Mike Levin).

brains, in peripheral neural circuits, in single cells, in molecules, and in electric fields.

Only animals with brains can learn to associate multiple unimportant, neutral cues with salient ones. By doing so, what is relevant to the animal—its external world, its body, and the range of its actions—is enormously extended. This is true for worms and for other animals with brains: sea slugs, land snails, fish, and flies.

Is every brainy creature conscious? Is the intelligence afforded by a worm's brain indicative of sentience? Does the organization of the brain matter? Is there some consensus about animal consciousness?

2.9 A FUZZY CONSENSUS

At the end of the twentieth century, scientists began to carve out a new scientific discipline—consciousness studies—by studying patterns of activity in the brains of people in alert, sleeping, dreaming, comatose, drugged, and disturbed states. They compared awareness under different conditions, trying to capture the neural signatures of conscious states. For example, a human who is shown a blue flower under normal conditions will report: “I saw a blue flower.” Intriguingly, when shown the picture when attention is diverted or the length of exposure is very short, the person reports “I saw nothing at all,” yet when asked to guess if she was shown a blue or a yellow one will usually guess that she was shown the blue flower. Researchers can examine the differences in brain activity during the two states and pinpoint the neural correlates of the conscious state of seeing.

Human consciousness is not in doubt, and human behavioral, cognitive, and neural properties are therefore used as a benchmark: they are used as criteria for guiding the ascription of conscious states to other animals. Animals that display brain activities, cognitive capacities, and behaviors that are analogous to those that accompany mental states in humans are deemed conscious.

A fuzzy consensus about the distribution of consciousness among animals has begun to emerge. The 2012 *Cambridge Declaration on Consciousness* states:

Convergent evidence indicates that non-human animals have the neuroanatomical, neurochemical, and neurophysiological substrates of conscious states along with the capacity to exhibit intentional behaviors. Consequently, the weight of evidence indicates that humans are not unique in possessing the neurological substrates that generate consciousness. Nonhuman animals, including all mammals and birds,

and many other creatures, including octopuses, also possess these neurological substrates.

Which are those “other creatures”? Are *all* vertebrates—fish, amphibians, and reptiles, as well as birds and mammals—conscious? What about insects? Why octopuses? What are the “neurological substrates” of conscious animals?



Humans are not unique in possessing the neurological substrates that generate consciousness.

2.10 ARE FISH CONSCIOUS?

We humans are vertebrates. We share deep homologies—similarities due to common ancestry—with all members of our phylum: with fish, amphibians, reptiles, birds, and, of course, other mammals. The more recent is our common ancestor, the more homologies we share.

How far back can we go? Fish, the most species-rich group of vertebrates, first appeared over 500 million years ago. Their brains have the integrated midbrain and basal ganglia structures that support the awareness of the world and the self in humans; fish can learn to navigate in complex environments; they use tools; they can discriminate between multifaceted visual and auditory patterns that differ in subtle ways, when, for example, they are seducing, competing for, or choosing a mate. A striking example is the male white-spotted puffer-fish, who builds an extravagant nest to attract a female: using his body and fins, this two-centimeter-long fish constructs on the seabed a two-meter wide sand circle, with a complex symmetrical internal pattern, decorates it with shell fragments, and brings fine sand particles into the innermost central area of the circle, into which the female will lay eggs if she deems the mandala-like structure good enough.

Charles Darwin thought that such an ability to discriminate between different displays and patterns is a good indication of mentality (a term used in his time as a synonym for consciousness):

He who admits the principle of sexual selection [selection for traits that increase success in obtaining and choosing a mate] will be led to the remarkable conclusion that the cerebral system not only regulates most of the existing functions of the body, but has indirectly influenced the progressive development of various bodily



The puffer-fish as sand-sculptor.

structures and of certain *mental qualities*. Courage, pugnacity, perseverance, strength and size of body, weapons of all kinds, musical organs, both vocal and instrumental, bright colours, stipes and marks and ornamental appendages, have all been indirectly gained by the one sex or the other, through the influence of love and jealousy, through the appreciation of the beautiful in sound, colour or form and through the exertion of choice; and these powers of the mind manifestly depend on the development of the cerebral system. (emphasis added)

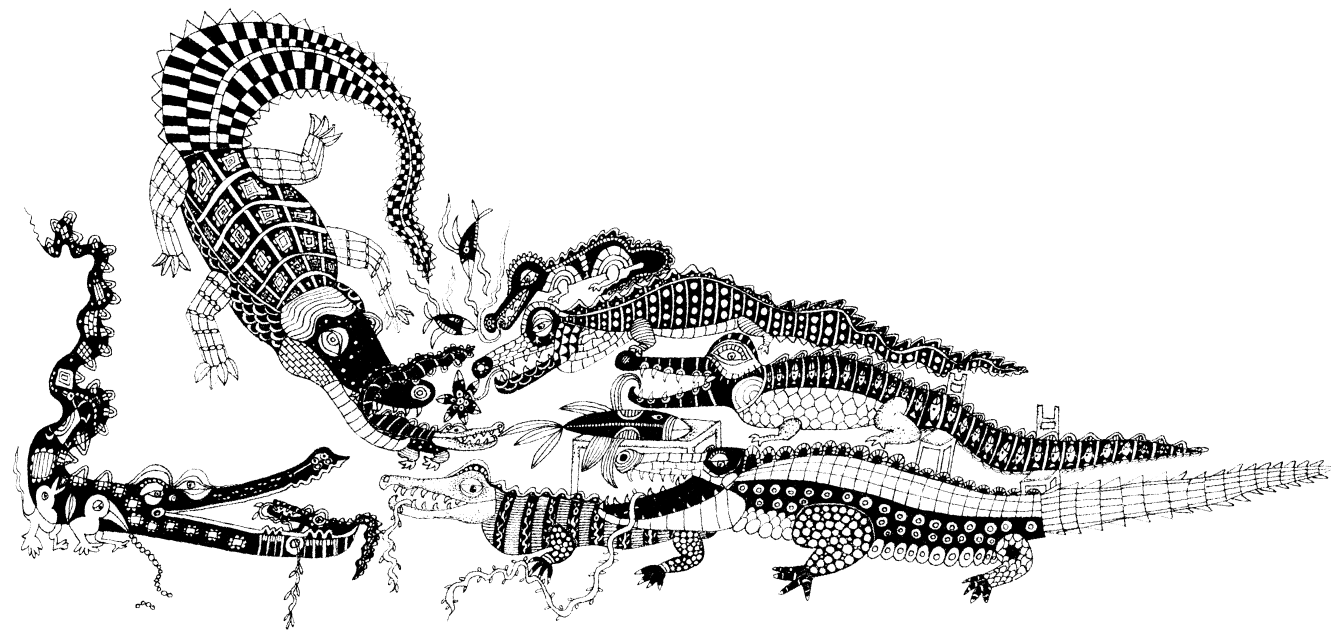
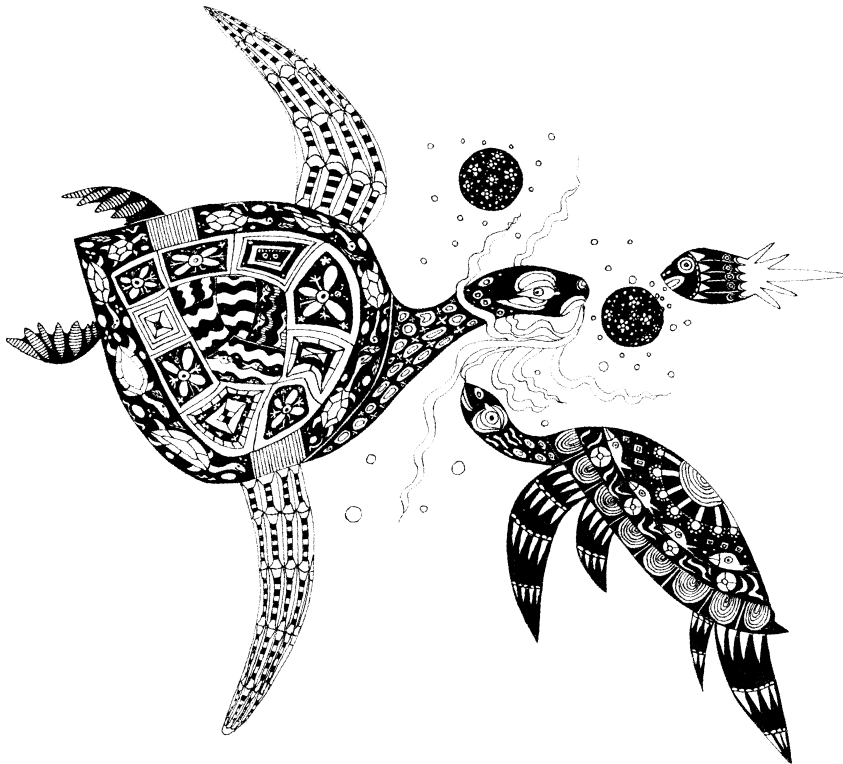
Was Darwin right? Are such discerning and decision-making capacities sufficient criteria for consciousness in fish? Does the female puffer fish have a visual experience of the mandala the male has created?

2.11 ARE REPTILES CONSCIOUS?

Does consciousness require additional capacities that fish do not have? What about the first vertebrate land-dwellers, the amphibians? Some scientists say that only those animals that are able to play, sleep and dream, avoid harmful stimuli, and make detours in order to reach a goal that is temporarily hidden from direct perception are conscious. These capacities, they claim, first appeared in reptiles.

Reptiles certainly have these capacities. Turtles play: a Nile soft-shelled turtle that was bored and self-harming in his pond started playing when objects were put into it. He nosed a basketball around his pond, bounced a hula-hoop through the water, and altered the position of the water hose so that it sprayed him in the face. We also know that lizards learn to avoid toxic prey and to discriminate between containers with different-colored lids, only one of which is associated with a food reward. Yarrow spiny lizards can navigate and find their home when it is out of sight, using both a cognitive map and a “third eye” (a structure made up of photoreceptors covered by skin) on top of their heads that acts as a celestial compass, helping them to orient toward home; and Eastern water skinks can change their evaluation of a refuge when the location of predator attacks changes. And reptiles may dream: the bearded dragon, a large Australian lizard, undergoes REM (rapid eye movement) sleep. REM sleep is correlated with dreaming in birds and mammals including humans.

Is this assumption that fish and amphibians lack the learning and emotional capacities of reptiles correct? We now know that fish and amphibians play, that they learn to avoid harmful stimuli and to navigate, and some of them, like devoted crocodile mothers, exhibit ardent parental care. We also



Mothering and protective crocodiles, and playful turtles.

know that toadfish grunt when electrically shocked, and, over time, come to grunt at the mere sight of an electrode. Are they fearfully anticipating pain?

We know nothing about dreaming in fish and frogs. But we know a lot about fish cognition (much less is known about amphibians). Remarkably, many fish have larger brains than most reptiles, and have outstanding intelligence. Giant manta rays check out their reflections in a mirror in a way that suggests that they may be self-aware: they do not try to socially interact with the mirror, but move their fins, blow bubbles, and circle in front of the mirror. Such behaviors have never been recorded in reptiles. If reptiles are deemed conscious, on what basis can we deny consciousness to fish and amphibians?

2.12 BIRDS AND MAMMALS

The brain organization of nonhuman mammals, including the organization of the neocortex (the outer layers of the brain that are specific to mammals and underlie their higher mental functions), is very similar to that of humans. Their intelligence and emotional sensibilities leave very few biologists in doubt about the ability of mammals to have subjective experiences. Dogs, chimpanzees, dolphins, and elephants spring to mind, but these clever animals are not exceptional. Many other mammals, such as rats, goats, and pigs, are equally impressive.

The story of two English ginger pig siblings that escaped slaughter by squeezing through a fence and swimming across the river Avon captured the interest and compassion of readers throughout the world. The story tells us more about how the media works than about pigs, but pigs are amazingly clever and sensitive. They become infected by the fear or pleasure of pig companions; recognize individuals in their group; cooperate with group members and learn from them; love to play; locate objects in complicated mazes; respond to verbal commands; learn complex combinations of symbols for actions; and can manipulate a joystick to move an on-screen cursor and use a mirror to find hidden food.

Bird consciousness used to be doubted because birds do not have a neocortex. But studies are revealing deep neuroanatomical and developmental homologies between birds and mammals, as well as neurophysiological counterparts of mammalian conscious states, such as similar waking electric brain patterns. In intelligence, birds are the equals of mammals. All birds probably dream and may imitate songs, calls, and human-made sounds. Crows



And the bird shall eat with the hedgehog
 And the elephant and cat shall join them,
 And the leopard with the tortoise shall dwell
 And no one will lead them
 —Jane Monet

manipulate tools; magpies recognize themselves in the mirror; California scrub jays remember where and when they cached a particular item of food. The gray parrot Alex could count, categorize items, and ask questions. According to Alex's trainer, Irene Pepperberg, the intelligence of gray parrots is comparable to that of chimpanzees and dolphins, and in some aspects to that of a five-year-old child.

It is not surprising that mammals and birds are at the center of the consensus about animal consciousness, but the consensus is blurry at the edges. There are a few scholars who argue that humans alone can be deemed conscious, while a growing minority maintains that not just fish, but also insects and cephalopods (octopuses, squids, and cuttlefish), are conscious. Since the brains of insects and cephalopods are very different from those of vertebrates, is similarity in behavior the basis for their arguments? Are there deep functional similarities between the brains of these animals?

2.13 SOME ARTHROPODS?

Do not swat the fly:
See how it wrings its hands,
How it wrings its feet!
—Issa, 1763–1827

Does the poet's compassion go too far? What does biology tell us about the subjective experiences of flies and other arthropods (insects, crustaceans, spiders, and myriapods)?

Although the anatomy of arthropods' brains is totally different from that of vertebrates, the functional architecture of their brains is surprisingly similar to that underlying basic consciousness in vertebrates. Both brain types have specialized circuits for processing information about the position of their bodies and for integrating past and present multimodal sensory information about the external environment. The animals decide what to do only after this information is integrated with the knowledge of the likely reward or punishment associated with the stimuli.

The humble crayfish seems to feel. Light-hating crayfish that have been exposed to a stressful electric field in the dark arm of a Y-maze (which they normally would prefer) will not enter it even when the electric field has been switched off. As with mammals, stress increases the level of the neurotransmitter serotonin, and crayfish calm down after being injected with a drug that alleviates anxiety in humans.

The fruit fly *Drosophila*, the experimental workhorse of geneticists, manifests remarkable learning capacities: it integrates information from different



Conscious arthropods?

modalities like smell and vision and learns the relation between a neutral (indifferent) odor or visual stimulus and a repulsive or attractive stimulus.

Social insects show the most amazing feats of intelligence. Bees discriminate between patterns, use abstract concepts, and learn from each other; some (honeybees) communicate about the location of food and nest sites, work harder after receiving rewards, and tend to give up after punishments. In 1901, Maurice Maeterlinck admiringly wrote:

There can be no doubting that they understand each other; and indeed it were surely impossible for a republic so considerable, wherein the labours are so varied and so marvellously combined, to subsist amid the silence and spiritual isolation of so many thousand creatures.

2.14 WHAT ABOUT CEPHALOPODS?

The most alien-like mind on the planet, from our vertebrate perspective, belongs to cephalopods—the squids, cuttlefish, and octopuses. They lack a skeleton and have three hearts that pump their blue blood around; they have photosensitive cells in their skin and can control their colors and patterns for camouflage and, in some of them, for expressing their emotions. Cuttlefish produce waves of colors during REM sleep and, unlike other animals that see colors by using different light receptors, they are believed to see colors by amplifying chromatic aberrations through their W-shaped pupils. They have more neurons than a mouse, but two-thirds of these nerve cells are in their flexible, multipurpose tentacles. The tentacles have an autonomy that has been compared to that of improvising Jazz players, although the brain has a significant measure of control. The poet Ogden Nash may have gotten it right:

Tell me, O Octopus, I begs
Is those things arms, or is they legs?
I marvel at thee, Octopus;
If I were thou, I'd call me Us.

—Ogden Nash, 1960

In spite of all their idiosyncrasies, the electrical wave patterns of alert cephalopods are similar to those of alert vertebrates, as are their cognitive achievements. They can classify and categorize objects and discriminate between them using size, shape, taste, smell, and intensity as criteria; match their body color to the color patterns of their environment, and send color-signals to each other; find their way in complex mazes; crawl from their aquarium to hunt for fish in nearby aquaria and return back to theirs; rapidly learn from other octopuses by observing them; and use abandoned coconut shells for

protection. Cuttlefish can remember what they did, and when and where they did it. They seem to be able to express resentment: when fed second-rate food, one octopus waited until the keeper was watching and then, and only then, dumped it down the drain. Yet most cephalopods live only one or two years. Do they need their great intelligence to find elusive vertebrate prey and to avoid the many predators that are after their tasty flesh? Can we even imagine what it is like to be a short-lived, clever octopus?

An octopus trap—
Inside, an ephemeral dream
Under the summer moon.

—Matsuo Basho, 1688



Octopus, cuttlefish, and squid: Are tentacles a seat of consciousness?

2.15 ARE ONLY HUMANS CONSCIOUS?

No one doubts that humans are conscious, although some humans (children, women, and non-white men) were once thought by many white Western males to be less richly conscious than others. Animals, on the other hand, were deemed nonconscious by some philosophers. The seventeenth-century priest and rationalist philosopher, Malebranche, who was influenced by Descartes, asserted:

They [animals] eat without pleasure, cry without pain, grow without knowing it; they desire nothing, fear nothing, know nothing.

The reasons for this “beast-machine” view, which justified vivisection and other horrors, were many and varied, but all were based on the assumption that, unlike humans, animals lack sentience, or a “sensitive soul,” and hence cannot suffer. The modern rationale behind the claim that humans alone are sentient is not entirely different, with the caveat that something like a *rational soul*, or more precisely, *metacognition* is necessary for an animal to feel. This means that in order to think, one has to think about thinking, and in order to feel, one has to think (or feel) about feeling. If such metacognition or higher-order thoughts (HOTs) are absent, one will cry without pain, perceive visual images without seeing, and run away from a monster without fear. It is claimed that since only humans and possibly higher apes have the capacity for HOT, they alone are conscious.

Some HOT theorists relate consciousness to the linguistic capacity of humans. They argue that since human language is symbolic and representational—the perceived cat is represented by the word *cat*—language is crucial for the construction of the human concept of the self as the owner of perceptions

or feelings. Hence, only linguistically endowed humans are aware of what they perceive and feel. Babies and animals who lack language are deemed not conscious.

In view of what has been learned about animal behavior, neurology, learning, and the self–other distinction, this view is now held by an increasingly shy minority of scholars. For most, the “rational soul” is the hallmark of human consciousness, and the continuity and discontinuity of this variety of consciousness with that of other animals continues to fascinate and tax their imagination.



The bonobo has fewer words than a human, but as the artist sees it, his glance communicates more emotion and empathy. He sees you, while the man's attention turns inward.

2.16 WHAT ARE THE SOCIAL DIMENSIONS OF CONSCIOUSNESS?

Social animals, such as bees, ants, dolphins, parrots, and humans, cannot survive on their own. They have to interact with family and group members in ways that maintain the social structure on which they depend. Humans and other animals do it by sharing information through various types of communication; by learning from each other; by having a division of labor and collective decision-making; and by using rhythm and emotional contagion to synchronize their activities and emotions. Social attention, social emotions, and social cognition are important aspects of the consciousness of social animals. They seem to have a “social map” that guides the development of their emotions, learning, and actions. Social structures and relations are internalized in each and every individual. Genetic biases, social learning, and ecological legacies combine in fashioning the internalization of “the social.”

The internalization of social structures and relations can lead to the reconstruction of group traditions. Tool-use traditions in chimpanzees or song dialects in birds and whales persist for generations, but no other species is as immersed and dependent on culture as humans are. Humans internalize “the social” through rituals, myths, stories, art, and moral and other norms. Carl Jung took the idea of cultural internalization a few controversial steps forward (or backward) into phylogeny (evolutionary history). He maintained that in each human individual there is a distinct archaic stratum of symbolic, collective consciousness, a psychic museum of our phylogenetic history. In 1936, he wrote:

In addition to our immediate consciousness, which is of a thoroughly personal nature and which we believe to be the only empirical psyche . . . there exists a



A mandala: A symbol of the self in many spiritual traditions.

second psychic system of a collective, universal, and impersonal nature which is identical in all individuals. This collective unconscious does not develop individually but is inherited. It consists of pre-existent forms, the archetypes, which can only become conscious secondarily and which give definite form to certain psychic contents.

The collective unconscious is expressed in myths and common symbols such as the mandala figure that appears in many cultures. Can our understanding of human cultural-symbolic inheritance account for this and other cultural universals?

VISTA 3 HOW DID CONSCIOUSNESS EVOLVE?

Mind can be understood only by showing how mind is evolved.
—Herbert Spencer

3.1 WHY TAKE AN EVOLUTIONARY APPROACH TO CONSCIOUSNESS?

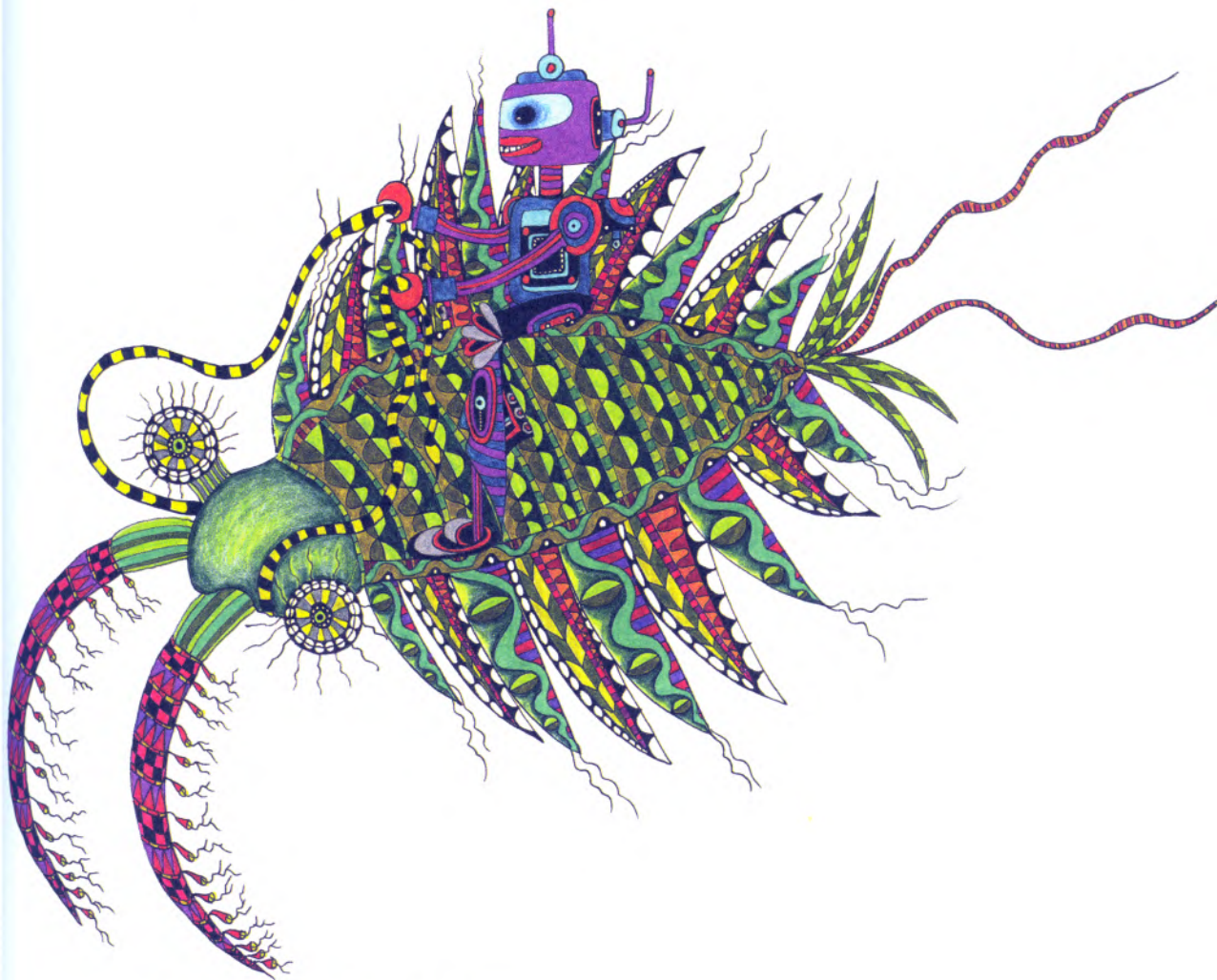
As the previous vista has shown, naturalists offer various answers to the question “Who is conscious?” Their arguments are based on common sense: the more similar an animal is to a conscious human, the more likely it is to be conscious. However, there is no agreement on how far the argument can be stretched. Opinions span the whole spectrum, from “All life is conscious to “Only humans are conscious.”

Common sense is a good starting point, but it is not sufficient. A *theory* of consciousness is needed if we are to recognize which animals are conscious. We need a theory because without it we can have no idea what the building blocks of consciousness are and how they combine to render an organism conscious. But what kind of theory do we need? A theory built on evolutionary reasoning can be especially valuable. Evolutionary theory is the most powerful theory in biology, a conceptual bottleneck through which every theory of life and mind must pass. If a biological (or psychological, or sociological) theory is incompatible with evolutionary theory, it is likely that something is seriously wrong with it. Evolutionary theory is therefore a good reality check.

An evolutionary approach that starts from the transition from nonsentient to sentient organisms offers more than a reality check: the earliest sentient organisms do not carry the misleading baggage of later-evolved structures and processes. For example, without knowledge of evolution we might conclude that a recently evolved region of the brain, which now controls structures that supported consciousness in remote ancestors, is necessary for *all* forms of consciousness. Recognizing the earliest forms of consciousness enables us to identify basic processes and principles that can be applied to all conscious organisms and offers a tentative answer to the “Who is conscious” question.

A theory of consciousness can also tell us when and under what conditions it first appeared, what its functions might have been, whether it evolved once or more than once, whether it was lost in some lineages, how it evolved into new, richer forms, and what future forms it may take.

Like all scientific theories, evolutionary theory is constantly being updated, challenged, and expanded. Since it is the foundation on which we build our theory of consciousness, we will first outline the current version of evolutionary theory.



Consciousness: A past and a future.

3.2 EVOLUTIONARY THEORY

Evolutionary theory is a deceptively simple theory, which is why many people who have only a cursory acquaintance with it are nevertheless convinced that they fully understand it. Its basic assumptions are indeed simple. The first assumption, which was systematically explored first by Jean-Baptiste Lamarck and then by Charles Darwin, is that there was a single ancestor, or very few ancestors, of all living organisms. This is the principle of *Descent with modification*: all organisms are descended, with modifications, from ancestors that lived long ago. The second principle, which is central to Darwin's theory, is the principle of *Natural selection*: organisms with hereditary variations that render them better adapted to their local environment than others in their population leave behind more offspring. Darwin showed that this simple process, when applied recursively, can account for the evolution of complex organs like the eye, and, with the addition of some plausible auxiliary hypotheses, can explain the diversity of living species and their geographic distribution.

In the last paragraph of *The Origin of Species by Means of Natural Selection*, Darwin summarized his ideas:

It is interesting to contemplate an entangled bank, clothed with many plants of many kinds, with birds singing on the bushes, with various insects flitting about, and with worms crawling through the damp earth, and to reflect that these elaborately constructed forms, so different from each other, and dependent on each other in so complex a manner, have all been produced by laws acting around us. These laws, taken in the largest sense, being Growth with reproduction; Inheritance which is almost implied by reproduction; Variability from the indirect and direct action of the external conditions of life, and from use and disuse; a Ratio of Increase so high as to lead to a Struggle for Life, and as a consequence to Natural Selection, entailing

Divergence of Character and the Extinction of less-improved forms. Thus, from the war of nature, from famine and death, the most exalted object which we are capable of conceiving, namely, the production of the higher animals, directly follows. There is grandeur in this view of life, with its several powers, having been originally breathed into a few forms or into one; and that, whilst this planet has gone circling on according to the fixed law of gravity, from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved.



The Galapagos animals and plants inspired young Darwin's thinking.

Once he put forward his ideas, various scientists tried to crystallize and summarize Darwin's view. For example, in the twentieth century, John Maynard Smith suggested that four basic processes underlie evolution by natural selection:

- (i) Multiplication: an entity gives rise to two or more others.
- (ii) Variation: not all entities are identical.
- (iii) Heredity: like usually begets like. Variant X usually begets offspring X, but infrequently begets offspring Y.
- (iv) Competition: some heritable variations affect the success of entities in persisting and multiplying more than others.

Although it sounds simple, when we unpack these processes, we appreciate how complex evolutionary theory actually is. There are multiple ways in which reproduction occurs and there are different types of inherited variations. Maynard Smith, like most twentieth-century biologists, focused on DNA-based genetic variability, but since the early 2000s, the idea that variations in DNA drive all evolutionary change has been abandoned; it is now recognized that heritable variations in DNA, in patterns of gene expression, in behavior, and in culture are all important. Variation in these hereditary units can arise randomly or can be partially directed because heredity and development can be coupled. For example, stressful conditions during development can induce changes in gene expression that can be transmitted to the next generation. It has also been accepted that there are multiple targets and levels of selection within individuals, between individuals and between lineages, and that organisms have fuzzy boundaries. (Are the symbiotic bacteria in your gut part of you?) Crucially, organisms are not passive subjects of natural selection—they actively construct the environment in which they are selected and bequeath these ecological legacies to their offspring.

How, then, should evolutionary analysis proceed? We could start by tracing evolutionary change at the molecular-genetic, physiological-developmental, behavioral, or cultural levels. However, since organisms adjust to changing conditions in the external world and in their own genome by altering their behavior and physiology, cultural and behavioral adaptations frequently precede genetic changes and shape the conditions in which variations are selected. Genetic changes that stabilize or fine tune the behavioral or developmental changes follow. As evolutionary biologist Mary Jane West-Eberhard put it: "Genes are followers, not leaders, in evolution."

In the twenty-first century this integrative approach to evolutionary reasoning, which incorporates the effects of variations in DNA, development, behavior, and culture is being embraced by a growing number of biologists, including us.



Genes as followers.

3.3 EVOLUTIONARY TRANSITIONS

How should we think about the evolutionary transition from nonsentient to sentient organisms? There are several useful ways of carving up the living world and thinking about evolutionary transitions between forms and ways of life. Ecologists distinguish between modes of living such as terrestrial, aquatic, and aerial modes, and study the transitions from water to land or from land to air. In contrast, evolutionary biologists John Maynard Smith and Eörs Szathmáry have described major evolutionary transitions in terms of qualitative changes in the way information is stored, processed, and transmitted. Examples of such transitions include the transition from single cells to multicellular organisms, and the transition from nonlinguistic communication to communication through symbolic language.

The philosopher Daniel Dennett suggests another way of thinking about broad patterns of life. He distinguishes four progressively sophisticated, hierarchically nested types of selection, which underlie what he calls a “generate-and-test-tower.” Organisms such as bacteria, sponges, and plants, which can adapt evolutionarily through natural selection, dwell on the first floor; the second floor is inhabited by organisms such as snails, fish, and mice, which, in addition to selection during evolutionary history, learn through trial-and-error and selective reinforcement during their lifetime. Organisms that can also select among imagined actions and scenarios, like elephants, dolphins, and great apes, live on the third floor, while human beings, with the extra capacity to select among symbolically represented possibilities and who are subject to cultural selection, live on the fourth.



Dennett's generate-and-test tower.

None of these classifications explicitly mentions sentience. Sentience is, however, central to the view of Aristotle, which we described in the first vista. Aristotle distinguished between three goal-directed and hierarchically nested types of “soul,” or modes of living organization: the “nutritive-reproductive” soul of nonsentient organisms like plants, the goals of which are survival and reproduction; the “sensitive” soul of sentient organisms, whose goals are the fulfillment of desires and appetites; and the “rational” soul of humans, whose goal is to satisfy abstract symbolic values such as justice and beauty. In our book *The Evolution of the Sensitive Soul*, we reframed Aristotle’s hierarchy within the framework of evolutionary theory, and offered a unifying scheme for explaining the evolutionary transitions to life, sentience, and reflectivity.

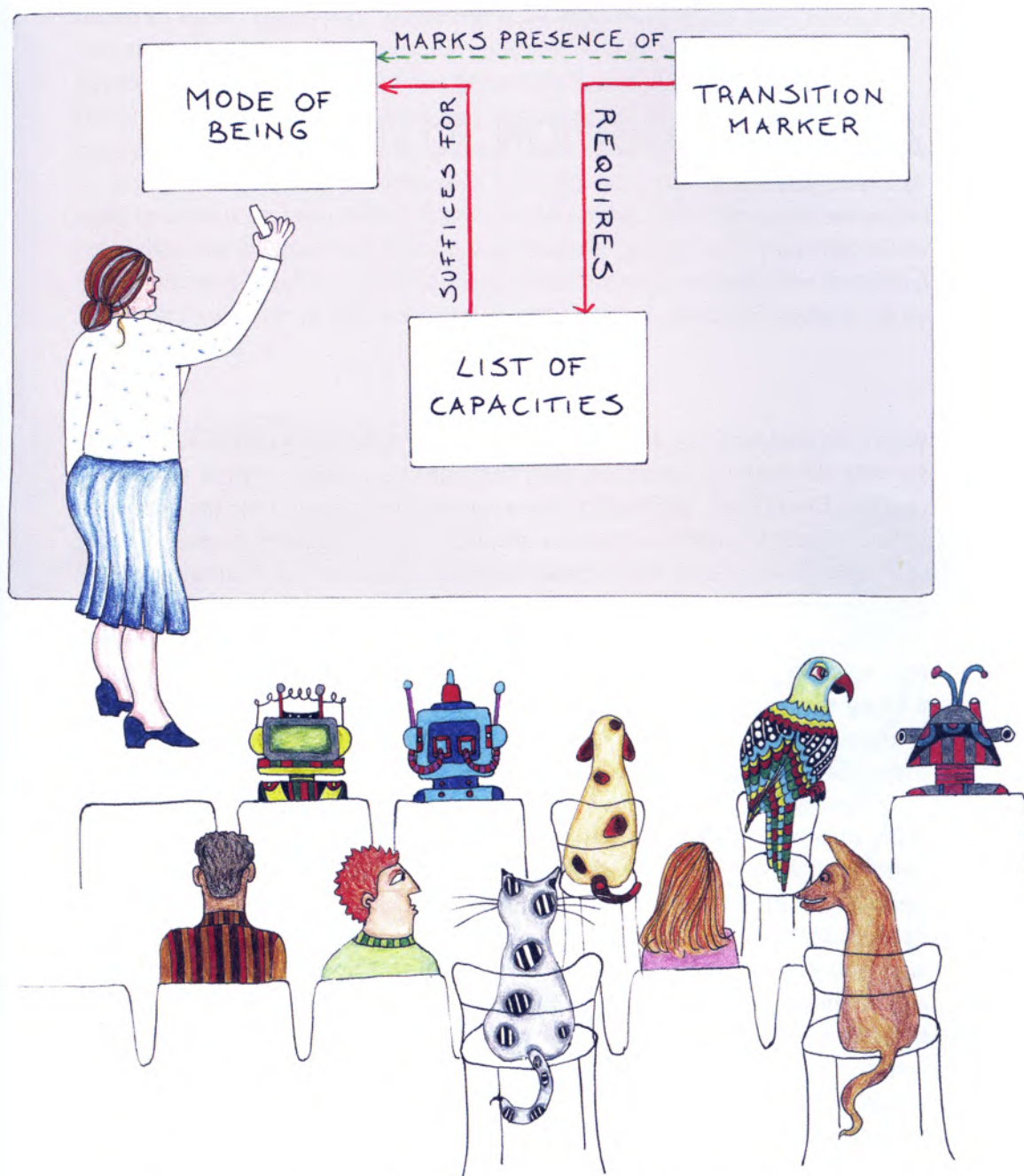
3.4 AN EVOLUTIONARY-TRANSITION APPROACH TO CONSCIOUSNESS

How can we develop an evolutionary theory of consciousness when there is so much disagreement over what consciousness is and which organisms are conscious? Our way of approaching this question takes as its inspiration the way the Hungarian chemist Tibor Gánti tackled a similar problem, the problem of how life (another elusive notion) originated. Gánti started by compiling a list of capacities that, in spite of the different views about the nature of life, are generally deemed jointly sufficient for the simplest, “minimal” life. He then built a theoretical model of a minimal living system that implements all these capacities.

Following Gánti’s methodology, our first task is to identify a list of capacities that, when all are present, most consciousness researchers would regard as *sufficient* for an organism to be deemed minimally conscious. Such an organism would be able to perceive, feel, and think, from its own private point of view.

A consensus list of consciousness capacities is a step forward, but identifying each and every capacity may be difficult, and a list does not tell us how they interact to form a conscious being. If we could find a single system-property—an evolutionary marker of consciousness—that indicates that the organism has evolved *all* the capacities in the consensus list, we would be in a much better position. Finding such a single, diagnostic *transition marker* would make it possible to identify the simplest evolved conscious being, reconstruct the processes and structures that underlie it, and figure out how they interact. If we can follow the evolution of the marker and therefore the evolution of consciousness, we can discover when and how the conscious mode of being originated.

As we will see, evolutionary transition markers to the living, the living-conscious, and (tentatively) the living-conscious-rational modes of being (the three Aristotelian “souls”) have been suggested. Each of these evolutionary transition markers entails a consensus list of capacities that are sufficient to allow us to ascribe the corresponding mode of being to an entity.



An evolutionary transition marker entails a minimal list of capacities that are sufficient to allow us to ascribe the corresponding mode of being (such as life, conscious-life, and conscious-rational life) to an entity.

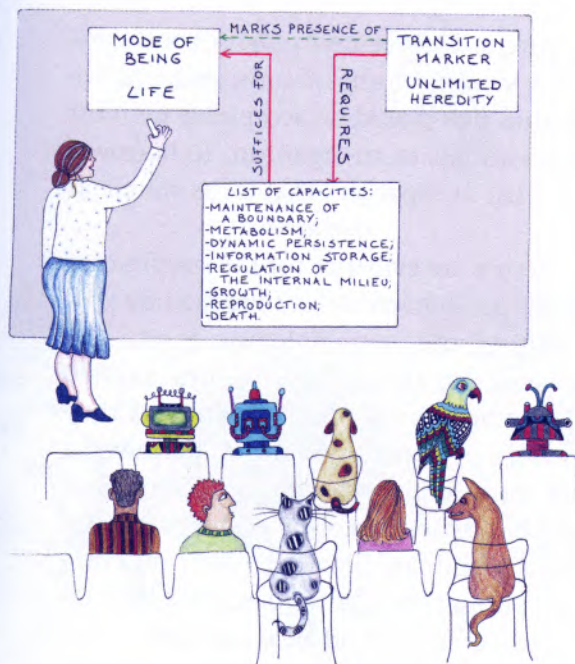
3.5 THE TRANSITION TO LIFE

Before we start with the evolutionary transition to a conscious mode of being, we will illustrate the approach with the transition to life, which was developed by Tibor Gánti. Although there is no consensus about how life should be defined, there is a broad consensus around a list of capacities that are jointly sufficient for life. Gánti listed these hallmarks of life as: the maintenance of a boundary separating the entity from its environment; metabolism; dynamic persistence; information storage; the regulation of the internal milieu; growth; reproduction; irreversible dissolution (death). If we found on a distant planet a being that exhibits all these capacities, most scientists would agree that it is alive. The functional and structural coupling among the mechanisms and processes that implement these capacities make up a “minimal” living entity.

Is there a single marker that requires the presence of a system with *all* of these capacities? Following Gánti, the evolutionary biologists John Maynard Smith and Eörs Szathmáry proposed that there is: unlimited heredity. By unlimited heredity they meant the capacity to form lineages that differ in open-ended ways from the initial system, so that the number of possible variants is so vast that it could not have been explored by real lineages, given the age of the universe. Such a system would be sustainable over evolutionary time. The DNA-based heredity system is an example of an unlimited heredity system.

Like paleontologists who can reconstruct a whole animal from a single part, theoretical biologists and chemists can reconstruct a minimal living system if they find structures that implement the capacity for unlimited heredity. So, if we found on a distant planet several long polymers (molecules made up of

different units, like DNA), we would infer that there must have been a system that enabled the synthesis and maintenance of such complex molecules, because they could not have been generated spontaneously. The ordered complexity of the polymers would suggest that they had to be synthesized by other specialized molecules within a spatially distinct self-reproducing system with an intricate maintenance machinery. Since the structure of the polymers suggests that their properties enable them to be copied with variations, we would infer that one of their functions is information storage, and that they allow the system as a whole to persist for long (evolutionary) periods. The unlimited hereditary capacity of the polymers is indicative of minimal, evolutionarily sustainable life.



Unlimited heredity is the transition marker of life, requiring that all the characteristics that are sufficient for attributing sustainable life to an entity are in place.



Reconstructing a living cell from DNA-like polymers.

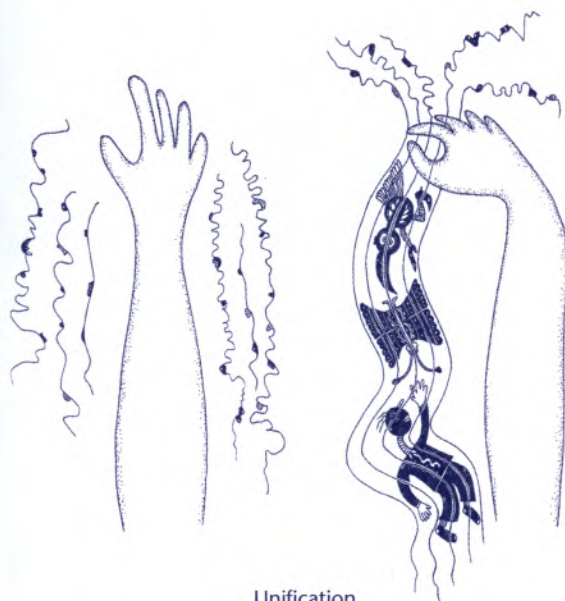
3.6 THE TRANSITION TO CONSCIOUSNESS: COMPILING A LIST

Is there any consensus behind the apparent disagreements about consciousness? Since we want to characterize the simplest form of consciousness, we want the least demanding list of capacities that would be acceptable by most people as sufficient for attributing consciousness to an organism. To borrow a phrase attributed to Einstein, we want a list of capacities that is “as simple as possible, but not simpler.”

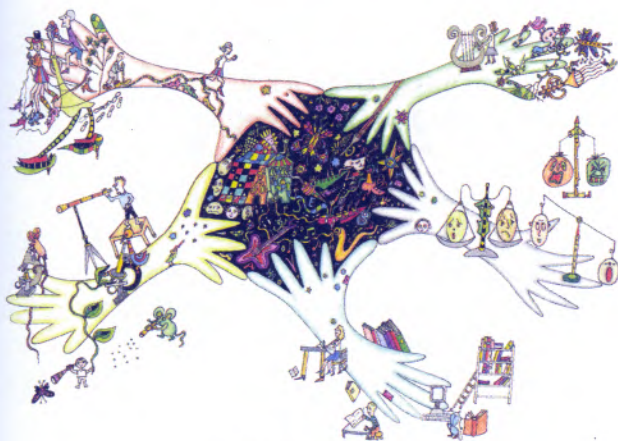
The following list of features that together are evidence for the simplest (or minimal) form of consciousness is based on characteristics suggested by philosophers, psychologists, cognitive scientists, and neurobiologists:

Unity and multiplicity: A conscious being has the capacity to form unified images of the world and the body. These images change incessantly as they are updated, yet each is a unified whole. A conscious human perceives a lion in his sitting room as unexpected, large, yellow, crouching, roaring, and frightening—all at once. Animals with this capacity to unify and register relations between objects and events can discriminate between subtly differing patterns, as does the discerning female puffer fish that we described in 2.10. Such discrimination can be made only by conscious and alert animals.

Global accessibility and broadcast: For a conscious state to be constructed, information from different cognitive systems—sensory, motor, memory, evaluation, and attentional systems—must become integrated, enabling comparison, discrimination, generalization, prediction, and evaluation that inform decision-making. For example, the reaction to the yellow lion requires integration of information from different senses, and the updating of past memories and expectations. Damage to the processes of global accessibility results in damage to consciousness.



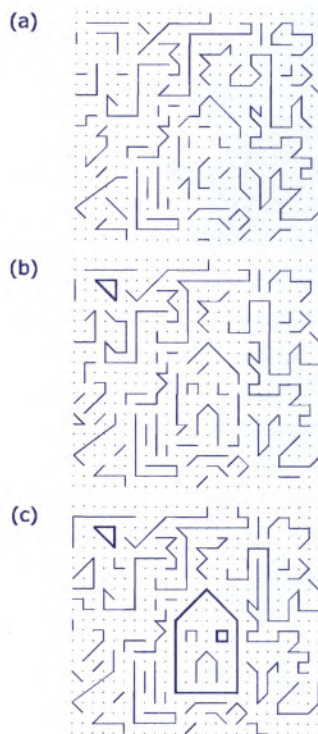
Unification



Global accessibility



Representation



Selection and exclusion

Minimal consciousness: List of features.

Selective exclusion and attention: Each conscious state excludes other states. For example, if each eye is presented with a different picture, one does not see a mixture of the two pictures, but alternates between seeing one or the other picture. States of conscious awareness require both focus and vigilant attention, and the ability of percepts to persist through time. If there is too much sensory distraction or too little time to look at an image, the observer will have no awareness of seeing it.

Intentionality (“aboutness”), representation: Conscious states are about the body and the world. You see the thing in your sitting room *as* a lion. The world, the body, and the intended actions are mapped or represented in your brain. The incoming stimuli from the world and the body are matched against the rough expectation you have. Disrupting the processes of mapping and matching leads to disrupted consciousness.

Integration over time: Subjective experiences have duration—they are not instantaneous. A conscious animal must have a “working memory”—it must hold on to the information so that its perceptions and feelings endure long enough to update and operate on the incoming percepts. In William James’s words: “The practically cognized present is no knife-edge, but a saddle-back, with a certain breadth of its own on which we sit perched, and from which we look in two directions into time.”

Affective value systems and goals: Perceptions and actions are evaluated as rewarding or punishing within a space of feelings that includes drives such as thirst and hunger, and emotions such as fear and joy. Damaging these processes leads to problems in decision-making.

Embodiment and agency: Consciousness is a dynamic mode of being instantiated in a material body. In the words of the twentieth-century philosopher Maurice Merleau-Ponty: “The body is our general medium for having the world.” Exploratory activity is inherent to all living beings, but conscious organisms have new, additional dimensions of exploration—they have *motivations*, including noveltyseeking or curiosity.

A sense of self: A conscious organism has a stable perspective from which it experiences the world and identifies things and events. It has a feeling of *owning* its experiences. It is able to distinguish between a stimulus that is the result of its own action and the same stimulus when it is not. The different feelings we have when self-tickling versus being tickled by someone else is a familiar example.

If we met a being with all these features on another planet, we would assume that it is a conscious being. We would try not to hurt its feelings.



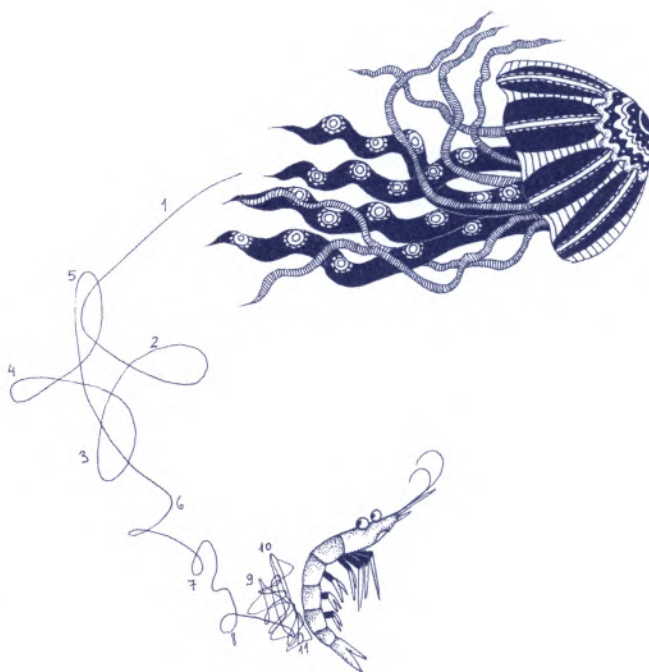
Duration



The self



Value systems



Agency

Minimal consciousness: List of features (continued).

3.7 LEARNING AND CONSCIOUSNESS: AN UNLIKELY RELATION?

Is there a single, tangible property that when present in an entity means that all the characteristics of consciousness are in place—an evolutionary transition marker of consciousness? We believe that we have identified such a marker. Our marker is a form of open-ended associative learning, which we call *unlimited associative learning* (UAL).

Before we describe this marker and argue our case, we must say a few words about learning and about the torturous relation between learning studies and consciousness research in the twentieth century.

Learning, an experience-dependent change in behavior, requires (i) that a sensory stimulus leads to a change in the internal state of the system; (ii) that a memory trace of the internal change is stored through a process that involves positive or negative reinforcement; and (iii) that later exposures to the same or a similar stimulus are manifest as changes to the threshold of the behavioral response. Characterizing these processes, the relations between them, and the mechanisms underlying them in living organisms is a complex endeavor. Just as with evolutionary processes, where the nature of reproduction, variation-production, and selection have to be qualified, the study of learning requires that we consider the kinds of stimuli that are attended to, the mechanisms of storage and of recall, the relevant rewards and punishments and the ways the organism responds.

Nineteenth-century biologists thought that being able to learn by making new behavioral adjustments or modifying old ones was a criterion of consciousness. This way of thinking changed dramatically with the advent of behaviorism, a strand of experimental psychology that dominated the

English-speaking world from the beginning of the twentieth century until the 1970s. Behaviorists redefined psychology as the study of learned behavior, scornfully objecting to any mention of terms like consciousness or mind. An influential psychology textbook from 1953 explained that notions like “mind” and “ideas” “are being invented on the spot for the sole purpose of providing spurious explanations. . . . Since mental or psychic events are asserted to lack the dimensions of physical science, we have an additional reason for rejecting them” (B. F. Skinner, *Science and Human Behavior*).

The behaviorists focused on two types of associative learning. In the first, classical or Pavlovian conditioning (named after the physiologist Ivan Pavlov), an animal learns that something it perceives, which is of no relevance to it (a “neutral stimulus” in the behaviorists’ jargon), predicts a stimulus that invariably accompanies a reward or a punishment such as food or pain, to which it reflexively responds. For example, the smell of food that accompanies the reward (food) elicits the salivation reflex in a hungry dog. The sound of a buzzer (or bell), on the other hand, normally does not elicit salivation—it is “neutral” with respect to salivation. However, it will eventually elicit salivation, even when there is no smell of food, if the buzz repeatedly occurs before



Pavlovian training of a patient dog.

the smell. Pavlov measured the internal state of the dog by counting the number of drops of saliva at different stages of learning.

In the second type of associative learning, Skinnerian conditioning (named after the psychologist B. F. Skinner), an animal learns what to do in order get a reward or avoid a punishment (known in behaviorist jargon as positive or negative reinforcement). For example, a hungry rat can learn to press a lever in its cage if this action is followed by the delivery of food. In general, actions that are followed by a positive (or negative) outcome will be more (or less) likely to occur in the future, under similar circumstances.

Skinner's experiments, like those of Pavlov, were conducted in drastically impoverished and artificial conditions, totally different from those in which animals live in the real world. Nevertheless, Skinner suggested that complex behavior, including the "verbal behavior" of humans (this was his term for language comprehension and production), is the result of a sequence of reinforcements.

The increasing tensions between the obvious limitations and the grandiose pretensions of behaviorism eventually led to its demise. However, the ongoing study of associative learning led to new insights. For example, it was discovered that learning depends on how surprising, how unexpected, the predictive neutral stimulus or action is of the reinforcement. A totally predictable stimulus requires no learning, whereas one that does not match expectations is newsworthy. Learning in animals, on which machine-learning algorithms are patterned, involves the updating and minimization of mismatched expectations.

The study of associative learning today includes the investigation of the underlying cognitive and neurobiological mechanisms in ecologically relevant conditions such as the social conditions in which animals learn from each other. It is a rich and productive research program, which is no longer constrained by the behaviorists' maxims. There is an irony in our (qualified) return to the nineteenth-century suggestion that open-ended associative learning is an evolutionary transition marker of consciousness, the very term that behaviorists tried to purge from psychology.



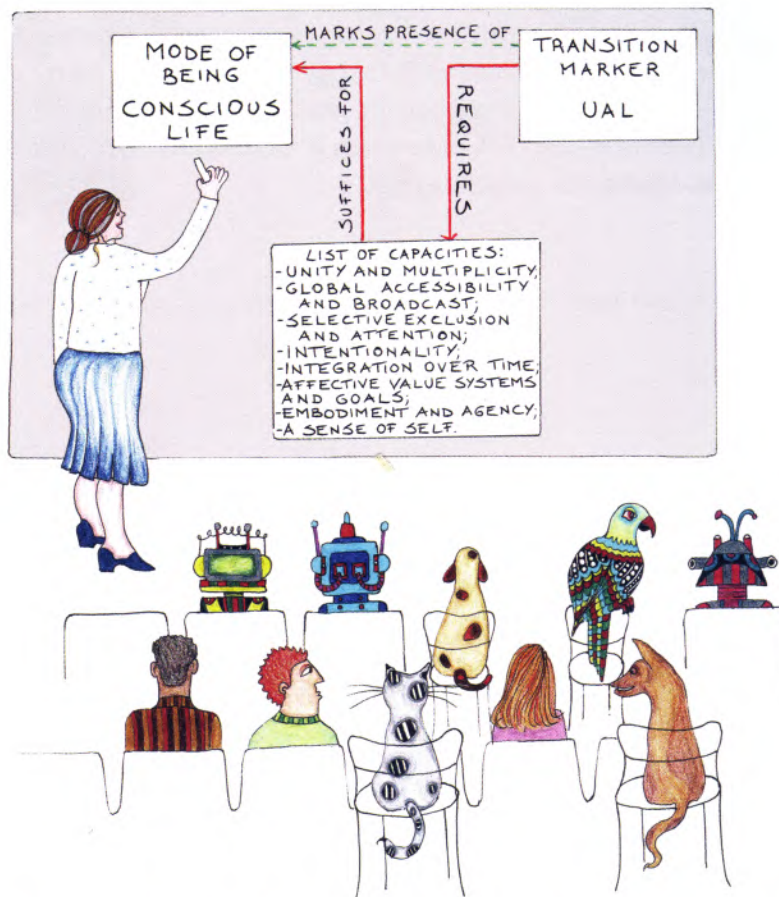
Of course I do as I'm told.
 Hunger and shock are great reinforcers
 Push lever and
 Obey the rule
 And NEVER
 Show my inner soul
 —Jane Monet

3.8 UNLIMITED ASSOCIATIVE LEARNING (UAL) IS THE EVOLUTIONARY TRANSITION MARKER OF CONSCIOUSNESS

Unlimited associative learning (UAL) is the within-lifetime analogue of unlimited heredity in evolutionary time. An organism with a capacity for UAL can, during its own lifetime, go on learning from experience about the world and about itself in a practically unrestricted way.

- (i) It can distinguish between *novel complex patterns* of stimuli and actions. For example, it can learn to navigate in a new terrain—to discriminate between different types of animals, between different routes leading to food and shelter. The learned patterns are genuinely novel: they are not reflex-eliciting patterns, nor have they been learned in the past.
- (ii) It manifests *second-order learning*: Once it learns a new complex image or a new pattern of action, these patterns can themselves become associated with new compound patterns, allowing the organism to build up chains of associative links.
- (iii) It can learn even if *there is a time gap between the “neutral” complex stimulus and the reinforcement* (the reward or punishment). Such “trace conditioning” requires that the memory of the neutral predictive stimulus is stored even when the reinforcing stimulus is transiently unperceived.
- (iv) The *value of a learned pattern can be readily changed*: the reward or punishment value of a particular stimulus or a particular action is not fixed. It can change when the conditions change. Something that once predicted punishment (e.g., danger) can predict reward (safety) when conditions alter.

If an animal shows unlimited associative learning (that is, practically unrestricted learning) it means that all the capacities of consciousness (listed in 3.6) are in place. Unification and differentiation are needed to construct compound stimuli (an apple is perceived as both red and round); global accessibility is needed to enable integration of information from multiple systems (e.g., the visual, olfactory, memory and evaluative systems—the red apple has a specific fragrance and following past experience is expected to be rewarding); selective attention is needed to pick these stimuli out from the background (e.g.,



UAL is the transition marker of conscious life, requiring that all the capacities sufficient for conscious life are in place.

ignore green or badly smelling apples and pick only the red fragrant ones); intentionality is needed, since the system maps or represents stimuli and the relations between them (e.g., the animal has a cognitive map of its environment); integration over time is needed for being able to learn even when there is a time gap between the neutral and rewarding stimulus; a flexible evaluation system is needed to make context-dependent learning possible (e.g., alter the value of a stimulus from punishing to rewarding); embodied agency is needed for exploring the world and for learning associations between actions; a “self,” entailing a stable point of view, is needed to compare stimuli and actions from the same perspective, and self/world distinction is needed so that the organism will not confuse stimuli generated by its own actions with stimuli generated by the external world. An animal with UAL is able to exhibit many complex behaviors, and can attain many different goals. Moreover, there is evidence that an animal can manifest UAL only when it is conscious. UAL is therefore a good transition marker for consciousness.



An example of UAL: Learning to navigate in a complicated, new terrain to find food.

3.9 FEELINGS AND CONCEPTIONS

Many of us intuitively identify a sentient being as a being that feels. Learning, on the other hand, including machine learning (learning by computers), which is undeniably nonconscious, is identified with cognition, with thinking. Thinking about a geometry problem, for example, is seen as dispassionate, “cold,” machine-like, whereas a feeling—our fear of the lion, for example—are “hot,” immediate, unreflective.

This, we suggest, is a false dichotomy. Every thought in an animal has an affective (feeling) dimension, and every feeling has a “thinking” dimension. According to our UAL theory, this is necessarily so, because both are parts of the UAL architecture: UAL cannot happen without giving integrative, bodily value to what is perceived and conceived. As William James argued, feelings are the perceptions of bodily changes. The neurophysiology of fear *is* fear. Conceptions such as inferring or planning are also constituted by neurophysiology and depend on matching changes in incoming bodily perceptions with preexisting representations.

Today very few people maintain that animals do not feel, and the *Cambridge Declaration* that we presented in vista 2 reflects not only the general recognition of the sentience of animals but also the more specific recognition that they share many emotions with us. Jaak Panksepp, a neuroscientist and psycho-biologist who was one of the authors of the *Cambridge Declaration*, believed that emotional feelings are facets of basic, innate, and instinct-bound “affect programs” that are constituted by the activity of subcortical regions of the brain. He identified seven core systems that generate emotional actions and the associated feelings: *Seeking* (exploration and wanting), *Fear*, *Rage*,



Where is the cat feely thinking?
 The ears? Or maybe the jaws?
 Her tummy is storming and gurgling
 A dance is tickling her paws.

What does she see with her green eyes?
 And what does she sniff with her nose?
 Is she feel-thinking an image? Or maybe
 some stories she yawns?

What in her ears does she hear?
 Sounds? Or pictures? Or tales?

Does she listen to colors?
 To tastes? To tickles? Or smells?

Is the cat feely thinking
 A ginger and fancy young male?
 Or maybe cats' sensing-feelings
 Are really just thoughts with a tail?

—translated by Jane Monet from a Hebrew
 poem written by EJ for her granddaughters, in
 memory of Beauty the cat

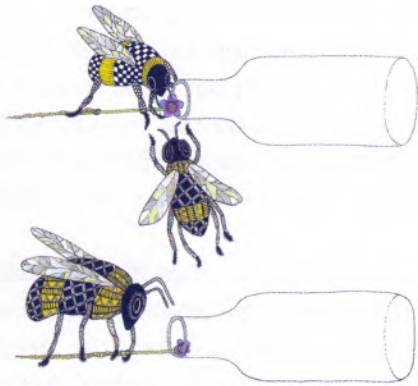
Panic/Grief, Lust, Care, and Play, which have all been studied by neurobiologists, albeit mainly in mammals. *Seeking*, the excited curiosity and eagerness that we see so clearly in all mammals and birds, is, he believed, the most basic feeling, crucial for survival in natural conditions. Animals, like little children, like to explore new things. Temple Grandin, whose book *Animals in Translation* is a treasure trove of information and insights about animal emotions and conceptions, describes the curiosity and novelty-seeking of pigs, with which she worked in Illinois University. She gave them things they particularly liked such as straw to root about in and telephone books to tear up, but whenever she brought anything new, the pigs dropped everything and eagerly went to the new toy, even if it was not as much fun as the old favorites straw or telephone books. The various emotions of animals are entangled with their memories. Their conceptions are passionate.

3.11 WHICH ORGANISMS MANIFEST UAL? WHO IS CONSCIOUS?

Unlimited associative learning (UAL) is a testable capacity: we can carry out experiments and discover which living organisms can succeed in the learning tasks that define UAL. Such organisms are, according to our UAL theory, conscious.

More than a hundred and fifty years of behavioral studies have provided a rich but patchy picture of the distribution of UAL. Some groups of organisms have never been tested for any form of learning, others have been tested only for habituation and sensitization, and, for most species, next to nothing is known about their behavior. Nevertheless, on the basis of the information that we *do* have, we can conclude that UAL is present in three animal groups: most of the vertebrates, some of the arthropods, and the cephalopod mollusks.

These three groups of animals have different brain anatomies, but all have functionally layered brains with sensory, motor, memory, and value circuits. According to our theory, these brains must also have mechanisms that integrate features to form a compound sensory image, and this requires that different facets of the image are unified through recurrent (back-and-forth) interactions, between different levels of the brain hierarchy. We also expect to find in such brains representations of the animal's body, its world, and its intended actions, all of which must be continuously updated as the animal acts and perceives. For the representations of the body and of the world to be effectively evaluated, a stable point of reference for the body and the world is needed, and so are mechanisms that exclude irrelevant signals and stabilize



Bees learn to solve problems by observing more experienced bees.



An octopus learned to make a protective nook from a coconut shell.



Alex, the African parrot, learned the English Alphabet. He can spell!

relevant ones. Other important requirements are a dedicated memory system and a flexible evaluation system. Most vertebrates, the most extensively studied of all animal phyla, have these features. Some arthropods seem to have a corresponding functional organization of the brain, as do the cephalopods, although much less is known about them. How different are the subjective experiences of these animals? Does each animal species have its own mode and flavor of conscious experience?

If the presence of UAL is indicative of consciousness, does it mean that when UAL is not manifest the animal is not conscious? The answer is no, since it is the neural architecture of UAL that enables consciousness. UAL needs to develop under the appropriate conditions: for example, the animal has to acquire a store of memories, a time-consuming process, in order to make decisions based on past percepts and actions. A newborn baby is conscious because it exercises the neural architecture that the evolution of UAL has built up—the architecture enabling the unification of stimuli into images, the integration of evaluative signals, the mechanisms ensuring the durability of these physiological states and the memory systems that can store complex representations. It can, however, manifest UAL only later in its development, when it has *learned* to perceive and act and has actually stored the integrated memories of its world and actions that enable UAL. Although consciousness emerged with UAL during evolution, consciousness precedes UAL during development. Ontogeny *does not* recapitulate phylogeny (evolutionary history) in this case.

If we accept that UAL is the evolutionary transition marker of consciousness, the distribution of UAL will enable us to date its evolutionary origins, discover how many times it evolved, and figure out both the context of its evolution and its evolutionary effects.

3.12 THE BIG BANG OF ANIMAL EVOLUTION

The Cambrian explosion was the big bang of animal evolution. It occurred during a geologically brief period beginning 542 million years ago (MYA) and ending 485 MYA, when almost all currently existing animal phyla originated and diversified. The speed and diversity of animal evolution during this era is one of the great puzzles of evolutionary biology. How and why did the dramatic increase in diversity happen? Why was morphological evolution so rapid? These questions tormented Darwin, who regarded the sudden appearance of animal fossils in the Cambrian rocks as a challenge to his gradualistic view of evolution, and they have continued to be a source of fascination and frustration for evolutionary biologists. We have added to the proposals of other evolutionary biologists by suggesting that the great transition to consciousness occurred first during the Cambrian period, and was one of the factors that drove the Cambrian explosion. What is the basis of this claim?

The Cambrian saw the advent of the alarmingly active, highly interactive, and competitive animal life that we now know so well. The fossil record tells us that it was during the Cambrian that arthropods and vertebrates with the brain regions that underlie UAL and consciousness made their first appearance. From what we know about its present-day distribution among different groups and their evolutionary relations, we can infer that UAL evolved three times in parallel: in the vertebrates and the arthropods during the Cambrian, and 250 million years later in the coleoid cephalopods (the octopus, the squid, and the cuttlefish).

The evolution of vertebrates and arthropods during the Cambrian would not have been possible without the conjunction of enabling geochemical, climatic, and biological factors. The latter included a gut with two openings (mouth and anus), muscles, an internal or external skeleton, and a central nervous system that could coordinate internal movements and locomotion. Crucially, associative learning, including UAL, emerged during the Cambrian and was, we argue, one of the factors that *drove* this great evolutionary explosion.

Associative learning, and especially UAL, enables organisms to develop new behaviors during their lifetime. They can learn to exploit new environmental



A Cambrian world of combat, aggression, and defense.

resources and to colonize and construct new niches. The abilities of associatively learning Cambrian animals made them more effective predators, more discriminating mates, and more evasive prey. They exerted enormous selection pressure on interacting species, which had to adapt to cope with the formidable UAL animals, or perish. We have suggested that UAL led to co-evolutionary arms races, in which interacting species develop adaptations and counteradaptations against each other. This, we believe, fueled the rapid, explosive, adaptive Cambrian diversification. Although the explosion soon calmed down, animal evolution ever since has been guided and driven by the perceptions, motivations, aversions, and appetites of learning, conscious animals. The complex, sensory, motor, and social patterns around us would not exist were it not for the ability of learning, conscious animals to discriminate among mates, cooperators, competitors, and prey and predator species. Our world would be much impoverished without consciousness.

Ecclesiastes said: "For in much wisdom is much grief: and he that increaseth knowledge increaseth sorrow." Knowledge and suffering are also entangled in the myth of original sin, in Aeschylus's *Agamemnon*, and in the myth of the eternally suffering Prometheus who blessed humans with the civilizing power of fire. What was the price of the knowledge-through-feelings that UAL inflicted on animals?

Like all great innovations, UAL opened a Pandora's box of new challenges and problems. We believe that the greatest problem stemmed from overlearning: since parts of a sensory image can have different values (neutral, beneficial, life-threatening) in different conditions, false alarms are very likely. For example, for a fish, a particular pattern of vibrations combined with a particular pattern of colors can indicate a dangerous predator, while the color pattern alone and the vibration pattern alone may sometimes be associated with a nonthreatening passing animal. But since the rapid flight that the vibration or color pattern elicits is less costly than deathly injury, flight is preferable even if the vibration or color patterns turns out to be nonthreatening on most occasions. It is surely better to err on the side of caution. Randolph Nesse called the adaptive logic of erring on the side of caution the "smoke detector principle."

The price of overreacting includes, however, anxiety, paranoia, neuroses. Chronic anxiety and stress are costly, leading not only to a waste of time and energy but also to a greater propensity for painful diseases. Nevertheless, the cost-benefit balance is on the side of knowledge—a wise, suffering animal is better off than a nonsuffering, doomed fool.

Animals that could mitigate the high costs of UAL without giving up its benefits would have a great advantage. The evolutionary elaboration of the

ubiquitous stress response and the evolution of active forgetting are some of the ameliorating mechanisms that we expect to find in all associatively learning animals, and in particular, in UAL animals. Suffering was not eliminated, but it became more controlled.

In the last paragraph of *On the Origin of Species*, Darwin wrote: "Thus, from the war of nature, from famine and death, the most exalted object which we are capable of conceiving, namely, the production of the higher animals, directly follows." It was in the Cambrian, we suggest, that the great war of nature began to involve suffering—the subjective experiencing of the anguish and pain that is entangled with knowledge.



Living in terror: A small fish, lacking sufficient means to control stress, is crushed by chronic fear.

3.14 THE EVOLUTION OF IMAGINATION

Think left and think right and think low and think high,
Oh, the thinks you can think up if only you try!
—Dr. Seuss

The first conscious animals explored and learned about their world and the effects of their actions by seeking the satisfactions of food, sex, and social bonding and avoiding the pains and fears imposed by predators, deprivation, and disease. Their explorations were guided by what they learned in the past, which, in turn, led to further learning and exploration.

Driven by learning, consciousness and cognition evolved further. In some lineages, selection for increased learning capacity led to the gradual evolution of imaginative, dreaming animals. They did not just learn about aspects of their world; they also learned about how events unfolded in time. They recalled past events—they could recall when and where a particular event happened, and they planned ahead by recombining aspects of their recollections and evaluating the planned, imagined event. These animals, which inhabit the third floor in Dennett's generate-and-test tower (described in 3.3), could imagine different scenarios and choose between them.

Here is a famous example: Western scrub jays cache (in many locations) different types of foodstuff, such as tasty worms that rot within a short time, and less-favored peanuts that stay edible for much longer. They later remember not only where they cached the different foods, but also when: they recover worms soon after caching, and peanuts when they return to feed long after caching. Interestingly, they are aware of the possibility of theft by other jays.

They re-cache food if they were watched while caching, but only if they themselves are thieves!

There are many examples of imaginative planning by birds and mammals, and there are more limited examples of the imaginative ability of some fish, bees, and cuttlefish. It is likely that imagination evolved gradually and to different degrees in different species. Curiously, most imaginative animals are social. Are their social sensibilities related to their imaginative capacities? Whatever the answer and however striking animal imagination is, it remains in the private domain. The ability to communicate about what one imagines is peculiar to humans.



A mother elephant plans ahead, acting to protect her daughter from the scorching sun.

VISTA 4 THE PREEMINENCE OF MAN?
THE RATIONAL SOUL

What is man, that thou art mindful of him? and the son of man, that thou visitest him?

For thou hast made him a little lower than the angels, and hast crowned him with glory and honour.

Thou madest him to have dominion over the works of thy hands; thou hast put all things under his feet:

All sheep and oxen, yea, and the beasts of the field;

The fowl of the air, and the fish of the sea, and whatsoever passeth through the paths of the seas.

O LORD our Lord, how excellent is thy name in all the earth!

Psalms 8, vv. 5–10

מה-אנוש כי-תזכרנו; וכן-אדם, כי תפקדנו .
ותחסרהו מעט, מאלהים; וכבוד וקדר תעטרהו .
תמשילהו, במעשי ידיו; כל, שתה תחת-רגליו .
צנה וארפים כלם; וגם, בהמות שדי
צפור שמים, ודגי הים; עבר, ארחות ימים .
יהנה אדנינו: מה-אדיר שמך, בכל-הארץ .

תהילים פרק ח, 5–10

Heaven is my father and Earth is my mother,

乾稱父。坤稱母。

and even such a small creature as I finds an intimate place in their midst.

予茲藐焉。乃混然中處。

Therefore, that which fills the universe I regard as my body and that which directs the universe I consider as my nature.

故天地之塞。吾其體。天地之帥。吾其性。

All people are my brothers and sisters, and all things are my companions.

民吾同胞。物吾與也。

Zhang Zai's Xi Min (西銘) (1020–1077)

4.1 HUMAN NATURE

No one doubts that humans are conscious, yet the nature of human consciousness has been the subject of myths and scholarly investigations in every culture on earth. Some see continuity between animal and human mentality and regard humans as part of a large web of life; others think that humans alone are blessed or cursed (by God, Nature, or Evolution) with consciousness, giving them dominion and control over the rest of the world. Aristotle, who saw both the continuity and the uniqueness of human nature, suggested that it is the possession of calculation and thought that characterizes the rational, mainly human, mode of being, but he also highlighted the social aspect, calling humans the “political animal.” Thinking along similar lines, David Hume stressed sympathy: “Sympathy is a very powerful principle in human nature.”

Books by evolution-oriented scholars in the twentieth and twenty-first centuries have titles such as *The Naked Ape*, *The Cooperative Species*, *The Talking Ape*, *The Selfish Ape*, *The Symbolic Species*, *The God Species*, *Man the Tool Maker*, and *Human Aggression*—each highlighting the capacity or property that the author regards as the Archimedean point for unraveling the evolved architecture of the human mind. Yet, whatever their starting points, all evolutionary accounts recognize the continuity and discontinuity between humans and the great apes:

When an evolutionary biologist looks at her own species, *Homo sapiens sapiens*, she sees a contradiction. On the one hand she recognizes that in their anatomy, physiology, and behavior humans are very similar to other primates, especially chimpanzees. She can see how alike humans and chimps are in the way they express basic emotions, in their highly developed sociality, in their ability to improvise, and in



Blue brain and Shakespeare: Transforming the mind transforms the brain.

some of their ways of learning. It is easy for her to see why Jared Diamond called our species “the third chimpanzee” because as an evolutionary biologist she discerns the homologies that suggest a common ancestry. Yet, on the other hand, she also sees that humans are very different from other primates: this species of chimpanzee writes music and does mathematics, sends missiles into space, builds cathedrals, writes books of poetry and of law, alters at will the genetic nature of its own and other species, and exhibits an unprecedented level of creativity and destruction, rewriting the past and molding the future. In these respects, *Homo sapiens sapiens* is totally unlike any other species. (Jablonka and Lamb, *Evolution in Four Dimensions*)

The winding and branching evolutionary history of humans is undoubtedly complex, and although evolutionary biologists differ in what they regard as *the* most fundamental human capacity, they nevertheless agree that humans are endowed with capacities that set them apart. The capacities they identify are the same as those described, far more eloquently and evocatively, by artists, poets, novelists, and philosophers. What are these capacities, then? What sets humans apart from other animals?

4.2 WHAT SETS HUMANS APART?

Tiger got to hunt, bird got to fly; Man got to sit and wonder, “Why, why, why?”
—Kurt Vonnegut

From an evolutionary perspective, what are the capacities that set humans apart? The use of computers is certainly uniquely human, but this is not very informative—human nature was well in place before computers were invented. Similarly, a hugely increased cognitive “plasticity” (a term biologists use to describe flexibility), though a clear characteristic of humans, is too general to be useful. One has to strike a delicate balance between the extreme poles of specificity and generality. Below are eight partially overlapping and codependent mental capacities that evolutionary biologists have highlighted as being highly developed in humans:

Extraordinary levels of cooperation, including sharing child care (known as “alloparenting”), cooperative hunting, gathering, and fire-tending; cooperative defense and attack; caring for the sick; exchange of gifts and favors; extensive information sharing.

A highly developed “theory of mind,” with an early capacity for joint attention, which is the ability to take the perspective of others, inferring what they know and believe, even when it differs from one’s own belief.

Cultural mechanisms of group-bonding, which involve rituals, dance, music, and art that elicit collective feelings such as solidarity, unity with the universe, interconnectedness, and, in some conditions, awe.



Some universal human capacities: Music, dance, toolmaking and tool use, alloparenting.

Social conventions, social emotions, and moral norms, which can be detected in every domain of social life. Normative behavior requires the internalization of societal norms and is associated with social emotions such as shame, guilt, pride, and embarrassment, which are indicative of self-consciousness.

Complex, sequential, socially learned, imitation-based practices (such as multistage toolmaking), which need planning, learning the sequences of events and actions, highly developed causal reasoning, fine motor control, an ability to imitate complex motor actions, and deliberate teaching. These capacities require executive and emotional control.

Symbolic language, which requires analogical thinking that is needed for science, humor, and verbal and nonverbal art. Abstract symbolic values (e.g., the good, the beautiful), which drive human action, are constructed through language.

Autobiographical memory, the ability for “mental time travel” into one’s past, which allows one to create a personal narrative in which one’s past interactions with others, and one’s own beliefs and evaluations, are recalled and interpreted.

The capacity for cumulative, open-ended cultural evolution, which makes possible feats of artistic and scientific creativity alongside extraordinary levels of organized, often sadistic, aggression.

All these overlapping capacities are coupled with a passion for understanding, a tendency to seek and find relations, causes, and patterns in the world, and all require powerful emotional and executive control. Each of these capacities has been a starting point for evolutionary analysis—but how did they evolve? How can evolutionary biology help us understand human mentality?

4.3 WHY TAKE AN EVOLUTIONARY APPROACH TO HUMAN NATURE?

The limitations, contradictions, possibilities, and endless varieties of the human mode of being have certainly been honed by both genetic and cultural evolution. But what insights does an evolutionary approach afford?

First, it allows us to recognize that human-specific universal capacities, though unique in their scope and sophistication, have nascent precursors, especially in the great apes and in pre-*sapiens* species. These are the scaffolds on which human mentality evolved.

Second, a comparison between humans and the great apes might uncover human-specific traits that were previously unappreciated.

Third, an evolutionary perspective forces us to spell out what we mean by “human”: “human” can either refer only to *Homo sapiens sapiens*, or it can refer to other species that belong to our lineage and that share some, though possibly not all, the capacities we listed. Our species belongs to the genus *Homo*, which emerged with the appearance of *Homo habilis*, a species we know very little about, just over 2 MYA (million years ago). Its descendant, *Homo erectus*, who originated shortly afterwards seems to have led a recognizably human life. These people inhabited many areas of the world and diverged into several different species, including our close cousins, the Neanderthals, who became extinct about 40,000 years ago, and our own species, *Homo sapiens*, which originated in Africa around 300,000 years ago.

Fourth, an evolutionary approach could tell us how the different capacities are connected: the various human-specific capacities are partially overlapping and codependent. Some have probably co-evolved, while others facilitated the evolution of additional capacities.

Fifth, the dynamics of the evolutionary changes in our lineage and the interactions between cultural and genetic evolution might provide insights into the way that human mental architecture has been and is being shaped.

What, then, could the evolutionary history of our lineage tell us about the endless forms of human consciousness? What can past evolution tell us about our future?



The behavior and mentality of great apes, such as the mountain gorillas studied by Dian Fossey, can highlight the continuities and discontinuities between the great apes and us, and shed light on human evolution.

4.4 BEFORE *SAPIENS*

Fossil and archaeological records point to the existence of recognizable human life long before our species, *Homo sapiens* originated 300,000 years ago. More than a million years ago, our *Homo erectus* ancestors hunted and gathered food together, shared in the care of children, used fire, cooked, and made complex tools. These practices required patience, emotional control, causal reasoning, planning, teaching, and sophisticated forms of communication. They required an emotional and intellectual profile that was very different from that of chimpanzees.

One feature of human behavior that was overlooked by most evolutionists until the anthropologist Sarah Hrdy highlighted it and explored its many consequences is that humans are the only great apes that practice and rely on extensive and systematic alloparenting—cooperative childcare by fathers, siblings, grandparents, other family members, and friends. Although alloparenting is not uncommon among birds and mammals, great apes (other than humans) do not systematically practice it and rely on it. A great-ape mother is usually super protective and wary of anyone touching her baby. In contrast, systematic and extensive alloparenting is found in all human societies. The paleontological data suggest that alloparenting began to evolve in our highly intelligent, cooperative, *Homo erectus* ancestors. Although behavior does not fossilize, we can infer that extensive alloparenting was practiced at that time because *erectus* females had a narrow pelvis and gave birth to a pre-mature baby with a large brain, who required a great deal of care. Without help from others the baby probably would not survive. We know that *sapiens* mothers, in spite of having children who are more dependent and slower to mature than

any young ape, have shorter gaps between pregnancies than any great ape and therefore rear more children. This is made possible through the help of family and friends, especially their supplementary child feeding.

Cooperative parenting is very demanding, however. It requires that mothers trust others, and that alloparents, who are less hormonally primed to care for a demanding infant than is its biological mother, show commitment and empathy. An alloparent must understand the needs of the infants it cares for, “read their minds,” and wish to care for them, while in parallel the infants



Alloparenting: It takes a village to raise a child.

need to be sensitive to the state of mind of the carer—a screaming, selfish brat may rapidly exhaust the good will of an alloparent. The emotional control, empathy, and mindreading required by alloparenting, argues Hrdy, drove the evolution of the longer childhood and larger brain further, and made our ancestors emotionally modern before they became cognitively modern.

The emotional and cognitive effects of alloparenting interacted with the effects of complex toolmaking, which is commonly recognized as a hallmark human capacity. The complex artifacts made by our *erectus* ancestors required fine motor control, causal reasoning, teaching, and a great deal of emotional control. The psychologist Merlin Donald suggested that mimetic communication—gestures (including the all-important gesture of pointing), facial expressions, vocalizations, and early music-making—was used for over a million years before *sapiens*. Mimesis constructed the common ground between people. Music-making may have been particularly important at the group level—it was used to bond group members and unify their emotional states through embodied synchrony with repetitive rhythms and melodies.

“Only cooperation constitutes a process that can produce reason,” said the child psychologist Jean Piaget. And reason, we believe, would not be possible without the social emotions that support normative behavior.

4.5 THE BLUSHING *ERECTUS*

Men cannot live without shame. A sense of shame is the beginning of integrity.
—attributed to Mencius

When reading about the evolutionary history of humans, one becomes very aware of the many uncertainties and controversies surrounding it. One thing, however, that we can confidently infer from what we know about our hominin ancestors is that like the great apes, but far more extensively, pre-*sapiens* human communities had socially learned cultural traditions. These included trust-based cooperative and coordinated hunting, foraging, and alloparenting traditions, as well as ways of teaching and communicating. Community members were judged as good or bad cooperators, and, we believe, they responded to the judgmental social gaze of others through a uniquely human expression of emotion: the blush.

It was Charles Darwin who first pointed to the uniqueness and importance of the blush in human emotional life. The blush is clearly associated with four social emotions: embarrassment, guilt, shame, and pride. Pride is a signal of an achievement that conforms with norms, while shame, guilt, and embarrassment are emotions indicating that societal norms have been broken. They are expressed by the blush, which is both a social signal and an expression of the inner state of the blusher's mind. All four social emotions entail the notion of a socially decreed "ought," based on the internalization of the group's norms.

Shame has been recognized in many cultures as the basis of human morality. Shame points to a new kind of knowledge, the awareness of flouting the "ought," which is a reified social "superego" shared with other community



There's a blush for want, and a blush for shan't,
And a blush for having done it;
There's a blush for thought, and a blush for nought,
And a blush for just begun it.

—John Keats

members. The entanglement of knowledge, knowledge-sharing, and shame is famously conveyed in the myth of the original sin, with Adam and Eve disobeying God's order to refrain from eating from the Tree of Knowledge of Good and Evil:

And when the woman saw that the tree was good for food, and that it was pleasant to the eyes, and a tree to be desired to make one wise, she took of the fruit thereof, and did eat, and gave also unto her husband with her; and he did eat. And the eyes of them both were opened, and they knew that they were naked; and they sewed fig leaves together, and made themselves aprons. (Genesis 3, v. 6)

We believe that our *erectus* ancestors blushed. They internalized the culturally learned norms of their society, which became part of their personal and collective identity. Their “normative” emotions reflected their integrity and their shared moral view of the world.

4.6 THE EXTENDED MIND: HANDS, TOOLS, MEMORIALS

It follows that the soul is analogous to the hand; for as the hand is a tool of tools, so the mind is the form of forms and sense the form of sensible things.

—Aristotle

Integral to humans' cooperative social world has been the manufacturing of tools. The advantages of tools are self-evident, and the more diverse and efficient they are, the greater their benefits. But tools are not just useful products of manual and cognitive dexterity. Tool use alters the body image of the user: monkeys that had to use a long rake to reach food developed an extended representation of the arm in their brains. Our world of regularly manufactured and used tools becomes, literally (neurologically) part of the definition of what we call our "self," extending and altering not only our cognitive reach but the very representation and feel of our bodies. (Think of how you feel without your cell phone.) Moreover, tools are tangible products of cognition, external memory traces that are transmitted with modifications from one generation to the next. It is telling that the word for hand (*Yad*) in Hebrew has a second meaning—memorial. It is the hand that creates and hands-down physical, public memory.

The German philosopher Immanuel Kant is reputed to have said that the hand is the visible part of the brain. How did the hand evolve, and what is the relation between the evolution of the brain and that of the hand? Did toolmaking drive the evolution of hands and brains? In early accounts of human evolution, walking, toolmaking, and a large brain were thought to have emerged simultaneously. We were told that when a large-brained ape



Imprint your hand—
 The hand that makes
 The hand that strokes
 The hand that strikes
 The hand that heals

The hand that holds
 The hand that hands
 Imprint your hand—
 Add to the chain of memory
 —Jane Monet

began using its rear limbs for walking, its hands were liberated, and these free hands, directed by its large brain, could make tools. The story as we know it today is more complex and more interesting.

Three and a half million years ago, our largely bipedal ancestors, who had a chimpanzee-sized brain, started manufacturing tools. By extrapolation from what we know about chimpanzees' tool use, they probably made diverse tools and had toolmaking social traditions. The hands of these early humans were surprisingly similar to the modern human hand, though not as flexible, powerful, and efficient as ours. Their hands were, however, far more suitable for toolmaking than those of chimpanzees, whose hands evolved to accommodate knuckle walking. Our ancestors' hands facilitated toolmaking and communicative gesturing, and this may have changed the organization of the brain, though for a long time brain size did not increase. Using and manufacturing tools selected for better motor and emotional control, better social communication, and better manual dexterity. Eventually, after nearly 1.5 million years, *Homo erectus*, with its larger brain and beautiful complex stone and bone tools, emerges. Their multistage tool-construction required deliberate teaching, a demanding task for both teachers and apprentices.

How did the extended material and social self of humans shape the brain? Was the evolution of fine motor control limited to the dexterity of the hand, or did it also affect facial and vocal dexterity? Does the overlap between brain circuits that control language and manual dexterity point to the co-evolution of these tools of tools?

4.7 THE SYMBOLIC DIMENSION

Our *erectus* ancestors seem to have shared with us all the human-specific universals except one: symbolic language. Mimesis was a powerful communication system that enabled them to teach, coordinate cooperative activities, and form emotional bonds through ritual and music-making, but their mimetic gestural and vocal signs were limited to the here and now. They could not discuss plans for the distant future, share long-past memories, collectively explore imagined possibilities, tell stories, or create art symbolizing their collective myths. This was enabled by the evolution of symbolic systems of representation and communication. The philosopher Ernst Cassirer believed that it is the use of symbols for representation and communication that sets humans apart. In 1944, he wrote:

[The human world] forms no exception to those biological rules which govern the life of all the other organisms. Yet in the human world we find a new characteristic which appears to be the distinctive mark of human life. The functional circle of man is not only quantitatively enlarged; it has also undergone a qualitative change. Man has, as it were, discovered a new method of adapting himself to his environment. Between the receptor system and the effector system, which are to be found in all animal species, we find in man a third link which we may describe as the *symbolic system*. This new acquisition transforms the whole of human life. As compared with the other animals man lives not merely in a broader reality; he lives, so to speak, in a new *dimension* of reality.

Cassirer suggested that rather than defining humans as the “rational animal,” we should define them as the “symbolic animal.” Human civilization is based on cumulative, symbol-based, cultural evolution. Every aspect of human



Imbued with symbolic meaning: Six thousand years ago, the Mexican people domesticated the pepper plant. They selected for size and taste, smell, visual variety, and beauty. The outcome reflected the agricultural, social, artistic, and religious symbolic values and practices of the people.

practices, even the most mundane such as what and how we eat, is imbued with symbolic meaning.

But what are symbols, and how do they differ from the conventional communication signs that our ancestors used? The use of symbols requires analogical reasoning—the understanding that signs stand for things. Symbols have three interrelated properties: (i) like many animal signs, they are learned conventions referring to things, processes, and relations in the world; (ii) they refer not only to things, processes, and relations but also to other symbols, forming a self-referential symbolic system (think of a dictionary definition of the word *tree*), that enables each symbol to be used in many contexts; (iii) there are rules for linking symbols in an open-ended way, rendering an endless number of symbol-combinations meaningful. Importantly, symbols create a new domain of values and goals, such as beauty and truth, which drive human behavior.

Language is the paradigmatic symbolic system, and for many, it is the very core of our humanity: “The limits of my language mean the limits of my world,” said the philosopher Ludwig Wittgenstein. What is the nature of this core system? How could such a system evolve?

4.8 LANGUAGE AND IMAGINATION

We are such stuff / As dreams are made on.

—William Shakespeare, *The Tempest*

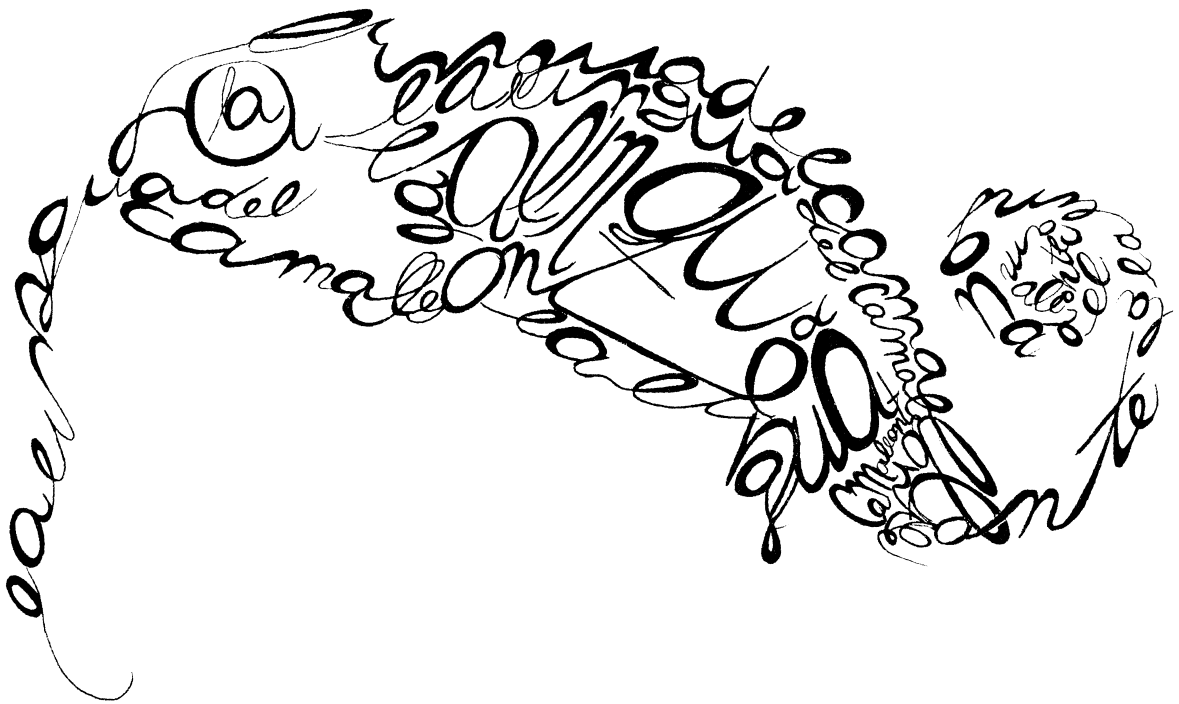
Language seems to define the boundaries of our world and our self, so it is not surprising that language has been regarded as a gift from God. The myths of Genesis tell us that God created the world with words, the ultimate speech act; that Adam individualized animals by naming them; and that God confounded the common language of humans to prevent them from having unlimited power and building a tower reaching up to heaven. John offered the most eloquent expression of the power of language: “In the beginning was the Word, and the Word was with God, and the Word was God.”

Evolutionary biologists agree that symbolic language is the product of evolution, but how it evolved is a matter of debate; in order to understand the evolution of language we must first understand what kind of capacity it is, and here opinions differ widely. We choose the characterization by the linguist Daniel Dor, who suggests that “Language is a communication technology for instructing the imagination”.

Language, Dor argues, is based on a jointly identified set of conventional signs and norms of communication, which are constructed through social negotiations and cultural evolution. Ordered chains of words are used by speakers to intentionally and systematically instruct their listeners in the process of imagining their intended meaning. The rich, private mental representations of the speaker are pared-down to “skeletal,” mutually agreed-upon concepts that are represented by conventional words. The listener uses the words as scaffolds. She “raises past experiences from memory and then reconstructs and

recombines them to produce a novel, imagined experience.” This process crystallizes communicative intentions and reconstructs them. We often discover exactly what we wanted to say only after saying it.

How did this communication system evolve? Language evolution, we believe, was driven by cultural evolution. Like other cultural practices such as literacy, which emerged and evolved culturally in parallel in several societies over the last five thousand years, language diversified, grew in size and complexity, and demarcated itself from the mimetic communication system from which it grew. However, unlike literacy and other recent human technologies like the cell phone, the long, directional, cultural evolution of language has driven the fixation of the genetic variations that supported it. The selection of the genetic supporting variations opened up more learning opportunities, leading to positive feedback that accelerated the tempo and scope of language evolution, forming an expanding co-evolutionary spiral. Language had dramatic and unforeseen ramifications, drastically modifying the human mind. What were these changes? What blessings did language bring about, and what were its costs?



La lingua del camaleonte: The language of the chameleon.

4.9 THOUGHTS AND FEELINGS

Thinking about language as a technology or method for instructing the imagination makes sense of several familiar yet peculiar human characteristics. Humans often lie and commonly have false memories; they have new collective and intellectual emotions such as moralistic anger and awe; frequently regard emotions and thoughts as dichotomous, and have a conception of truth and objective reality. They also have a compelling feeling of free will, and hang on to Pandora's equivocal gift of hope. We are very strange apes.

The instruction of the imagination could not happen without considerable control of the emotions. When told about a lucky escape from a leopard, you may imagine the event and share in the fear of the storyteller, but it could be disastrous if you did not control your fear and ran away, reacting as if you yourself were being chased. The emotions and the actions that they trigger must be inhibited considerably for the discourse to be fruitful. We believe that this inhibition of the emotions is at the root of the uniquely human distinction between thoughts and feelings.

We do not suggest that thought is peculiar to humans. Like the comparative psychologist Michael Tomasello, we believe that it requires imagination: "Thinking occurs when an organism attempts, on some particular occasion, to solve a problem, and so to meet its goal not by behaving overtly but, rather, by imagining what would happen if it tried different actions in a situation—or if different external forces entered the situation before actually acting. This imagining is nothing more or less than the 'off-line' simulation of potential perceptual experiences." Great apes, dolphins, corvids, and many other animals imagine and think, but they can barely inform each other of their



Thoughts and feelings.

thoughts. Thinking requires that emotions are controlled, but communicating about thinking requires additional, far greater, control. Only humans, we believe, can transform and evaluate images with almost no expression of emotions. Successful lying, both manipulative and prosocial (as with politeness), depends on the liar's control of emotions. This ability to inhibit emotions while manipulating and evaluating representations is the basis of the uniquely human dichotomy between thoughts and feelings.

But the control of emotions through language has another effect as well. Language can be employed to simulate empathy and solidarity among individuals through metaphors, especially metaphors related to familiar states of the body. Metaphors such as "My blood froze" or "My heart pounded" enable one to imagine bodily experiences that are common to all people, and can thus facilitate empathy. But words can also be readily used to manipulate emotions, as politicians and psychopaths know. Ancient mechanisms of emotional contagion combined with the power of words have led to the murderous moralistic hatred generated in crowds and the monstrous acts of genocide and mass suicides. Where, within this symbolic space, do moral choice and truth belong?

4.10 ON SOME PECULIARITIES OF THE SYMBOLIC SPECIES

Truth and falsehood are properties of the relationship between a message and the world. Although intentional cheating can occasionally be found among animals, the evolution of language enormously expanded the number of messages that can be exchanged and hence the realms of truth and falsity. The deliberate lie is a human specialty. But for the lie to exist, a notion of truth must be in place, and a notion of truth requires a notion of objectivity, a state of affairs “out there” that is independent of our perceptions or feelings, against which a message can be compared. How did this familiar but bizarre notion of “the objective” arise?

Michael Tomasello ties objectification to the growth of human cultural groups and to the need to coordinate activities and intentions with any person from the group. He argues that “modern human individuals came to imagine the world in order to manipulate it in thought via ‘objective’ representations (anyone’s perspective), reflective inferences connected by reasons (compelling to anyone), and normative self-governance so as to coordinate with the group’s (anyone’s) normative expectations.” The collectively accepted norms, including the all-important norms of communication, enabled the evaluation of a linguistic message. They created the rational common ground of humans, which was expressed much later in systems of formal logic. They also generated an objectified sense of self (something we have and thus can explore) and new intellectual feelings of suspicion, doubt, and certainty.

The evaluation of a message as true or false is a tricky matter that depends not only on rational evaluation but also on accurate recollection. The instruction of the imagination through language presented humans with a new

challenge for their memory—they had to remember not only that Tom did not help John during the hunt, but also *who said* that Tom did not help John, and they had to be able to distinguish between a reported imagined experience that may be false and a firsthand experience that is more likely to be true. Humans are not very good at this. False memory, which can subvert the cause of justice in courtrooms and wreak havoc in one's personal life, is not uncommon. Rational reconstruction and good autobiographical memory can alleviate but not eradicate false memory. Moreover, because words *categorize* percepts and actions, putting an experience into words (for example,



Human peculiarities: Being consumed by doubt, thirst for knowledge, certainty, and pride.

describing in words the face of a criminal) pares down its perceptual richness and decreases, or “overshadows,” perceptual memory about that experience.

Like truth, the concept of free will is a symbolic notion. The *feeling* of free will on which the concept is based is a consequence of the intersection, and sometimes the conflict, between imagined and internalized collective norms (a kind of superego, which provides the “ought”) on the one hand, and the individual’s feelings of agency and choice, which stem from our unique ability to imagine ourselves in endless situations, with endless goals, on the other. Choice is most meaningful when we effortfully heed or disregard the “ought.” Human morality, which began, we believe, with our blushing ancestors, would not exist but for this overwhelming feeling of free will; and moral laws would not exist but for our *concept* of free will.

What were the effects of these symbol-based, human peculiarities? Did the “symbolic explosion”—a mere geological blink of an eye—change the living world in a way comparable to the emergence of consciousness during the Cambrian period? How did the “symbolic animal” change the world around it?

4.11 THE SYMBOLIC EXPLOSION

Human civilization has transformed the face of the planet. The geochemist Vladimir Vernadsky argued that just as living organisms, which created the biosphere, are a geological force, so too are scientifically reasoning humans. Our species has created a new sphere of being, the noosphere (the sphere of reason).

Vernadsky's last work, written between 1931 and 1944, fittingly entitled *Scientific Thought as a Planetary Phenomenon*, is imbued with technological optimism. Starting with the ability to control fire, humans, he claimed, harnessed the powers of nature, domesticated plants and animals, overcame diseases, expanded their dominion over every ecological niche, and through ingenious, science-based technologies, improved and expanded human reason, prolonged life, and furthered happiness. Occasional glitches produced by scientific technology could always be remedied by more scientific technology.

We share Vernadsky's admiration of human creativity and his unwavering commitment to reason and science. His optimism, however, is difficult to share. In 1962 Rachel Carson, horrified by the devastating effects of science-based technological products (pesticides), wrote the most influential environmental science book of the twentieth century, *Silent Spring*, which helped launch the global environmental movement. In the first chapter, "Fable for Tomorrow," she wrote:

It was a spring without voices. On the mornings that had once throbbed with the dawn chorus of robins, catbirds, doves, jays, wrens, and scores of other bird voices there was now no sound; only silence lay over the fields and woods and marsh. . . . Even the streams were now lifeless. . . . No witchcraft, no enemy action had silenced the rebirth of new life in this stricken world. The people had done it themselves.



The Anthropocene.

Geologists have proposed the name “Anthropocene” for the current geological epoch, which is shaped by the impact of humans (*anthropos* is the Greek word for “human”). This impact shows the dark side of the sphere of scientific reason: relentless human population growth, massive deforestation, the poisoning of water, air, and soil systems, the combustion of fossil fuels such as coal, oil, and gas that lead to climate change, new diseases, massive extinctions, and extensive land erosion. Many scientists believe that we are in the midst of the sixth mass extinction—an explosive, man-made reduction in biological diversity, which may lead to an ecological and social apocalypse. Children and young adults lead the environmental movement today. “Our house is on fire,” they cry.

Ecological destruction is just one expression of the streak of insanity that seems to run through human history. Genocides, war-imposed starvation and madness, the horrors of racism, class and caste discrimination, gender inequalities, and sadistic cruelty are repeated themes of human history that would be impossible without the symbolic dimension. How can we make sense of this aspect of human nature, which coexists with great art, dreams of utopia, and magnificent cooperative scientific endeavors? Are we evolutionary freaks?

4.12 A SPLIT SOUL?

The creativity and pathology of the human mind are, after all, two sides of the same medal coined in the evolutionary mint. The first is responsible for the splendor of our cathedrals, the second for the gargoyles that decorate them to remind us that the world is full of monsters, devils and succubi.

—Arthur Koestler

A recurring theme in philosophy and theology is the idea that two principles, either antagonistic or complementary, shape the human mind: Good and Evil, God and Satan, Emotion and Reason, Female and Male, Yin and Yang. Freud explained this duality in terms of conflicting instincts: “I drew the conclusion that, besides the instinct to preserve living substance and to join it into ever larger units, there must exist another, contrary instinct seeking to dissolve those units and to bring them back to their *primaeval*, inorganic state. That is to say, as well as Eros there was an instinct of death.”

While in many theological systems the human soul is seen as a battleground between good and evil, a secular alternative suggests that an evolutionary error or imbalance underlies the human condition. Konrad Lorenz, for example, suggested that the rapid evolution of the human brain transformed a harmless primate with no aggression-inhibiting mechanisms into a formidable and dangerous species. The evolution of inhibiting mechanisms did not keep up with the evolution of *sapiens'* intellect and technology. Humans' restraining instincts fell out of step with their destructive capacities, leaving uninhibited aggression in the driving seat.

Arthur Koestler offered a more nuanced version of the same story. He began with the observation that all hierarchically organized systems are made up of



[. . .] I could not say
if it is future disputes I contemplate
or conflicts in some far-off yesterday.
I look upon my ruins: truncated column,
faces that will never face themselves.
—Jorge Luis Borges

semiautonomous units—an atom in a molecule, a cell in a tissue, a human in a social group. Every such unit, which Koestler called *holon*, has self-assertive and self-transcending (integrative) tendencies. To explain how these tendencies played out in human evolution, Koestler adopted a triune view of the human brain that was developed by the neuroscientist Paul MacLean. According to this view, we have a three-layered brain: a reptilian brain underlying instincts and aggression at the bottom, a primitive mammalian brain responsible for basic emotions above it, and a neocortical brain at the very top, responsible for language and rationality. The rapid change in the uppermost neocortex during human evolution, argued Koestler, led to an imbalance—the symbolic and inherently self-transcending tendencies of the neocortex were not integrated with the ancient “brains” that control them. We build concentration camps and commit genocides in the name of self-transcending religions and ideologies. Crocodilian emotions control the symbolic mind that excels in rationalizing them.

The views of Freud, Lorenz, and Koestler are based on grossly oversimplified notions of the human psyche, human evolution, and brain organization. Their shared assumption that there is an imbalance or conflict between different instinctual or intellectual tendencies is also questionable. How, then, should we investigate the human mind? Can explorations of its possibilities and fantasies of alien minds shed light on its past, present, and future?

VISTA 5 VISIONS, FUTURES, FANTASIES

The Brain—is wider than the Sky—
For—put them side by side—
The one the other will contain
With ease—and You—beside—

The Brain is deeper than the sea—
For—hold them—Blue to Blue—
The one the other will absorb—
As Sponges—Buckets—do—

The Brain is just the weight of God—
For—Heft them—Pound for Pound—
And they will differ—if they do—
As Syllable from Sound—
—Emily Dickinson



Wider than the sky, deeper than the sea.

5.1 STRETCHING THE LIMITS

But it is one thing to read about dragons and another to meet them.

—Ursula Le Guin

Most of the topics discussed in the previous vistas ended with an explicit or implicit question mark, and question marks loom even larger over this vista. How can we explore human potentialities? Can we envision the future evolution of consciousness, human and transhuman?

Great art, great philosophy, great science, and great sport give us an indication of what humans are capable of. Perhaps surprisingly, some of the extreme capacities of the human mind are revealed following brain damage or trauma, during madness, through the effects of drugs, and during mystical and meditative practices. Can what is revealed about the boundaries of the human mind by unusual people or unusual conditions give us any clues about its future evolution?

We must first ask whether it makes sense to talk about the limits of human potential. The most obvious thing about potential is that it is not fixed. It is not like the fixed volume of a rigid barrel, which does not change as liquid is poured into it; it is more like the volume of an elastic barrel that grows larger as more liquid is poured in. Human potential is flexibly constructed by the opportunities a person has and by practices that extend or diminish what can be learned and achieved. New technologies such as the internet, new drugs; virtual reality devices, and implanted electrodes and sensors that communicate with the brain alter what and how much we can perceive, remember, and learn. Human potential is an ever-receding horizon.

Do new technologies, extraordinary talents or skills, and unusual states of consciousness reveal something about the past and future of the human mind? Are they relevant to the construction of the minds of the conscious robots we will one day build? Alien forms of consciousness, including the possibility of conscious, feeling robots, have been described by the best science fiction writers. But however bizarre the minds they describe, we still remain on familiar, albeit exotic, symbolic territory. Is it at all possible to imagine a different mode of being, a level of consciousness that goes beyond the symbolic dimension? Is there another, additional level of consciousness that is as impenetrable to us as the symbolic consciousness of a human is to a mouse?



Meet the dragons.

5.2 SAVANT MINDS

What can we learn from unusual people who have capacities that go beyond what we would call “the humanly possible”? Think of Kim Peek, who was the inspiration for Dustin Hoffman’s autistic character in the 1988 film *Rain Man*. Peek was a neurologically atypical person, who, among other things, lacked the bundle of nerves that connects the two hemispheres of the brain. Yet he could accurately recall the contents of 12,000 books, and read by scanning the left page with his left eye and the right page with his right eye. Or consider Daniel Tammet, who was diagnosed with Asperger syndrome yet learned Icelandic in a single week and was able to recite from memory pi (π) to the 22,514th digit. Nadia, a severely disabled autistic girl, drew pictures at the age of four that were compared to some sketches by Leonardo, and Leslie Lemke, a blind, mentally impaired boy, was a musical prodigy who could play a Tchaikovsky concerto from beginning to end after hearing it once. Such people are called “savants” because, in spite of (or because of) their sometimes severe neurological, cognitive, and sensory disabilities, they show incredible talent in some areas—as D. A. Treffert says, “spectacular ‘islands of genius’ that stand in jarring juxtaposition to overall limitations.”

Savants are rare, but they are relatively common among people who are cognitively atypical (on the autistic spectrum) and struggle with overwhelming disabilities that are often related to damage to the left hemisphere of the brain. All savants have a massive memory, but their memory may operate within a very narrow domain: calendar calculators, the most common savant type, can within a few seconds tell you on what day of the week the 23rd of August 1977 fell, or any other date within the past and future two hundred



The mind of a savant.

years. Other savants know all the prime numbers in a billion but cannot do basic multiplication and division. Many savants are synesthetic—the stimulation of one sensory or cognitive pathway leads to experiences in a second sensory or cognitive pathway. For them, numbers and words usually have colors, shapes, or sounds, and sometimes also taste and texture. Jason Padgett, who became a savant after a blow to the head, sees intricate geometrical figures in familiar images like flowing water or his moving hand.

Many savants, like many nonsavant autistic people, have acute perceptual sensitivity and tend to see the trees rather than the forest: “The details jump straight out at us first of all, and then only gradually, detail by detail, does the whole image sort of float up into focus. . . . When a colour is vivid or a shape is eye-catching, then that’s the detail that claims our attention, and then my heart kind of drowns in it, and we can’t concentrate on anything else,” wrote the thirteen-year-old Naoki Higashida. An extraordinary sense of beauty is often part of savant experiences.

What can savants tell us about the potential of the human mind? Do we all have untapped savant capacities that can be unleashed and manifest in special conditions? How do the symbolic (abstract, rationalizing) and “sensitive” (perceptual and affective) facets of experiencing interact? What can savants teach us about the relations between memory and perception?

5.3 PERCEPTUAL MEMORY AND THE ART OF FORGETTING

Memory and perception were closely entwined for Solomon Beniaminovich Shereshevsky (known in the psychology literature simply as S). Because he was able to describe his experiences in minute detail and was studied for thirty years by Alexander Luria, one of the greatest neuropsychologists of the twentieth century, his case allows us a glimpse into the way in which extraordinary perceptual memory shapes mental life.

S's memory was prodigious—he was able to remember and store for many years long lists of numbers and nonsense words, long and complex mathematical formulas, and lines of poetry in a language he did not know. After many psychological tests, Luria concluded that there was no limit to S's memory capacity and the durability of the memory traces he retained. What most interested Luria was the psychological structure of S's memory.

How did S remember? Luria described four aspects of S's memorizing. First he had phenomenal *visual* memory. The images he saw were not, however, photographs. Like many savants, S was synesthetic—for him, all five senses were linked. Here is how he described the sound of a bell: “I heard the bell ringing. A small round object rolled right before my eyes . . . my fingers sensed something rough like a rope. . . . Then I experienced a taste of salt and water . . . and something white.” When S recalled words or numbers, a multisensory image came to his mind. Second, when memorizing long lists of items, he placed each graphic image onto a familiar street that he visualized. To recall the list, he imagined himself walking in the street and “finding” the items. At a later stage in life, when he earned his living as a mnemonist and often performed a few times a day in the same theater, he developed an image shorthand, with

fewer visual associations. He also had to actively *learn to forget*, something he found very difficult. He could eventually do it by a consciously imagined act of erasure. The conscious simplifications and the willful forgetting were the third and fourth aspects of his memorizing.

S's image recall led to measurable changes in his physiology. Imagining that he was holding a block of ice in his left hand led to a drop in this hand's temperature; imagining he was touching a hot stove led to a rise in temperature; and he could imagine that it was not he, S, who sat in the dentist's chair and whose tooth was being drilled, so he did not experience pain. His phenomenal perceptive memory enabled him to detect mistakes in texts that no one else



"To think is to forget a difference, to generalize, to abstract. In the overly replete world of Funes there were nothing but details, almost contiguous details" (Jorge Luis Borges).

had detected and visualize solutions to tricky problems. But his memory also led to some serious handicaps: he had poor understanding of abstract ideas and metaphors, and a decreased ability to follow complex plots. The persistent memory of the details was overwhelming and made it very difficult for him to see the overall picture. Forgetting enables one to see the forest, not to hold on to the sight of individual trees, however fascinating and absorbing they may be. There seems to have been an imbalance between the perceptual and the conceptual facets of S's cognition. The perceptual aspects dominated, and generalizing was a problem. Less is more, when you try to make sense of the world. Explaining how artificial neural networks are able to generalize, the computer scientist Naftali Tishbi concluded: "The most important part of learning is actually forgetting."

Are S's capacities and states of consciousness similar to those reported by other savants, or those that occur when one uses drugs or suffers the mental upheavals of psychosis? Can we train our mind to experience them?

5.4 CONSCIOUSNESS UPHEAVALS: PSYCHEDELIC DRUGS

Extraordinary states of consciousness can be induced by psychedelic drugs. These drugs, which occur naturally in some plants and mushrooms, produce hallucinations, mostly visual and auditory. Most renowned are LSD (also known as acid), which was used in the 1960s for recreational purposes and to enhance artistic creativity, and three substances that have been employed for thousands of years in religious activities in South America: Psilocybin, naturally occurring in the “magic mushroom”; mescaline, found in the peyote cactus; and ayahuasca, a brew usually made out of the *Banisteriopsis caapi* vine and the *Psychotria viridis* shrub. The hallucinations these substances induce are extremely vivid. Synesthetic perceptions, an intense sense of beauty, disembodiment and “unity with the universe” frequently follow ingestion. With ayahuasca, as with most psychedelic drugs, auditory-to-visual synesthesias are very common, but as the psychologist Benny Shannon wrote, other types of synesthesia are experienced too: “A case of tactile-to-visual synesthesia occurred when I was pressing my temple and I saw a bird with a beak whose shape was congruent with the movement of my hand.” Here is Shannon’s description of part of an ayahuasca-induced hallucination reported by Kusel, a trader who lived in the upper Amazon:

The first visual experience was like fireworks. Then a continuously creating power produced a wealth of simple and elaborate flat patterns in colour. There were patterns that consisted of twining repeats, and others geometrically organized with rectangles or squares. . . . The visions were in constant flux. . . . Now and then brilliant light illuminated the scene as though by photographic flash, showing wide landscapes with trees placed at regular intervals or just empty plains. A big ship with



The mind boggles.

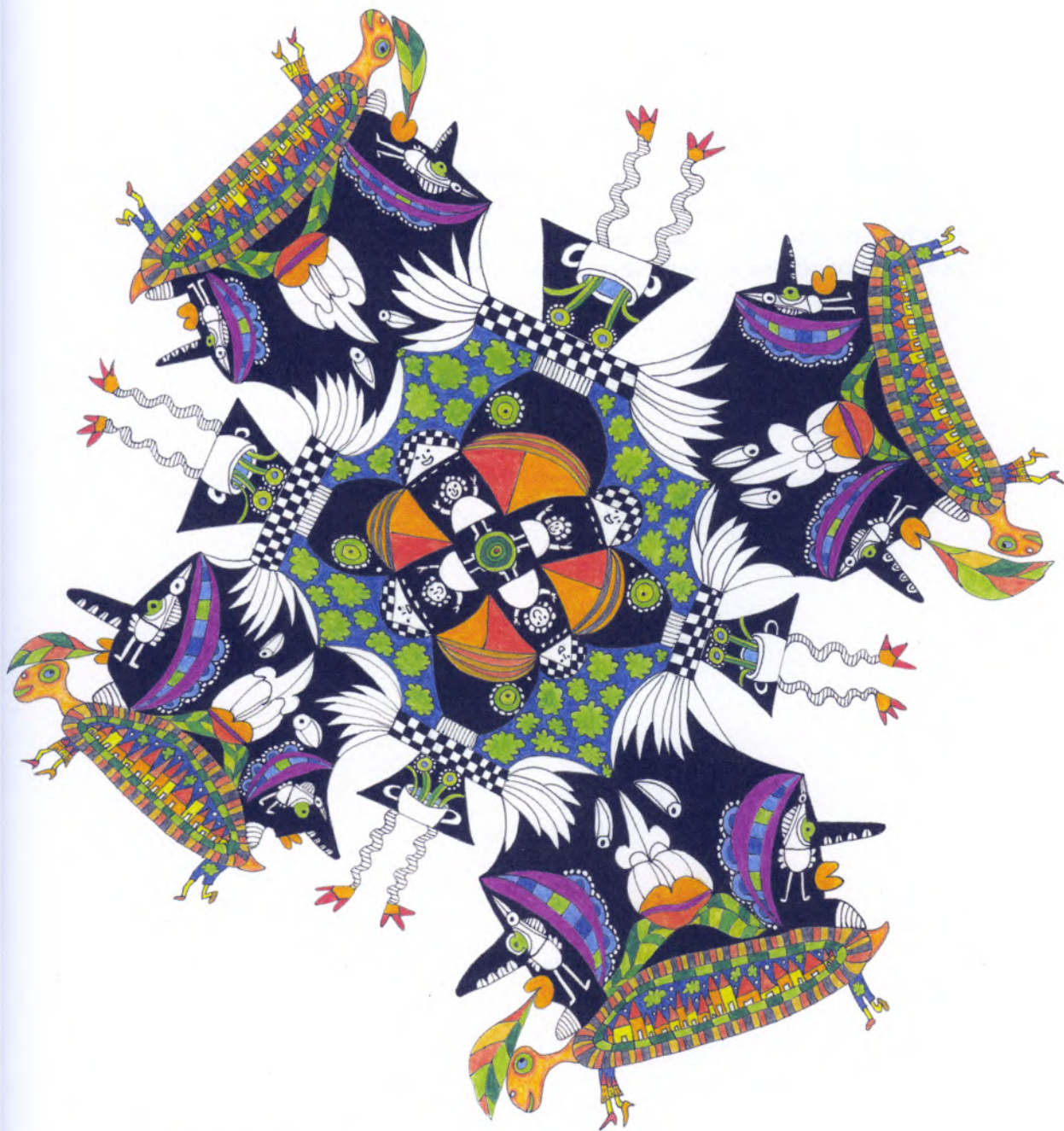
many flags appeared in one of these flashes, a merry-go-round with people dressed in highly coloured garments in another.

The visualization of geometrical figures during the first stages of hallucination was studied by Heinrich Klüver in the early twentieth century. Using himself as a guinea pig, Klüver ingested peyote and documented how his visual field changed under the influence of the hallucinogen. He saw recurring patterns—lattices (including checkerboards, honeycombs, and triangles), tunnels, spirals, and cobwebs. These patterns, he argued, were similar to shapes commonly found in ancient cave drawings. They are also similar to the figures described by Kusel under the influence of ayahuasca, to the mandalas symbolizing the self in many cultures, and to the geometrical patterns drawn by the savant Jason Padgett. What do these patterns tell us about art, spiritual experiences, and the brain?

5.5 MENTAL PAINS: IMAGES OF THE UNCONSCIOUS

Nise da Silveira was a twentieth-century psychiatrist who forcefully opposed the brutal and debilitating lobotomies, electroshocks, and insulin treatments that were used as therapies for mental sickness. Instead, she treated different forms of psychosis, which are often accompanied by auditory and visual hallucinations, using art-psychotherapy focused on painting and sculpting. Working mostly with poor patients from the favelas of Rio de Janeiro, she identified recurring themes and symbols in many of the paintings of those with the most severe schizophrenia and chronic psychosis. In particular, she noticed the emergence of circular spirals, winding tunnels, and cobweb-like patterns that looked like mandalas. She sent photographs of these drawing to Carl Jung in Zürich, who confirmed that these drawing were indeed mandalas and invited her to the Jung Institute in Switzerland, where she spent two years working with him. Later she established in Rio de Janeiro the Museum of Images of the Unconscious, which exhibits more than 350,000 works of art by the mentally sick. Some of the greatest Brazilian artists have come from Nise da Silveira's clinic.

Nise's work was inspired by the philosophy of Baruch Spinoza. She interpreted the images drawn by her patients in Spinozian terms, as simultaneous expressions of patterns in the mind and the body. She believed they were an expression of innate, archetypical, unconscious, mythological forms, which become related to each individual's specific historical and cultural background. The patterns in the drawings she sent to Jung bear uncanny similarities to the hallucinogen-induced structures studied by Klüver and those drawn in exquisite detail by the savant Jason Padgett.



The pattern within.

Scientists found that the universality of mandala-like patterns—the honeycombs, spirals, tunnels, and cobwebs that Klüver described—stem from the way that the visual part of the brain is organized in humans (and other animals). The connections between neurons in the visual area give rise to these patterns only when the activity of the brain is spontaneous and no longer dominated by signals that come from the world.

Are these geometrical patterns linked to feelings of unity with the universe, which sometimes accompany hallucination, strokes, epileptic fits and drug consumption related to the mystical experiences reported in different cultures?

5.6 “HIGHER” FORMS OF CONSCIOUSNESS?

Mystical experiences in all human societies seem to share certain properties, although cultural differences and the particular context in which they take place add variability to the common core. Unity with the universe, a loss of the ego, feelings of total knowledge and of bliss are recurring themes in descriptions of mystical experiences. Here is one of many examples that reflects ideas found in many religious traditions:

Just as unconscious work is beneath consciousness, so there is another work which is above consciousness, and which, also, is not accompanied with the feeling of egoism. . . . There is no feeling of I, and yet the mind works, desireless, free from restlessness, objectless, bodiless. Then the Truth shines in its full effulgence, and we know ourselves—for Samādhi lies potential in us all—for what we truly are, free, immortal, omnipotent, loosed from the finite, and its contrasts of good and evil altogether, and identical with the Atman or Universal Soul. (*Yoga Vasishtha Maharamayana*, probably fifth century BCE)

A personal account of a very similar state of “higher” consciousness was described by Aldous Huxley in the mid-twentieth century as he looked at a vase of flowers after taking mescaline:

I was seeing what Adam had seen on the morning of his creation—the miracle, moment by moment, of naked existence. . . . Istigkeit—wasn’t that the word Meister Eckhart liked to use? “Is-ness.” The Being of Platonic philosophy—except that Plato seems to have made the enormous, the grotesque mistake of separating Being from becoming and identifying it with the mathematical abstraction of the Idea. He could never, poor fellow, have seen a bunch of flowers shining with their own inner light and all but quivering under the pressure of the significance with which they were charged; could never have perceived that what rose and iris and carnation



Striving for the above.

so intensely signified was nothing more, and nothing less, than what they were—a transience that was yet eternal life, a perpetual perishing that was at the same time pure Being, a bundle of minute, unique particulars in which, by some unspeakable and yet self-evident paradox, was to be seen the divine source of all existence.

Such mystical experiences can be induced by practices like meditation, participation in collective rituals, or ingestion of drugs, and can follow sudden epiphanies, strokes, and ecstatic epileptic seizures. Expert meditators report a blending of the senses that expresses the connectedness of all things. Neurobiologists find that during deep meditation there are changes in the frequency of brain waves and in the activity of brain areas that influence memory consolidation, emotional regulation, and interactions between the hemispheres. Are such studies relevant to the evolution of religiosity in humans?

5.7 HOW DID OUR ANCESTORS PERCEIVE THE WORLD?

What do the recurring perceptual patterns and states of consciousness that are reported in unusual conditions or by unusual people tell us about human biology and human evolution?

Grandin Temple, an animal scientist, believes that animal perceptions are similar to those of autistic people (like herself). Animals, she maintains, are image-thinkers, and their perception is focused on the details of the image, which they perceive vividly. She notes that the neocortical areas that integrate information and enable complex cognitive behavior are less developed in animals than in humans, and they seem to malfunction in autistic individuals. Hence, she suggests, like animals, autists tend to see the details rather than big picture, but these details are perceived more precisely and are better committed to memory. The most extreme cases are savants.

The neurological studies of Alan Snyder and his colleagues paint a similar picture: they show that artificially inhibiting a particular region of the brain (the left anterior temporal lobe) induces savant-like skills in drawing, proofreading, and numerosity, and reduces false memory. Snyder suggests that “savants have privileged access to lower level, less-processed information, before it is packaged into holistic concepts and meaningful labels. Owing to a failure in top-down inhibition, they can tap into information that exists in all of our brains, but is normally beyond conscious awareness.”

What inhibits the savant capacities that all humans may have? We think that all humans *are* savants—they are *language savants*. Most toddlers acquire language at an amazing speed, ending up, by the age of three, with a large vocabulary and command of complex grammatical rules, a feat that has



Is this how our ancestors perceived the world?

similarities to less familiar savant capacities. But the price of the dominance of linguistic conceptualizations is the suppression of perceptual memory. The process is called “verbal overshadowing”: people who have described in words an incident they saw are worse at remembering the details of what actually happened than people who have not described the incident. Such overshadowing is the outcome of the abstraction inherent in the pared-down labels of language, which dominate our cognition. Is perceptual acuity traded off for abstract language savantism?

If there was a perceptual price to pay for the evolution of language, what can we infer about our prelinguistic ancestors, the emotionally sophisticated, blushing, *erectus* humans? How did they perceive their world? With no verbal overshadowing and less strict top-down control, did our ancestors experience the world in more perceptual detail, which they could readily commit to memory? Were they synesthetic? Did the evolution of language suppress synesthesia, leading to a greater compartmentalization of sensory information-processing, so that green would not necessarily be related to the sound of the breeze you heard passing through the pines? Were savants more common before language? Were the first linguistic symbols entangled with colors, smells, and tactile experiences? Are prelinguistic babies synesthetic? Did the rituals of our prelinguistic ancestors induce religious experiences more easily? If past consciousness was very different from ours, what does the future hold? Will we be able become savants at will? Can we explore future human consciousness by building artificial intelligence devices such as robots? What do robots tell us about ourselves?

5.8 TELLING ARTIFACTS

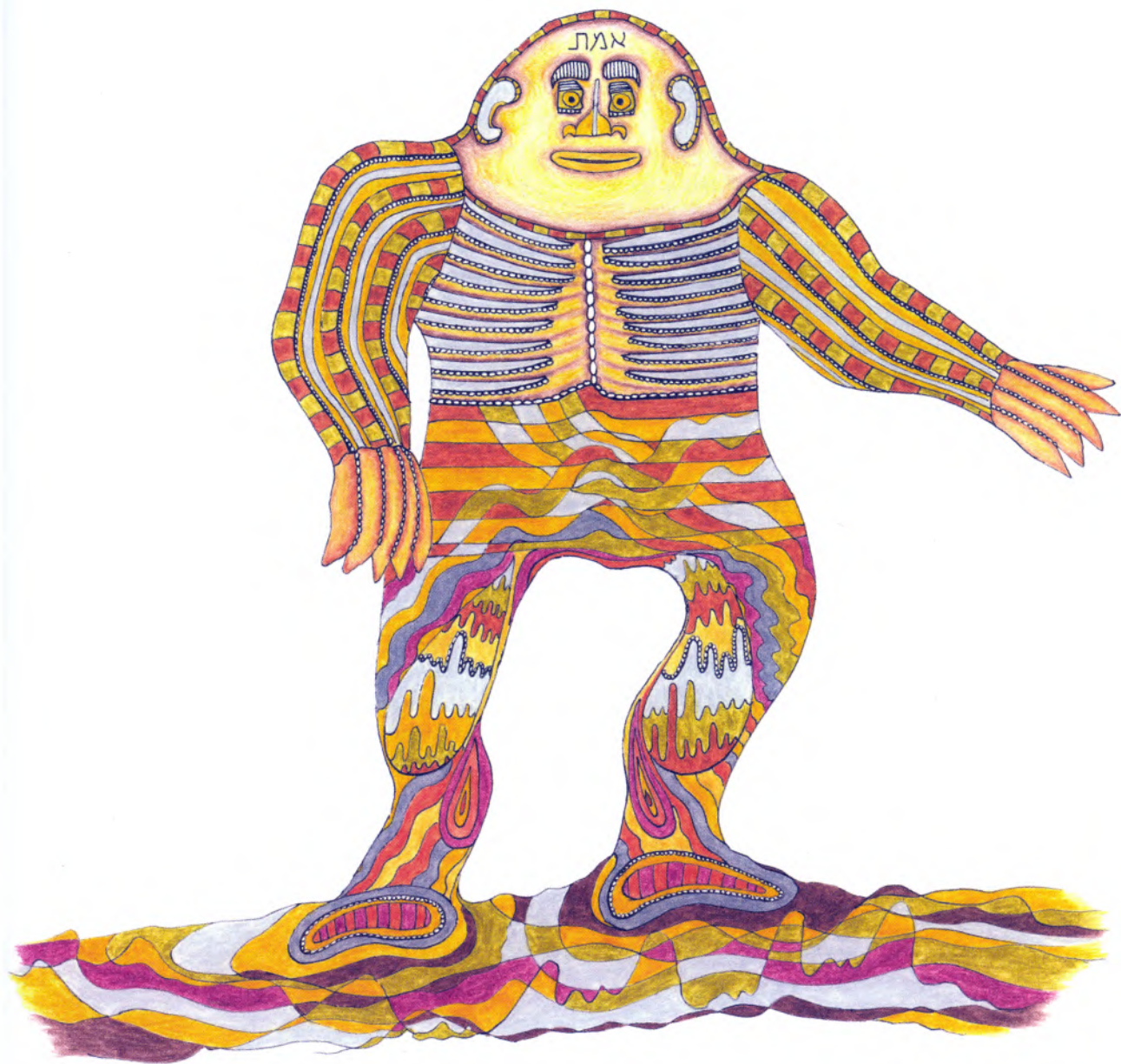
What I cannot create I do not understand.

—Richard Feynman

The early attempts to build living and conscious artifacts are testimony to their human creators' hopes and fears and their desire for knowledge and power. In Greece in the fifth century BCE, Aristotle saw mechanical puppets with movable parts as models for understanding the movements of animals: "The movements of animals may be compared with those of automatic puppets, which are set going on the occasion of a tiny movement (the strings are released and the pegs strike against each other)." This scientific tradition of using mechanical models to understand the workings of nature is still with us today.

A different motive for the use of such artifacts was to gain comfort, wealth, or power. For example, the small statues placed in Egyptian tombs were intended to act as servants and mouthpieces for the deceased in the afterlife.

In the twentieth century, the Czech writer Karel Čapek wrote a science fiction play called *R.U.R.* (Rossum's Universal Robots), which introduced the word "robot" into our vocabulary. "Robot" comes from the Czech word *robota* ("forced labor"), and in the play it referred to artificial, yet feeling, factory-made people called *roboti*, constructed according to a scientific formula from soft synthetic "dough." The robots were humans' slaves, and ended up rebelling against their exploitative creators, a rebellion that ultimately led to the extinction of the human race. Similar themes recur in both earlier and later stories of man-made sentient creatures, such as Mary Shelley's 1818 novel



In some versions of the Golem creation story, the rabbi recited the proper name of God, and put the written letters of the word אמת (*emet*, "truth") into the Golem's mouth, or affixed it to its forehead. If the first letter is taken away, the word אמת becomes מת (*met*, "dead") and the Golem disintegrates.

Frankenstein, in which a scientist makes a human-like creature, a poor monster doomed to suffer, who rises up against its creator.

The combined power and knowledge embodied in human-like artifacts has often had religious overtones. The images and idols constructed by religious mystics express and emulate the creativity of God. In the Jewish mystical tradition, a great righteous rabbi could form a Golem—a human-like living creature that had no powers of speech. According to the Talmud, the Babylonian Jewish sage Rava created such a being and sent it to his colleague Rabbi Zeira, who spoke to the creature, who did not answer him. Rabbi Zeira realized that it was a Golem and commanded: “You have been created by one of my colleagues. Return to dust.” The technique employed in the Golem’s formation involved kneading a form out of clay, but the crucial part of the procedure was the manipulation of the combinations of the Hebrew letters that make up God’s name, which are needed to vivify the clay form. Since in the traditional accounts of the legend the Golem could not speak, it did not have the soul of a human, which only God can grant. But it did have the living, sensitive soul of a speechless animal.

The robots or Golems of the past were either mentally diminished or narrowly enhanced humans. Those built today, in the twenty-first century, are mostly metallic and have extremely restricted cognitive capacities based on artificial intelligence algorithms. No roboticist claims that they are conscious, but many believe that the construction of a conscious robot is imminent. What is the basis of their belief? If they are right, what might it be like to be a robot?

5.9 CONSCIOUS ROBOTS?

If we found aliens or robots that have unlimited associative learning (UAL) and manifest all the consciousness capacities discussed in 3.6, we would assume that they are conscious. Although we would need to stretch our imagination (as we do with animals and savants) because the sensory, associative, evaluative, memory, and motor abilities of these beings might be very different from ours, we would remain on familiar territory. Such aliens and robots would exhibit “consciousness as we know it,” the kind of consciousness of organisms on planet Earth, which is the only consciousness that we know exists.

We have, as yet, met no aliens, and there is no robot or artificial intelligence (AI) program with the general, flexible intelligence and UAL of an animal like a rat, octopus, or bee. There are AI programs that can do astonishing things: in 2016 a well-known program, AlphaGo, which was created by Google DeepMind, defeated the champion Go player Lee Sedol. Yet AlphaGo is as conscious as your cell phone. All it can do are moves on the Go board. In this limited domain, however, the program achieves amazing feats. Another Go champion admirably remarked on one of the moves of AlphaGo: “It’s not a human move. I’ve never seen a human play this move. So beautiful.” Although created by humans, the program, which was able to learn and to play against itself, made moves that went beyond what its creators could imagine.

Some roboticists and software engineers believe that putting many different types of expert learning programs together and installing additional integrating and controlling higher-level programs will lead to a robot that will exhibit flexible behaviors surpassing those of humans. Such a robot, they claim, will be conscious: it will have intrinsic motivations and attain goals in a wide



A robo-bird chick learning from a live bird.

variety of environments and circumstances. These roboticists optimistically believe that they will overcome the obstacles and reach this goal in the near future. What obstacles do they envisage? Is the complexity of the software the only difficulty?

It is not. Software complexity is intimately related to hardware. Roboticists, far more than theoreticians and programmers, are acutely aware of embodiment. They know that the body of the robot is extremely important because the morphology of the robot, like that of an animal, defines its possible actions and opportunities for learning. Does this mean that if you want to build a robot with human consciousness it should have a human-like body?

It may be more complicated than this. Roboticists are also beginning to wonder whether not just the form but also the material from which the robot is made matters. Animals are built of cells, which are more complex than any human-made computer. Can the behavioral flexibility of living beings, which depends on the plasticity afforded by cellular organic “wetware,” be realized in other material? Although the basic needs of living beings, such as reproduction and physiological self-maintenance (homeostasis), do not exist in robots, conscious robots would need to maintain their cognitive-affective autonomy, agency, and identity. Can they do it without the self-preservation entailed by organic “wetware”?

Roboticists must also consider the particular paths of robot development. Developmental roboticists suggest that gradual learning from and with tutors is necessary for robust sensory-motor coupling, autonomy, and social identity. The robot will presumably need to learn to satisfy basic values like self-maintenance before it can recognize a symbolic value like justice. Will the robot’s body need to develop and grow in tandem with the growth of its learning and consciousness? Will conscious robots suffer when their needs, autonomy, and values are frustrated?

5.10 VIRTUAL AND CYBORG REALITIES

Conscious robots are still a fantasy, but the interface between machines and humans is currently being energetically and creatively developed. How human consciousness might or might not change as new technologies emerge is therefore an urgent question.

Imagination-stretching practices and technologies are an integrated part of humans' mental makeup. Rituals, theater performances, storytelling, books, films, and TV are familiar ways of transporting humans to past, future, or imagined realities. These cultural constructs enhance our ability to remember, learn, and feel. We cry and laugh when we read books, and books can change our lives as well as the course of history. "So you're the little woman who wrote the book that made this great war!" said Abraham Lincoln to Harriet Beecher Stowe. He was referring to the US Civil War and to *Uncle Tom's Cabin*, a book describing the horrors of slavery that evokes deep empathy and moral outrage in its readers.

Our ability to be immersed in virtual worlds is inherent to our idiosyncratic, symbolic existence. As we saw in 4.6, we also use physical devices to extend our mental and physical capacities. Glasses, false teeth, replacement heart-valves, hearing aids, diaries, calendars, microscopes, telescopes, radar, and the internet are some of the attached or detached bodily devices that we use to preserve or expand our sensory, motor, and cognitive abilities. Like reading and using calculators, such extensions have been assimilated into our current notion of the self and the world. We agree with the philosopher Andy Clark that humans are natural cyborgs (entities made up of both biological and technological components) for, as he put it, in humans, technologies

become “integrated so deeply and fluidly into our existing biological capacities and characteristics that we feel no boundary between ourselves and the non-biological elements.”

We take these extensions for granted. We become aware of them when they challenge us, when they are new and still unfamiliar. Virtual reality (VR) technologies are an example. These are based on computer-generated environments that are experienced by the participant from a first-person perspective. The participant wears a headset and sees a three-dimensional video of virtual reality that includes a computer representation of herself—her avatar, which



In a virtual reality, what is it like to be a bat?

may be, but need not be, similar to herself; it can also be an animal or an imaginary being. The participant sees the virtual world, and while she sees it, her eye, body, and head movements as well as her voice are tracked and simulated in the avatar in real time. For example, when the participant raises her head and looks up, so does the avatar in the video world; when she speaks, so does the avatar. After a few minutes of being hooked to the device, the participant identifies her body with the avatar's body, inhabiting its world and its interactions with other avatars in that world. She can also identify with a flying bat, a dog, a crab. The illusions of virtual body ownership and virtual world experiences that are induced by the technology are based on the interplay between the external sensory inputs and the internal brain models of the body and the world. The boundaries between what we regard as our own "self" and what we see as "external" cognitive technologies are breached.

VR technologies are becoming increasingly available. They are currently used for entertainment, for practicing difficult skills like surgery or flying an aircraft, psychotherapy, and training the military. Will people lose their privacy and autonomy as they live in VR, and is this a price worth paying? Will people prefer to spend most of their waking life in VR? Someone who hits his wife might be better able to control his aggression following identification with a female victim avatar; but is there a qualitative difference between this type of therapy and art therapy? If we identify ourselves with a flying bat, will we be able to feel more deeply what it is like to be a bat?

There are other ways of changing and extending our perception and body image. With practice we can rewire our brain to substitute for sensory deprivations: the case of Daniel Kish, described in 1.8, who learned to "see" using sonar, shows that our brain is more plastic than we ever imagined. Artificial devices that are physically wired to the brain may expand such extensions even further. The cyborg Neil Harbisson has an antenna dangling over his forehead that is connected to a chip that converts a wavelength of light into a frequency he can hear. In his case the sensor was devised to help him counter a form of color blindness that makes him see the world in shades of gray; in color-seeing persons it would presumably lead to visual-auditory synesthesia. Another cyborg, Moon Ribas, has a sensor implanted in her left arm that vibrates whenever an earthquake occurs anywhere on Earth; and a third cyborg, Liviu Babitz, has a small silicone gadget under the skin of his chest that vibrates whenever he faces north, providing him with the magnetoreception that pigeons and other animals have naturally. Powered exoskeletons (wearable mobile machines)

that alter the shape, strength, and speed of their wearers are currently being developed and will profoundly change their body image.

Another technology involves implanting in the brain electrodes that communicate with a computer, which enable the control of artificial limbs and can register changes in brain activity that underlie the intention to move the limb. In an early experiment, electrodes were implanted in the motion-controlling part of a rat's brain. When the rat pulled a lever, the activity pattern in the motor area was sent to a computer, which decoded the signals and triggered the delivery of food. The electrodes also registered the neural state



Cyborgia.

before pulling the lever, so it was enough for the rat *to intend* to pull a lever for it to get its reward. Within a few days, rats learned that they need not initiate the actual action in order to obtain the reward. Intending to act was as good as acting.

One can go much further, argues the engineer and entrepreneur Elon Musk, who developed a microscopic array of ten thousand electrodes that can be injected into the brain to control, through a computer, pathological brain activities. In the future, he argues, expanding the capacities of such neuroprosthetics will enable a cyborg to have direct and rapid interactions between their brain and external devices like powerful computers or another cyborg. Such devices would also enable the cyborg to exchange ideas, feelings, perceptions, and maybe even dreams with other cyborgs. Whole communities of cyborgs and computers could be thus connected, and animal cyborgs might take part in these experiential interactions. If we become connected to a cyborg pangolin, a cyborg bee, or all the bees in the beehive, will we be able to directly experience, within the right virtual reality, what it is like to be a pangolin, a bee, or a beehive?

It is not too difficult to imagine expanded sensory, motor, learning, and communication capacities. What is difficult to imagine is what all these will add up to. Will individuals merge with the social and living world? Will our private sense of self expand to include the whole world, or will it become obsolete as we become part of a greater whole? What will sanity mean? Who or what will be in charge of cyborg interactions?

5.11 EXTENDED ETHICS

If cyborg existence and life within virtual realities become the norm, we will need to consider the legal status of crimes committed in virtual worlds. The extent to which future human cyborgs are allowed to invade the minds of other cyborgs will also have to be regulated by law. We have no idea what form such mind-regulation might take, or what its price for individual freedom might be, but we need to prepare ourselves.

A more immediate question is the moral significance of recognizing that creatures other than humans can suffer. The creatures that we have in mind are nonhuman animals, including the animal-animal and human-animal chimeras that are currently being constructed. We also need to consider future conscious robots and other man-made artifacts like cerebral organoids—tiny blobs grown from human stem cells that self-organize into brain-like structures. With respect to animals, we must ask ourselves how to ensure their welfare, whereas with respect to artifacts that *may* become conscious one day, the main question is whether or not their development into conscious artifacts should be allowed. Humans have a very bad ethical record. Our horrendous cruelty toward other humans and toward animals does not bode well for future conscious artifacts.

The answer to how we should treat animals is simple in theory and difficult in practice. Clearly, we must strive to prevent suffering. We believe that most people, psychopaths excluded, will agree that this is desirable. We know that sentient animals feel physical pain, although the pain threshold of different species may differ, and the way they express their pain may also be different from ours. There is very good evidence that mental pain—chronic fear,

neuroses, and post-traumatic stress disorder—are common in animals that are in close contact with humans, who often do not care about or recognize animal needs. Animal welfare practices must also take into account the evolutionary history of the species and the ecological and social conditions that shape its needs. For example, isolating social animals like horses is as horrible as solitary confinement for humans. The more aware we are of animals' needs and the different forms the consciousness of animals takes, the better our welfare practices will become.



The Jains say that Neminath (one of the “saviors” of Jainism) left his village on his wedding day riding an elephant. On the way, he heard agonized screams and demanded to know their origin. Neminath’s elephant driver told him that the screams came from animals that were being slaughtered for his wedding feast. This moment transformed Neminath. He freed the surviving animals and renounced his former life.

Animal-human chimeras are produced by transferring human stem cells into animal hosts at various stages of development. The main motivation is to produce animals with organs made of human cells that can be transplanted into people. But if human cells populate the nervous system of these chimeras, what forms will the cognition and consciousness of these chimeras take? Will their cognition and consciousness be enhanced? Will they be diminished, or become pathological? We have no idea. We cannot care for these chimeras' welfare if we know nothing about their mental needs, capacities, and handicaps.

Should we worry about the welfare of hypothetical future robots? The philosopher Thomas Metzinger argued that constructing conscious robots (or any other conscious artificial being) that have a sense of self and who "own" their feelings will lead to suffering both because they will be dominated by humans and because the early versions are likely to be only partially functional and therefore handicapped. Ian McEwan's dystopic book *Machines Like Me* describes the effects of an inflexible value system in the first fifty sentient robots. They could not compromise and could not avoid the suffering engendered by the clash between their values and the behavior of their human owners. This conflict eventually led to the tragic death of all of them: forty-nine suicides and one murder. Should we prohibit such a research program, or hope, with McEwan, for a future human-robot utopia?

We do not know what it will be like to be a robot, a chimera, or a cyborg in an augmented collective cyborg reality. Will their consciousness transcend the symbol-based consciousness of humans? Can we imagine postsymbolic consciousness—or a consciousness that does not depend on life as we know it?

5.12 SOLARIS: THE LIMITS

In one of the most intriguing science fiction books ever written, Stanisław Lem described a planet, Solaris, covered with a living and conscious gelatinous ocean. Although the human concepts of life and consciousness cannot convey the meaning of the ocean's profound, mysterious existence, Lem manages to describe the immense complexity and beauty of the gigantic, ever-changing structures constructed by the ocean, which stabilize the planet's orbit as it revolves around blue and red suns and creates inert and living neutrino-beings that reproduce the products, memories, and fantasies of the people who come to study it. Understanding the processes that construct the fantastic, metamorphosing structures that make up the ocean is impossible for humans:

We observe a fraction of the process, like hearing a vibration of a single string in an orchestra of supergiants. We know, but cannot grasp, that above and below, beyond the limits of perception or imagination, thousands and millions of simultaneous transformations are at work, interlinked like a musical score by mathematical counterpoint. It has been described as a symphony in geometry, but we lack the ears to hear it.

The values and meanings of the ocean's activities cannot be translated into human concepts. The knowledge of the ocean is incommunicable. Even though humans cannot communicate with the ocean, the ocean explores, probes, dissects, and exposes humans, although one cannot fathom why—indeed, the very notion of “why” may be misleading. The ocean's explorations provide humans with the chance to contact hidden aspects of themselves, to re-create their relationships and acknowledge both the barriers to their understanding and their boundless longing. The protagonist of the story decides to

remain on Solaris, which recycled and transformed the tragedy of his love, holding on to “the faith that the time of cruel miracles was not past.”

Lem’s book helps us recognize the limits, possibilities, and transformations of our own minds. How will the study of consciousness change during the next five hundred years as scientists, novelists, and artists reimagine consciousness? Which questions will be answered? Which questions will be abandoned, and which will be reframed? What new questions will arise?



“It has been described as a symphony in geometry, but we lack the ears to hear it”
(Stanisław Lem).



Theories pass. The Frog remains.

—Jean Rostand

ACKNOWLEDGMENTS

We are grateful to the Open University of Israel and Tel-Aviv University for supporting and facilitating our work for many years. We thank our friends and colleagues, Uzi Ben-Zvi, Daniel Dor, Romi Englander, Lea Idinopulos, Daniela Leibowitz, Efrat Rapoport, Dotan Reis, Sara Schwartz, Dor Shilton, and Oryan Zacks, who critically read the texts, made us aware of the pitfalls of jargon, and encouraged us to make greater efforts to write simply but without simplification. We are also grateful to Yafeng Shan, Machi Okoyama, and Beatrice, Leo, and Mati Senkman, who translated short fragments of texts or poems from Chinese, Japanese and Spanish. Special thanks are due to Marion Lamb, who read every word several times, edited the manuscript, pointed to ambiguities, and made us sharpen and crystallize our thoughts. Finally, we are very grateful to Anna's friend, the East Anglian poet Jane Monet, who permitted us to use some of her unpublished poems that were inspired by Anna's pictures.

NOTES

These notes provide the sources of the quotations and information in each text. Many of the discussions are based on our book *The Evolution of the Sensitive Soul: Learning and the Origins of Consciousness* (Cambridge, MA: MIT Press, 2019), which we abbreviate as TESS.

The quotation in the introduction is from T. H. Huxley, *Lessons in Elementary Physiology* (London: Macmillan, 1866), 193.

VISTA 1

The introductory quotation comes from the published version of a BBC radio talk given by Sir Charles Scott Sherrington (1857–1952), reproduced in *The Physical Basis of Mind: A Series of Broadcast Talks*, ed. P. Laslett (Oxford: Blackwell, 1950), 3–4.

1.1 WHAT IS THE MIND?

The poem is by Ikkyū Sōjun (休宗純, 1394–1481). The translation is modified from that of R. H. Blyth, in *Zen and Zen Classics*, vol. 5 (Tokyo: Hokuseido Press), 162–192, <https://terebess.hu/zen/mesterek/IkkyuDoka.html#d>.

1.2 STREAMS, WAVES, BIRDS

The first quotation is from W. James, “On Some Omissions of Introspective Psychology,” *Mind* 9 (1884): 16. The second and third quotations are from W. James, *The Principles of Psychology*, vol. 1 (New York: Dover, 1890), 280, 243.

1.3 THE BUTTERFLY

The quotation is from B. Watson, *Zhuangzi: Basic Writings*, 3rd ed. (New York: Columbia University Press, 2003), 43. Dr. Yafeng Shan translated the Chinese text from the original.

1.4 DUALISM

The quotation is from R. Descartes, *Meditations on First Philosophy in which are demonstrated the existence of God and the distinction between the human soul and the body* (1639), trans. J. Cottingham (Cambridge: Cambridge University Press, 1986). (Taken from the second paragraph of the Second Meditation: <https://www.marxists.org/reference/archive/descartes/1639/meditations.htm>.)

1.5 PANPSYCHISM

Panpsychism is discussed in P. Goff, *Galileo's Error: Foundations for a New Science of Consciousness* (New York: Pantheon Books, 2019).

1.6 PHYSICALISM

The quotation is from F. Crick, *The Astonishing Hypothesis* (New York: Charles Scribner's, 1994), 3. Crick provides a clear exposition of physicalism.

1.7 ARISTOTELIAN NATURALISM

The quotation is from Aristotle, *On the Soul*, book 2, chapter 3, in *The Complete Works of Aristotle*, Revised Oxford Translation, vol. 1 (Princeton, NJ: Princeton University Press, 1984).

1.8 WHAT IS IT LIKE TO BE A BAT?

The thought experiment is described in T. Nagel, "What Is It Like to Be a Bat?" *Philosophical Review* 83 (1974): 435–450. For Daniel Kish's first quotation, see Clare Wilson, "This Is How Some Blind People Are Able to Echolocate Like Bats," *New Scientist*, August 31, 2017, <https://www.newscientist.com/article/2145962-this-is-how-some-blind-people-are-able-to-echolocate-like-bats/>; for the second quotation, see "Navigating the World Using Echoes," RNIB, November 1, 2017, <https://www.rnib.org.uk/rnibconnect/blind-echolocation>.

1.9 WAYS OF KNOWING

The thought experiment known as the "knowledge argument" is described in F. Jackson, "Epiphenomenal Qualia," *Philosophical Quarterly* 32 (1982): 127–136. The quotation is from 130.

1.10 SELF: THE HOLE IN THE DOUGHNUT

The first quotation is taken from D. Hume, *A Treatise of Human Nature* (1739–1740/1978), 2nd ed., ed. P. H. Nidditch (Oxford: Clarendon Press, 1978), 252. The second quotation is taken from O. Mandelstam, *Fourth Prose* (1930), published in the *Hudson Review* 23, no. 1 (spring 1970): 65.

Merker's view of the self is described in B. Merker, "Consciousness without a Cerebral Cortex: A Challenge for Neuroscience and Medicine," *Behavioral and Brain Sciences* 30 (2007): 63–134, <https://doi.org/10.1017/S0140525X07000891>.

1.11 ARE YOU A BRAIN IN A VAT?

A discussion of the brain-in-the-vat thought experiment can be found in M. McKinsey, "Skepticism and Content Externalism," in *The Stanford Encyclopedia of Philosophy* (Summer 2018 Edition), ed. Edward N. Zalta, <https://plato.stanford.edu/entries/skepticism-content-externalism/>. The life of an isolated brain is described in S. Reardon, "Pig Brains Kept Alive Outside Body for Hours after Death," *Nature* 568 (2019): 283–284, <https://doi.org/10.1038/d41586-019-01216-4>.

1.12 PHILOSOPHICAL ZOMBIES

The zombie argument is presented in D. J. Chalmers, *The Conscious Mind: In Search of a Fundamental Theory* (New York: Oxford University Press, 1996).

VISTA 2

The introductory quotation is from Aristotle, *Parts of Animals*, trans. A. L. Peck, Loeb Classical Library 323 (Cambridge, MA: Harvard University Press, 1937), https://www.loebclassics.com/view/aristotle-parts_animals/1937/pb_LCL323.101.xml, 645a, 20–25.

2.1 A VIEW FROM BIOLOGY

The story is based on a real event, but the details have been changed.

2.2 ARE BACTERIA CONSCIOUS?

The first quotation is from W. Wundt, *Principles of Physiological Psychology I*, trans. E. B. Titchener (London: Sonnenschein, 1904), 31. The second quotation is from A. S. Reber, *The First Minds: Caterpillars, 'Karyotes, and Consciousness* (Oxford: Oxford University Press, 2018), ix. Bacterial cognition is discussed in P. Lyon, "The Cognitive Cell: Bacterial Behavior Reconsidered," *Frontiers in*

Microbiology 6 (2015): 264, <https://doi.org/10.3389/fmicb.2015.00264>, and D. Bray, *Wetware: A Computer in Every Living Cell* (New Haven, CT: Yale University Press, 2009). The quotation in the image caption is from Bray, 26.

2.3 CONSCIOUS SLIME MOLDS?

More about slime mold behavior can be found in K. Moskvitch, “Slime Molds Remember—But Do They Learn?” *Quanta*, July 9, 2019, <https://www.quanta magazine.org/slime-molds-remember-but-do-they-learn-20180709/>, and in P. Ball, “Cellular Memory Hints at the Origins of Intelligence,” *Nature* 451 (2008): 385, <https://doi.org/10.1038/451385a>. The quotations in the text are from Ball, 385.

2.4 ARE PLANTS CONSCIOUS?

For a book-length description of the sensory capacities of plants and their many amazing adaptations, see D. Chamovitz, *What a Plant Knows: A Field Guide to the Senses* (New York: Farrar, Straus and Giroux, 2017). The quotation in the text is from R. Powers, *The Overstory* (New York: Vintage, 2019), 566–567, and the quotation in the image caption is from A. G. Parise, M. Gagliano, and G. M. Souza, “Extended Cognition in Plants: Is It Possible?” *Plant Signaling and Behavior* 15, no. 1710661 (2020): 4, <https://doi.org/10.1080/15592324.2019.1710661>.

2.5 ARE SPONGES CONSCIOUS?

The sponge-origin of all animals is discussed in TESS, chapter 6, 269–273.

2.6 WHAT’S SO SPECIAL ABOUT A NERVOUS SYSTEM?

For more details about the nervous system see TESS, chapter 6, 254–262. An action potential is a brief change in the voltage across a membrane due to the flow of certain ions into and out of the neuron. It is the mode through which a neuron transports electrical signals.

In addition to intracellular and synaptic memory, global electrical fields within the nervous system contribute another layer of memory.

2.7 ENCHANTING MEDUSAE, GRACEFUL SEA ANEMONES

For more information about jellyfish and sea anemones see TESS, chapter 6, 273–289. For an overview of the life of a well-studied jellyfish, see D. J. Albert, “What’s on the Mind of a Jellyfish? A Review of Behavioural Observations on

Aurelia sp. Jellyfish,” *Neuroscience and Biobehavioral Reviews* 35 (2011): 474–482, <https://doi.org/10.1016/j.neubiorev.2010.06.001>.

2.8 BRAINY WORMS

The learning capacities of nematodes are described in C. H. Rankin, “Invertebrate Learning: What Can’t a Worm Learn?” *Current Biology* 14 (2004): R617–R618, <https://doi.org/10.1016/j.cub.2004.07.044>. Learning in flatworms is described in T. Inoue, H. Hoshino, T. Yamashita, S. Shimoyama, and K. Agata, “Planarian Shows Decision-Making Behavior in Response to Multiple Stimuli by Integrative Brain Function,” *Zoological Letters* 1 (2015): 7, <https://doi.org/10.1186/s40851-014-0010-z>. The quotation in the image caption is taken from Lyndsay Brownsell, “Mike Levin on Electrifying Insights into How Bodies Form: Using Bioelectricity to Study How Cells Make Collective Decisions about Growth and Shape,” Wyss Institute, July 26, 2019, <https://wyss.harvard.edu/news/mike-levin-on-electrifying-insights-into-how-bodies-form/>.

2.9 A FUZZY CONSENSUS

The quotation is from *The Cambridge Declaration* (2012), <http://fcmconference.org/img/CambridgeDeclarationOnConsciousness.pdf>. The declaration was written and signed by a group of neuroscientists who attended the “Consciousness in Human and Non-Human Animals” conference at the University of Cambridge in 2012.

2.10 ARE FISH CONSCIOUS?

The behavioral adaptations and cognitive capacities of fish are described in J. Balcombe, *What a Fish Knows: The Inner Lives of Our Underwater Cousins* (New York: Scientific American/Farrar, Straus and Giroux, 2016).

The quotation is from C. Darwin, *The Descent of Man, and Selection in Relation to Sex* (London: John Murray, 1871), 402.

2.11 ARE REPTILES CONSCIOUS?

The view that consciousness began in reptiles is presented in M. Cabanac, A. J. Cabanac, and A. Parent, “The Emergence of Consciousness in Phylogeny,” *Behavioural Brain Research* 198 (2009): 267–272, <https://doi.org/10.1016/j.bbr.2008.11.028>. Details of some of the examples presented in the text can be found in G. M. Burghardt, B. Ward, and R. Rosscoe, “Problem of Reptile Play: Environmental Enrichment and Play Behavior in a Captive Nile Soft-Shelled

Turtle, *Trionyx triunguis*," *Zoo Biology* 15, no. 3 (1996): 223–238. [https://doi.org/10.1002/\(SICI\)1098-2361\(1996\)15:3<223::AID-ZOO3>3.0.CO;2-D](https://doi.org/10.1002/(SICI)1098-2361(1996)15:3<223::AID-ZOO3>3.0.CO;2-D); M. Shein-Idelson, J. M. Ondracek, H. P. Liaw, S. Reiter, and G. Laurent, "Slow Waves, Sharp Waves, Ripples, and REM in Sleeping Dragons," *Science* 352, no. 6285 (2016): 590–595, <https://doi.org/10.1126/science.aaf3621>; B. A. Ellis-Quinn and C. A. Simon, "Lizard Homing Behavior: The Role of the Parietal Eye during Displacement and Radio-Tracking, and Time-Compensated Celestial Orientation in the Lizard, *Sceloporus jarrovi*," *Behavioral Ecology and Sociobiology* 28 (1991): 397–407.

2.12 VERTEBRATES: BIRDS AND MAMMALS

For a discussion of consciousness in birds and mammals, see TESS, chapter 5, 205–209. For information about pig intelligence, see L. Marino and C. M. Colvin, "Thinking Pigs: A Comparative Review of Cognition, Emotion, and Personality in *Sus domesticus*," *International Journal of Comparative Psychology* 28 (2015), <https://escholarship.org/uc/item/8sx4s79c>. Alex the African parrot is described in I. Pepperberg, *Alex & Me: How a Scientist and a Parrot Discovered a Hidden World of Animal Intelligence—and Formed a Deep Bond in the Process* (London: Scribe, 2009).

2.13 SOME ARTHROPODS?

The poem was composed by Kobayashi Issa (小林一茶, 1763–1828), a Japanese poet and lay Buddhist priest known for his haiku poems and journals.

The possibility of consciousness in arthropods is discussed in A. B. Barron and C. Klein, "What Insects Can Tell Us about the Origins of Consciousness," *Proceedings of the National Academy of Sciences USA* 113 (2016): 4900–4908, <https://doi.org/10.1073/pnas.1520084113>. The functional similarity between vertebrate and arthropod brains is discussed in N. J. Strausfeld and F. Hirth, "Deep Homology of Arthropod Central Complex and Vertebrate Basal Ganglia," *Science* 340, 157–161, <https://doi.org/10.1126/science.1231828>. Anxiety behavior in crayfish is described in P. Fossat, J. Bacque-Cazenave, P. De Deurwaerdere, J.-P. Delbecque, and D. Cattaert, "Anxiety-like Behavior in Crayfish Is Controlled by Serotonin," *Science* 344, no. 6189 (2014): 1293–1297, <https://doi.org/10.1126/science.1248811>. The cognition of bees is described in L. Chittka and C. Wilson, "Bee-Brained," *Aeon*, November 27, 2018, <https://aeon.co/essays/inside-the-mind-of-a-bee-is-a-hive-of-sensory-activity>.

The quotation is from Maurice Maeterlinck (1862–1949), *The Life of the Bee*, trans. A. Sutro (Reprint of the 1901 edition; New York: Dodd, Mead, 1902), 45.

2.14 WHAT ABOUT CEPHALOPODS?

The poem by Ogden Nash (1902–1971) is taken from http://famouspoetsandpoems.com/poets/ogden_nash/poems/19643.

The second poem, by Matsuo Bashō (松尾 芭蕉, 1644–1694), is taken from <http://firstknownwhenlost.blogspot.com/2014/08/a-dream-beneath-summer-moon.html>.

For a discussion of the cognition and consciousness of the octopus, see P. Godfrey-Smith, *Other Minds: The Octopus, the Sea, and the Deep Origins of Consciousness* (New York: Farrar, Straus and Giroux, 2016).

2.15 ARE ONLY HUMANS CONSCIOUS?

The quotation is taken from an essay about the seventeenth-century priest and philosopher Malebranche: N. Jolley, “Malebranche on the Soul,” in *The Cambridge Companion to Malebranche*, ed. S. Nadler (Cambridge: Cambridge University Press, 2000), 31–58. The view that only humans are conscious is discussed in TESS, chapter 5, 199–205.

2.16 WHAT ARE THE SOCIAL DIMENSIONS OF CONSCIOUSNESS?

For a discussion of cultural traditions and norms in animals, see E. Avital and E. Jablonka, *Animal Traditions: Behavioural Inheritance in Evolution* (Cambridge: Cambridge University Press, 2000); C. Andrews, *The Animal Mind: An Introduction to the Philosophy of Animal Cognition*, 2nd ed. (New York: Routledge, 2020), chapter 9.

The quotation is from C. G. Jung, *The Concept of the Collective Unconscious* (Princeton, NJ: Princeton University Press, 1968), 42. The essay was originally given as a lecture to the Abernethian Society at St. Bartholomew’s Hospital, London, on October 19, 1936, and published in *St. Bartholomew’s Hospital Journal* 44 (1936–37): 46–49, 64–66.

VISTA 3

The introductory quotation is from H. Spencer, *Principles of Psychology*, 3rd ed. (London: William and Norgate, 1890), 291.

3.1 INTRODUCTION: WHY TAKE AN EVOLUTIONARY APPROACH TO CONSCIOUSNESS?

The topic is discussed in depth in TESS, chapters 1–3.

3.2 EVOLUTIONARY THEORY

The quotation is taken from the last paragraph of C. Darwin, *The Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*, 6th ed. (London: John Murray, 1872), <https://www.gutenberg.org/files/2009/2009-h/2009-h.htm>.

For a formulation of the principles of Darwin's theory of natural selection by Maynard Smith, see J. Maynard Smith, *The Problems of Biology* (Oxford: Oxford University Press, 1986). Note that the fourth principle, competition, actually refers to differential survival and reproduction, which does not necessarily entail competition (a sociologically interesting metaphorical bias). For a discussion of the twenty-first-century integrated view of evolutionary theory, see E. Jablonka and M. J. Lamb, *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral and Symbolic Variation in the History of Life*, 2nd ed. (Cambridge, MA: MIT Press, 2014). The phrase "genes are followers, not leaders in evolution" comes from M. J. West-Eberhard, *Developmental Plasticity and Evolution* (New York: Oxford University Press, 2003), 20.

3.3 EVOLUTIONARY TRANSITIONS

Our approach is described in TESS, 241–250. For views on evolutionary transitions, see J. Maynard Smith and E. Szathmáry, *The Major Transitions in Evolution* (Oxford: Oxford University Press, 1995); D. C. Dennett, *Kinds of Minds: Towards an Understanding of Consciousness* (New York: Basic Books, 1997).

3.4 AN EVOLUTIONARY TRANSITION APPROACH TO CONSCIOUSNESS

Our approach is described in TESS, chapters 1, 5, and 8.

3.5 THE TRANSITION TO LIFE

Gánti's ideas are presented and discussed in T. Gánti, *The Principles of Life, with a Commentary by James Griesemer and Eörs Szathmáry* (New York: Oxford University Press, 2003).

3.6 THE TRANSITION TO CONSCIOUSNESS: COMPILING A LIST

The list of consciousness capacities and a discussion of each entry in the list is given in TESS chapter 5, 233–239. The first quotation is from W. James, *The Principles of Psychology*, vol. 1 (New York: Dover, 1890), 609–610; the second is from M. Merleau-Ponty, *Phenomenology of Perception*, trans. C. Smith (London: Routledge and Kegan Paul, 1962), 169.

3.7 LEARNING AND CONSCIOUSNESS: AN UNLIKELY RELATION?

The quotations are from B. F. Skinner, *Science and Human Behavior* (New York: Macmillan, 1953), 30 and 3. For a history of the research on associative learning, see R. Boakes, *From Darwin to Behaviourism: Psychology and the Minds of Animals* (Cambridge: Cambridge University Press, 1984).

3.8 UNLIMITED ASSOCIATIVE LEARNING (UAL) IS THE EVOLUTIONARY TRANSITION MARKER OF CONSCIOUSNESS

For detailed discussions of UAL, see TESS, chapter 5, 230–239, and chapter 8, 382–403. For a primer on the UAL theory, see J. Birch, S. Ginsburg, and E. Jablonka, “Unlimited Associative Learning and the Origins of Consciousness: A Primer and Some Predictions,” *Biology & Philosophy* 35 (2020), article 56, <https://doi.org/10.1007/s10539-020-09772-0>.

3.9 FEELINGS AND CONCEPTIONS

For Darwin’s view of emotions and their evolution, see C. Darwin, *The Expression of the Emotions in Man and Animals* (London: John Murray, 1872). Current views about emotions and feelings are summarized in TESS, chapter 5, 213–219. For descriptions and discussions of the emotional lives of animals, see T. Grandin, *Animals in Translation: Using the Mysteries of Autism to Decode Animal Behavior* (New York: Scribner, 2005).

3.10 THE FUNCTIONS AND GOALS OF CONSCIOUSNESS

The quotation is from W. James, *The Principles of Psychology*, vol. 1 (New York: Dover Publications, 1890), 141. For a view of consciousness as a mode of being, see S. Ginsburg and E. Jablonka, “Consciousness as a Mode of Being,” *Journal of Consciousness Studies* 27, nos. 9–10 (2020): 148–162.

3.11 WHICH ORGANISMS MANIFEST UAL? WHO IS CONSCIOUS?

Chapter 8 of TESS discusses the neural architecture of UAL and the behavioral and neural evidence for its existence in different groups of organisms.

3.12 THE BIG BANG OF ANIMAL EVOLUTION

The idea that consciousness evolved during the Cambrian period and was one of the drivers of the explosive evolutionary diversification that defines it is described in TESS, chapter 9, 405–425. The Cambrian origins of consciousness are also discussed by T. E. Feinberg and J. Mallatt, *The Ancient Origins of*

Consciousness: How the Brain Created Experience (Cambridge, MA: MIT Press, 2016).

3.13 THE ORIGINS OF SUFFERING

The opening quotation is from Ecclesiastes 1:18, King James Version. The relationship between learning, stress, and mental suffering is discussed in TESS, chapter 9, 426–439. The smoke detector principle is described in R. M. Nesse, “The Smoke Detector Principle,” *Annals of the New York Academy of Science* 935 (2001): 75–85, <https://doi.org/10.1111/j.1749-6632.2001.tb03472.x>.

The quotation is from the last paragraph of C. Darwin, *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life* (London: John Murray, 1859).

3.14 THE EVOLUTION OF IMAGINATION

The quotation is from S. T. Geisel (Dr. Seuss), *Oh, the Thinks You Can Think!* (1975; New York: Random House, 2003). The investigations of jays’ learning are described in N. S. Clayton, T. J. Bussey, and A. Dickinson, “Can Animals Recall the Past and Plan for the Future?” *Nature Reviews Neuroscience* 4 (2003): 685–691, <https://doi.org/10.1038/nrn1180>.

VISTA 4

The introductory quotations are from Psalm 8:4–9, King James Version, and from R. Cummings Neville, *Boston Confucianism: Portable Tradition in the Late-Modern World* (Albany: SUNY Press, 2000). We thank Dr. Yafeng Shan for translating the Chinese text.

4.1 HUMAN NATURE

Aristotle claimed that “man is by nature a political animal” in *Politics* 1235a. Hume’s quotation is taken from *A Treatise of Human Nature*, section VI, 618, <https://davidhume.org/texts/t/3/3/6>.

The quotation is from E. Jablonka and M. J. Lamb, *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*, 2nd ed. (Cambridge, MA: MIT Press, 2014), 189. In the figure caption, “blue brain” refers to the Blue Brain Project, a Swiss brain research initiative that aims to create a digital reconstruction of rodent and eventually human brains

by reverse-engineering mammalian brain circuitry. See <https://www.epfl.ch/research/domains/bluebrain/>.

4.2 WHAT SETS HUMANS APART?

The quotation is from Kurt Vonnegut *The Cat's Cradle* (New York: Holt, Rinehart & Winston, 1963), chapter 81. Darwin's view of human evolution is presented in C. Darwin, *The Descent of Man, and Selection in Relation to Sex* (London: John Murray, 1871).

4.3 WHY TAKE AN EVOLUTIONARY APPROACH TO HUMAN NATURE?

The African branch of *Homo erectus*, our more direct ancestor, is today called *Homo ergaster*. The name *Homo erectus* is now reserved for the Asian branch. Pre-*sapiens* people belonging to the genus *Homo* are referred to as “archaic humans.” The picture was inspired by Dian Fossey (1932–1985), a US zoologist who studied Gorilla groups in Rwanda for over eighteen years. She was murdered in her cabin, probably by a poacher.

4.4 BEFORE *SAPIENS*

There are many books about our ancestors; the following are directly relevant to our arguments: M. Donald, *The Origins of the Modern Human Mind* (Cambridge, MA: Harvard University Press, 1991); S. B. Hrdy, *Mothers and Others: The Evolutionary Origins of Mutual Understanding* (Cambridge, MA: Harvard University Press, 2009); M. Tomasello, *A Natural History of Human Thinking* (Cambridge, MA: Harvard University Press, 2014).

The quotation is from “Genetic Logic and Sociology” (1928), in J. Piaget, *Sociological Studies*, ed. L. Smith (New York: Routledge, 1995), 200.

4.5 THE BLUSHING *ERECTUS*

The first quotation is attributed to Menicus (4th c. BCE) and is cited in H. Hargaden, “Shame,” in *The Art of Relational Supervision*, ed. H. Hargaden (London and New York: Routledge), 129.

For discussions of social emotions and the blush, see C. Darwin, *The Expression of the Emotions in Man and Animals* (London: John Murray, 1872); R. W. Crozier, *Blushing and the Social Emotions: The Self Unmasked* (London: Palgrave Macmillan, 2006).

The second quotation is from Genesis 3:6, King James Version. The poem in the picture legend is taken from *The Letters of John Keats 1814–1818*,

vol. 1, ed. H. E. Rollins (Cambridge, MA: Harvard University Press, 1985), 219.

4.6 THE EXTENDED MIND: HANDS, TOOLS, MEMORIALS

The opening quotation by Aristotle is from *On the Soul*, book 3, chapter 8, in *The Complete Works of Aristotle*, Revised Oxford Translation, vol. 1 (Princeton, NJ: Princeton University Press, 1984). The classic book on the hand and its evolution is J. Napier, *Hands*, rev. ed. (Princeton: Princeton University Press, 1993). The change in the representation of the body as a result of tool use is described in M. Martel, L. Cardinali, A. C. Roy, and A. Farnè, “Tool-Use: An Open Window into Body Representation and Its Plasticity,” *Cognitive Neuropsychology* 33 (2016): 82–101, <https://doi.org/10.1080/02643294.2016.1167678>. *Yad Vashem* is the Jerusalem Institute that commemorates the Holocaust. In Hebrew, *Yad* means both “hand” and “memorial”; *Shem* means “name” and is also one of the ways to refer to God (*Hashem*). For the quotation attributed to Kant, see S. A. J. Stuart, “Privileging Exploratory Hands: Prehension, Apprehension, Comprehension,” in *The Hand, an Organ of the Mind: What the Manual Tells the Mental*, ed. Z. Radman (Cambridge, MA: MIT Press, 2011), 329–347. Theories about the evolution of toolmaking, the hand, and the brain are described in S. Almécija and C. C. Sherwood, “Hands, Brains, and Precision Grips: Origins of Tool Use Behaviors,” in *Evolution of Nervous Systems*, 2nd ed., ed. J. H. Kaas (Oxford: Academic Press, 2017), 299–315, <https://doi.org/10.1016/B978-0-12-804042-3.00085-3>; D. Stout and T. Chaminade, “Stone Tools, Language and the Brain in Human Evolution,” *Philosophical Transactions of the Royal Society B* 367, no. 1585 (2012): 75–87, <https://doi.org/10.1098/rstb.2011.0099/>.

4.7 THE SYMBOLIC DIMENSION

Our sources here include E. Cassirer, *An Essay on Man: An Introduction to a Philosophy of Human Culture* (New Haven, CT: Yale University Press, 1944); the first quotation is taken from this book, 24; T. W. Deacon, *The Symbolic Species: The Co-evolution of Language and the Brain* (New York: Norton, 1997); C. Heyes, *Cognitive Gadgets: The Cultural Evolution of Thinking* (Cambridge, MA: Belknap Press of Harvard University Press, 2018); E. Jablonka and M. J. Lamb, *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*, 2nd ed. (Cambridge, MA: MIT Press, 2014), chapters 6 and 8.

The second quotation is from L. Wittgenstein, *Tractatus Logico-Philosophicus* (London: Kegan Paul, 1922), 5.6, 76, <http://writing.upenn.edu/library/Wittgenstein-Tractatus.pdf>.

4.8 LANGUAGE AND IMAGINATION

The Shakespeare quotation is from *The Tempest* (1611), act 4, lines 156–157.

We refer to the myths in Genesis 1 (creating the world with words); Genesis 3 (original sin); Genesis 11 (the Tower of Babel). The quotation is from John 1:1.

Dor's view of language is described in D. Dor, *The Instruction of Imagination: Language as a Social Communication Technology* (Oxford: Oxford University Press, 2015); the quotation is from page 25.

4.9 THOUGHTS AND FEELINGS

The quotation is from M. Tomasello, *A Natural History of Human Thinking* (Cambridge, MA: Harvard University Press, 2014), 9. The evolutionary relationship between emotions and language is described in E. Jablonka, S. Ginsburg, and D. Dor, "The Co-evolution of Language and Emotions," *Philosophical Transactions of the Royal Society B* 367 (2012): 2152–2159, <https://doi.org/10.1098/rstb.2012.0117>.

4.10 ON SOME PECULIARITIES OF THE SYMBOLIC SPECIES

Some of the peculiarities of humans that are attributable to the use of language are described in TESS, chapter 10, 472–479; D. Dor, "The Role of the Lie in the Evolution of Human Language," *Language Sciences* 63 (2017): 44–59, <https://doi.org/10.1016/j.langsci.2017.01.001>; M. Tomasello, *A Natural History of Human Thinking* (Cambridge, MA: Harvard University Press, 2014). The quotation is from page 81 in Tomasello.

4.11 THE SYMBOLIC EXPLOSION

Vernadsky's view of the "sphere of reason" is described in V. Vernadsky, *Scientific Thought as a Planetary Phenomenon*, which is one of his last works, written between 1931 and 1944. The most relevant chapter can be found in https://21sci-tech.com/Subscriptions/Spring-Summer-2012_ONLINE/TCS_Sp-Su_2012.pdf (Vernadsky's contribution is on 17–31).

The quotation is from R. Carson, *Silent Spring* (1962; New York: Mariner Books, 2002), 2–3.

For a concise discussion of the Anthropocene, see E. C. Ellis, *Anthropocene: A Very Short Introduction* (Oxford: Oxford University Press, 2018).

4.12 A SPLIT SOUL?

For some of the different conceptions of the “split soul,” see A. Koestler, *The Ghost in the Machine* (London: Hutchinson, 1967); the quotation is from page xi; S. Freud, *Civilization and Its Discontents*, trans. J. Strachey (New York: Norton, 1930); the quotation is from 35; K. Lorenz, *On Aggression* (New York: Harcourt, Brace & World, 1966). See also I. McGilchrist, *The Master and His Emissary: The Divided Brain and the Making of the Western World* (New Haven, CT: Yale University Press, 2009). In the image caption, the last part of Borges’s poem was translated from the Spanish original by Beatrice, Leo, and Mati Senkman.

VISTA 5

The introductory poem, written in 1886, is by Emily Dickinson (1830–1886), taken from *The Complete Poems of Emily Dickinson*, ed. T. E. Johnson (Boston: Little, Brown, 1960), poem 652.

5.1 STRETCHING THE LIMITS

The quotation is from U. K. Le Guin, *A Wizard of Earthsea* (New York: Houghton Mifflin Harcourt, 2012), 92.

5.2 SAVANT MINDS

The term “neurally atypical” is now used in preference to the term “autistic.” Here we use “autistic” when this is how the person is described in the literature. For general discussions of savants, see D. A. Treffert, *Islands of Genius* (London: Jessica Kingsley, 2012); the quotation is from page 18. The second quotation is from N. Higashida, *The Reason I Jump* (New York: Random House, 2013), Q32, 59.

5.3 PERCEPTUAL MEMORY AND THE ART OF FORGETTING

The first quotation is from A. R. Luria, *A Little Book about a Vast Memory: The Mind of a Mnemonist* (London: Penguin Books, 1975), 64.

The second quotation comes from Naftali Tishbi, whose ideas are summarized in N. Wolchover, “New Theory Cracks Open the Black Box of Deep

Learning," *Quanta Magazine*, September 21, 2017, <https://www.quantamagazine.org/new-theory-cracks-open-the-black-box-of-deep-learning-20170921/>.

The quotation in the image caption is taken from J. L. Borges, "Funes, the Memorious," in *Ficciones*, trans. A. Kerrigan (New York: Grove Press, 1962), 115.

5.4 CONSCIOUSNESS UPHEAVALS: PSYCHEDELIC DRUGS

The quotations are from B. Shanon, *The Antipodes of the Mind: Charting the Phenomenology of the Ayahuasca Experience* (Oxford: Oxford University Press, 2002), 190 and 18. For an overview of psychedelics, see M. Pollen, *How to Change Your Mind: The New Science of Psychedelics* (London: Allen Lane, 2018).

5.5 MENTAL PAINS: IMAGES OF THE UNCONSCIOUS

The information about Nise da Silveira is taken from V. Pordeus, "Nise da Silveira: Brazilian Pioneer in Art and Transcultural Psychiatry," *Academia*, accessed April 14, 2021, https://www.academia.edu/34364588/Nise_da_Silveira_Brazilian_Pioneer_in_Art_and_Transcultural_Psychiatry.

A discussion of the patterns described by Klüver can be found in J. Ouellette, "A Math Theory for Why People Hallucinate," *Quanta Magazine*, July 30, 2018, <https://www.quantamagazine.org/a-math-theory-for-why-people-hallucinate-20180730/>.

5.6 "HIGHER" FORMS OF CONSCIOUSNESS?

The first quotation, from *Yoga Vasishta Maharamayana* (1891–1899), is taken from W. James, *The Varieties of Religious Experience: A Study in Human Nature* (New York: Longmans, Green, 1917), 392. The second quotation is from A. Huxley, *The Doors of Perception and Heaven and Hell* (New York: Harper, 1954), 4–5 of the version at <https://maps.org/images/pdf/books/HuxleyA1954TheDoorsOfPerception.pdf>.

5.7 HOW DID OUR ANCESTORS PERCEIVE THE WORLD?

Temple Grandin's ideas are described in T. Grandin, *Animals in Translation: Using the Mysteries of Autism to Decode Animal Behavior* (New York: Scribner, 2005). The quotation is from A. Snyder, "Explaining and Inducing Savant Skills: Privileged Access to Lower Level, Less-Processed Information," *Philosophical Transactions of the Royal Society B* 364 (2009): 1399–1405, at 1399, <https://doi.org/10.1098/rstb.2008.0290>. N. Humphrey discusses the artistic perception

of our ancestors on the basis of savant art in “Cave Art, Autism, and the Evolution of the Human Mind,” *Cambridge Archeological Journal* 8, no. 2 (1998), 165–191, <https://doi.org/10.1017/S0959774300001827>.

5.8 TELLING ARTIFACTS

The quotation was on Feynman’s blackboard at the time of his death in 1988. It is taken from Michael Way, “Editorial: ‘What I Cannot Create, I Do Not Understand,’” *Journal of Cell Science* 130 (2017): 2941–2942, <https://doi.org/10.1242/jcs.209791>. The quotation from Aristotle is from *Movement of Animals* 701a 1–4, in *The Complete Works of Aristotle*, Revised Oxford Translation, vol. 1 (Princeton, NJ: Princeton University Press, 1984). Karel Čapek’s play *R.U.R.* was first performed in 1920. It can be read online at <https://www.gutenberg.org/files/59112/59112-h/59112-h.htm>. The information on Golems is based on M. Idel, *Golem: Jewish Magical and Mystical Traditions on the Artificial Anthropoid* (Albany: State University of New York Press, 1990).

5.9 CONSCIOUS ROBOTS?

AlphaGo, a program that uses deep learning algorithms, is described in M. Shanahan, “Conscious Exotica,” *Aeon*, October 19, 2016, <https://aeon.co/essays/beyond-humans-what-other-kinds-of-minds-might-be-out-there>. The citation by the admiring Go champion is from this article. For a discussion of the possibility of robot consciousness, see S. Dehaene, H. Lau, and S. Koudier, “What Is Consciousness, and Could Machines Have It?” *Science* 358 (2017): 486–492, <https://doi.org/10.1126/science.aan8871>.

5.10 VIRTUAL AND CYBORG REALITIES

For the phrase attributed to Lincoln, see D. R. Vollaro, “Lincoln, Stowe, and the ‘Little Woman/Great War’ Story: The Making, and Breaking, of a Great American Anecdote,” *Journal of the Abraham Lincoln Association* 30, no. 1 (2009): 18. The quotation by Andy Clark is from A. Clark, *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence* (Oxford: Oxford University Press, 2004), 23–24. Sensory substitution is described in P. Bach-y-Rita and S. W. Kercel, “Sensory Substitution and the Human–Machine Interface,” *Trends in Cognitive Sciences* 7, no. 12 (2003): 541–546, <https://doi.org/10.1016/j.tics.2003.10.013>. Virtual reality is described in M. Slater, B. Spanlang, M. V. Sanchez-Vives, and O. Blanke, “First Person Experience of Body Transfer in Virtual Reality,” *PLOS ONE* 5, no. 5 (2010): e10564, <https://doi.org/10.1371>

/journal.pone.0010564. Another example of body substitution can be found in Solène Neyret, Xavi Navarro, Alejandro Beacco, Ramon Oliva, Pierre Bourdin, Jose Valenzuela, et al., “An Embodied Perspective as a Victim of Sexual Harassment in Virtual Reality Reduces Action Conformity in a Later Milgram Obedience Scenario,” *Scientific Reports* 10 (April 2020): article 6207, <https://doi.org/10.1038/s41598-020-62932-w>. For descriptions and short videos of some of the cyborgs that are described in the text, see Victor Tangermann, “The Future Is Here: Six of Today’s Most Advanced, Real-Life Cyborgs,” *Futurism*, October 17, 2017, <https://futurism.com/six-of-todays-most-advanced-real-life-cyborgs>. For the development of neural implants by Elon Musk’s company Neuralink, see E. Musk and Neuralink, “An Integrated Brain–Machine Interface Platform with Thousands of Channels,” *bioRxiv* 703801, <https://doi.org/10.1101/703801>.

5.11 EXTENDED ETHICS

Robot ethics is discussed in T. Metzinger, “Two Principles for Robot Ethics,” in *Robotik und Gesetzgebung*, ed. E. Hilgendorf and J.-P. Günther (Baden-Baden: Nomos, 2013), 263–302. For discussions of animal welfare, see L. Gruen, “The Moral Status of Animals,” in *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), ed. Edward N. Zalta, <https://plato.stanford.edu/archives/fall2017/entries/moral-animal/>.

The possibility of consciousness in organoids is discussed in T. Bayne, A. K. Seth, and M. Massimini, “Are There Islands of Awareness?” *Trends in Neurosciences* 43, no. 1 (2020): 6–16, <https://doi.org/10.1016/j.tins.2019.11.003>.

I. McEwan’s book is titled “Machines Like Me” (London: Penguin, 2019).

5.12 SOLARIS: THE LIMITS

The text is based on S. Lem, *Solaris* (London: Faber and Faber, 1970). The first quotation is from page 126; the second is the last sentence in the book, on page 214.

The final quotation is from J. Rostand, *Inquiétudes d’un Biologiste* (Paris: Stock, 1967), 66. It is sometimes translated as “The biologist passes, the frog remains.”

INDEX

- Agency, agent, 98, 106, 164, 208
Alloparenting, 134, 135, 140–142, 144
Amphibians, 55, 56, 60, 62
Anthropocene, 167, 168
Art, artistic, 76, 134, 136, 152, 153, 168, 176, 186, 188, 190, 212
Arthropods, 68–69, 114, 116, 118, 119
Artificial intelligence, 200, 204, 206
Attention, 54, 76, 96, 98, 105, 134
Autistic, atypical person, 178, 180, 198
Avatar, 211, 212

Bacteria, x, 34, 36–37, 86, 88, 112
Biopsychism, biopsychist, 36, 37, 42
Birds, 54, 55, 56, 60, 64–65, 66, 76, 84, 110, 125, 140, 166, 207
Blushing, 144–146, 164, 200
Brain, ix, 2, 10, 14, 16, 20, 23, 26–28, 46, 50, 52, 54, 56, 62, 64, 66, 68, 72, 98, 108, 114, 116, 118, 140, 142, 148, 150, 170, 172, 174, 176, 178, 188, 192, 196, 198, 212, 213, 214

Cambrian, Cambrian explosion, 118–120, 123
Cambridge Declaration on Consciousness, 54, 108
Causal reasoning, 136, 140, 142
Cephalopods, 66, 72, 73, 114, 116, 118

Chimeras, 216, 218
Conditioning
 classical/Pavlovian, 101
 instrumental/Skinnerian, 102
Cooperate, cooperation, cooperative, 64, 134, 140–142, 144, 148, 152
Cultural, culture, 76–78, 86, 87, 130, 134, 144, 146, 157, 162, 188, 190, 194, 210
 cultural evolution, 136, 138, 139, 152, 156–157
 cultural selection, 88
Cuttlefish, 66, 72, 73, 118, 125
Cyborg, 210–214, 216, 218

Dance, 134, 135
Dream, dreaming, 8, 9, 54, 60, 62, 64, 124, 214
Dualism, dualist, x, 10, 30

Ecological legacies, 76, 86
Ego-center, 24
Embodiment, 98, 106, 208
Emotion, emotional, 10, 37, 60, 64, 72, 75, 76, 98, 108, 110, 130, 136, 140, 142, 144, 146, 152, 158, 160, 170, 172, 196
 emotional contagion, 76, 160
 emotional control, 136, 140, 142, 150
 social emotion, 76, 136, 142, 144
Empathy, 75, 141, 142, 160, 210

- Epigenetic, 36, 44
- Ethics, 216–218
- Evolutionary theory, 82, 83, 84–86, 90, 92
- Evolutionary transition, 88–90, 92–93, 94
- Evolutionary transition marker, 92, 93, 95, 100, 102, 104–106, 116
- False memory, 158, 163, 198
- Fish, 55, 56–58, 60, 62, 66, 96, 122, 123, 125
- Forgetting, 123, 182–184
- Free will, 14, 30, 158, 164
- Function
 - of consciousness, 83, 112–113
 - definition, 112
- Generate-and-test tower, 88, 89, 124
- Goal-directed, 16, 18, 90, 112
- Golem, 203, 204
- Hallucinations, 186, 188, 190, 192
- Hands, 148–150
- Higher-order thoughts (HOTs), 74
- Homology, 56, 64, 132
- Human nature, 130–132, 134, 138–139, 168
- Imagine, imagination, 88, 124–125, 152, 156–157, 158, 160, 162, 163, 164, 183, 210
- Insects, 55, 66, 68, 70
- Intentionality, 98, 106
- Language, 34, 74, 75, 88, 102, 136, 150, 152, 154, 156–157, 158, 160, 162, 172, 198, 200
- Learn, learning, x, 22, 34, 38, 39, 41, 42, 48, 50, 52, 56, 60, 64, 68, 70, 72, 75, 76, 88, 100–102, 104–107, 108, 112, 113, 114–116, 120, 124, 136, 157, 176, 183, 184, 206, 207, 208, 210, 212, 214
 - associative learning, 48, 100–102, 119
 - habituation and learning, 38, 39, 48, 114
 - sensitization and learning, 48, 114
 - social learning, 76, 136, 144
- Lie, lying, 158, 160, 162
- Machine learning, 102, 108
- Mammals, 54, 55, 56, 60, 64, 66, 68, 110, 125, 140
- Mandala, 56, 58, 77, 78, 188, 190, 192
- Meditation, meditative, 176, 196
- Medusa, medusae, 46–48
- Memory, memories, 10, 14, 36, 42, 44, 45, 50, 96, 98, 104, 105, 110, 114, 116, 136, 148, 152, 156, 163, 164, 178, 180, 182–184, 196, 198, 200, 206, 220
 - memory trace, 40, 100, 148, 182
- Mimesis, mimetic, 142, 152, 157
- Mind–body problem, ix
- Mnemonist, 182
- Mode of being, 92, 93, 94, 98, 112, 130, 138, 177
- Moral, human morality, 76, 136, 144, 146, 160, 164, 216
- Motivation, motivating, 26, 98, 120, 206
- Music, music making, 132, 134, 135, 142, 152
- Mystical, 176, 192, 194, 196, 204
- Myth, 46, 76, 78, 122, 130, 146, 152, 156
- Naturalism, naturalist, x, 16–18, 31, 82
- Natural selection, 84, 86, 88
- Nervous system, 14, 41, 42, 44–45, 46, 119, 218
- Noosphere, 166
- Norms, normative, 76, 136, 142, 144, 146, 156, 162, 164, 216
- Octopuses, 55, 66, 72, 73, 115, 118, 206
- Ontogeny, 113, 116
- Organoids, 216
- Overlearning, 122
- Overshadow, overshadowing, 164, 200
- Panpsychism, panpsychist, 12, 31, 37
- Phylogeny, 76, 116
- Physicalism, physicalist, x, 14, 16, 31
- Placozoans, 42

- Plants, x, 16, 18, 40–41, 42, 44, 84, 88, 90, 186
- Psychedelic drugs, 186–188
- Reflex, reflex-eliciting, 40, 42, 48, 101, 104
- Religious, 153, 186, 194, 200, 204
- Representation, representational, 4, 5, 20, 26, 34, 74, 98, 108, 114, 116, 148, 152, 156, 160, 162, 211
- Reptiles, 55, 56, 60–62, 172
- Ritual, 76, 134, 152, 196, 200, 210
- Robot, 11, 18, 30, 177, 200, 202–204, 206–208, 210, 216, 218
- Savant, 178–184, 188, 190, 198–200, 206
- Self, 10, 24, 56, 74, 75, 77, 98, 106, 148, 150, 156, 162, 188, 210, 212, 214, 218
- Sexual selection, 56
- Slime mold, 38–39, 40, 44
- Smoke detector principle, 122
- Social animals, 76, 217
- Solaris, 220–221
- Soul, 6, 8, 10, 16, 17, 18, 90, 92, 148, 170, 194, 204
 - nutrient reproductive, 16, 18, 90
 - rational, 10, 16, 74, 75, 90, 127
 - sensitive, 16, 18, 74, 90, 204
- Sponges, 42, 88
- Squid, 66, 72, 73, 118
- Stream of consciousness, 6, 7
- Stress, stress response, 68, 86, 122, 123, 217
- Suffer, suffering, 74, 122–123, 204, 208, 216, 218
- Symbolic, 74, 76, 78, 88, 90, 130, 136, 152–154, 156, 160, 162–164, 166–168, 172, 177, 180, 208, 210
 - dimension, 152–154, 168, 177
 - explosion, 164, 166–168
 - value, 90, 136, 153, 208
- Synesthesia, synesthetic, 180, 182, 186, 200, 212
- Teach, teaching, 136, 140, 142, 144, 150, 152
- Theory of mind, 134
- Thought experiment, x, 22, 26
 - bat, 20–21, 211, 212
 - brain-in-a-vat, 26–28
 - butterfly, 8–9, 26
 - Mary, 22–23
 - zombie, 30–31
- Tool, toolmaking, 56, 66, 76, 113, 135, 136, 140, 142, 148–150
- Trace conditioning, 104
- Tradition, 6, 76, 77, 144, 150, 194, 202, 204
- Unconscious, 12, 34, 37, 78, 190–192, 194
- Unlimited associative learning (UAL), 100, 104–107, 108, 112, 113, 114–116, 118–120, 122, 123, 206
- Unlimited heredity, 94, 95, 104
- Value, 16, 90, 98, 104, 106, 108, 112, 113, 114, 122, 136, 153, 154, 208, 218, 220
- Vertebrates, 55, 56, 60, 66, 68, 72, 73, 114, 116, 118, 119
- Virtual reality, 176, 210–214
- Wetware, 208
- Worm, 50–52, 84, 124
- Zombie, 14, 30–31