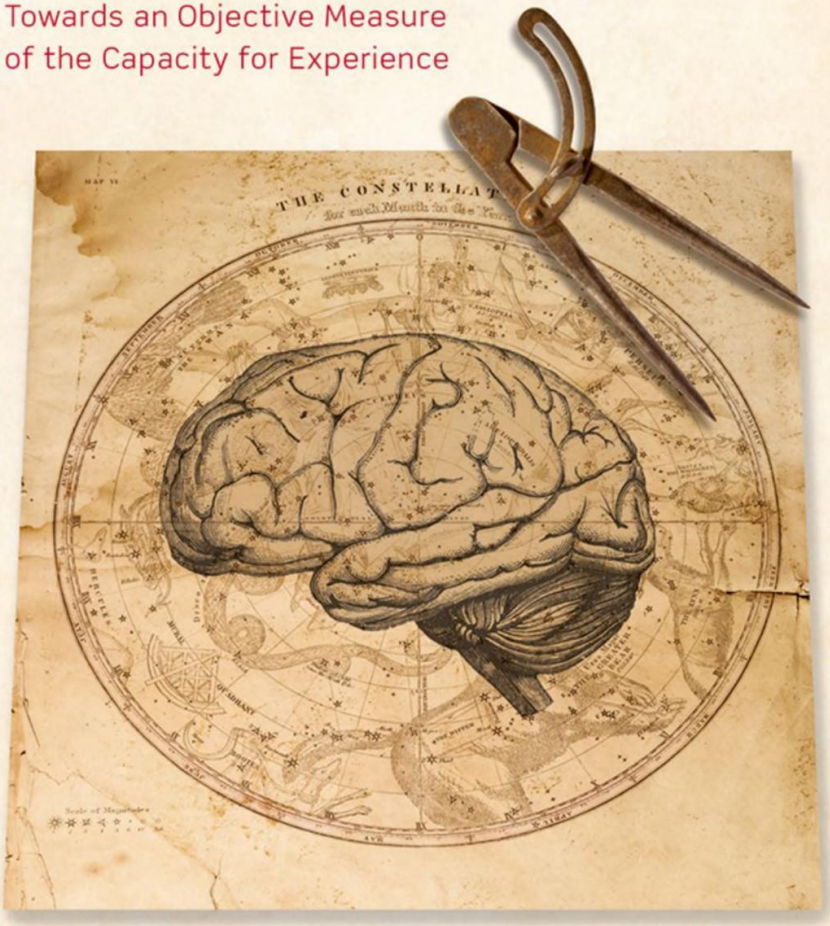
OXFORD

SIZING UP CONSCIOUSNESS

Towards an Objective Measure of the Capacity for Experience



MARCELLO MASSIMINI & GIULIO TONONI

SIZING UP CONSCIOUSNESS

Towards an Objective Measure of the Capacity for Experience

MARCELLO MASSIMINI & GIULIO TONONI

TRANSLATED BY FRANCES ANDERSON

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

Originally published as *Nulla di più grande* di Marcello Massimini e Giulio Tononi,
Baldini e Castoldi 2013

First Edition published in 2018

Impression: 1

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data

Data available

Library of Congress Control Number: 2018933552

ISBN 978–0–19–872844–3

Printed and bound by
CPI Group (UK) Ltd, Croydon, CR0 4YY

The translation of this work has been funded by SEPS
SEGRETARIATO EUROPEO PER LE PUBBLICAZIONI SCIENTIFICHE

Via Val d'Aposa 7 – 40123 Bologna – Italy
seps@seps.it – www.seps.it

CONTENTS

CHAPTER 1

A BRAIN IN YOUR HAND

Our Planet, Our Home

Sometimes events of an extraordinary nature force us to face the actual dimension of things and leave an indelible mark. Take, for example, the astronauts of the Apollo mission, who lived the experience of setting a foot on the Moon; these men were neither poets nor philosophers, they were engineers and military pilots, men of action. They reached the Moon after years of physical training, saturated in technology, chained to rigid sequences and endless checklists, prepared for any eventuality except one . . . the mind-shattering experience of existential flight. At a certain point in time these men saw the Earth, their home planet, rising over the Moon's horizon; in that instant they became vividly aware of a new sense of the world that was difficult to share with other people. There are not many ways of expressing the power of the experience in seeing the Earth from the Moon; in fact the astronauts all use more or less the same words:

> You look back at the Earth from the Moon, and you can put your thumb up to the window and hide the Earth behind your thumb. Everything you have ever known is behind your thumb . . . (Jim Lowell)

> I put up my thumb and shut one eye, and my thumb blotted out the planet Earth. I didn't feel like a giant. I felt very, very small. (Neil Armstrong)

> The Earth reminded us of a Christmas tree ornament hanging in the blackness of space. As we got further and further away it diminished in size. Finally it shrank to the size of a marble, the most beautiful marble you can imagine. That beautiful, warm, living object looked so fragile, so delicate, that if you touched it with a finger it would crumble and fall apart. Seeing this has to change a man. (James B. Irwin)

These words may seem naive, but on reflection they are rather deep and touching. There are no attempts here to expound complicated concepts; they express simple, sincere, and genuine astonishment. There they were, men who for years had dreamed of discovering the Moon, and all of a sudden, totally unexpectedly, they discovered the Earth. They saw their planet from a distance (approximately 300 thousand km) in a new perspective. All joys, sufferings, separations, dreams, everything distilled in a colored marble that you can hide behind your thumb. Back home, the astronauts mulled over their experiences on the Moon, and talked about them on the press and on television. Even when the overwhelming impression of the first sight of the Earth from the Moon started to fade, they still remembered the strange sensation of realizing that the bright sky that spans the human tragedy and comedy is, in fact, just a thin curtain and beyond it is darkness, immense and icy. The astronauts tell how their initial feelings of anguish soon transformed into liberation, serenity, and wonder. In a split second they were able to shrug off not only the fatigue of their mission, but also the weight of ages of tormented history. Many of them tried to express the profound and constructive affection they felt for that fragile home and its inhabitants, others went so far as to suggest that this spectacular sight of the Earth from the Moon, if accessible to all (including those in positions of power), would convince mankind to make a fresh start [1].[1] Archibald McLeish, an American poet, after having seen a photograph of the Earth observed from the Moon, brought back by the Apollo mission, attempted to interpret the state of mind of the astronauts with these lines published on the *New York Times* on December 25, 1968:

> To see the earth as it truly is, small and beautiful in that eternal silence where it floats, is to see ourselves as riders on the earth together, brothers on that bright loveliness in the eternal cold, brothers who know now they are truly brothers.

Unfortunately, we cannot say that these words have had a major impact on our attitude toward the planet. However, we are left with the impression that the sight of the Earth from the Moon succeeded in putting

1 This cognitive shift reported by a number of astronauts is known as the overview effect; for a detailed description of this phenomenon see reference [1].

things in their correct perspective, even if for a very short time. It was like the perfect incarnation of the Copernican revolution, only 400 years later. Of course, it is difficult to describe the sensations that the astronauts felt in the exact moment when they touched the wonderful modesty of the Earth with a fingertip. Indeed, it is hardly worth the effort; certain facts unleash their power only when they hit our senses directly. In 1968 everyone knew that the Earth is round and that it lies on the periphery of an immense universe. However, knowing this is one thing, feeling it is another.

Our Brain, "Our" Home

Today much is known about consciousness and its relationship with the body. We know that it does not depend on the lungs, liver, or heart, but on a handful of neurons in the skull. It only takes a small lesion in the occipital cortex to lose the perception of color or of faces. For each of us, a traumatic brain injury or a small dose of anesthetic may obliterate the universe as we know it, including ourselves. Even if you have never experienced coma or general anesthesia, your nightly descent into sleep is your own personal experience of how delicate the relationship between consciousness and brain matter is. Every night, when we fall asleep, something changes in the way our neurons function and suddenly we do not exist anymore—to all extents and purposes, the universe as we know it dissolves into nothing. In fact, we die every night. This is extraordinary, but we don't often think about it, perhaps because we cannot experience unconsciousness. When we are unconscious, we are not present. As Epicurus pointed out long ago, there is little point in worrying about when we are not there, before birth or after death. When death comes, we go. Maybe we are not overly concerned about the annihilation of deep sleep because we know (or think we know) that the universe will come back in the space of a few seconds upon waking up. Maybe we take consciousness and our brain for granted, just as we take for granted the planet upon which we live and other things that we cannot appreciate in their true dimension because we are completely immersed in them.

If we decide to face the question of the brain-consciousness relationship straight on, the essence of the mystery eludes us. Even those who,

armed with good intentions and a strong dose of curiosity, attack the vast literature that has accumulated on the subject, soon discover that it is far too easy to get lost. Start with the name itself. A disconcerting plethora of notions and definitions lie behind the term "consciousness." In certain contexts the term is used to indicate moral conscience, in others to signify awareness of self, and in others again it indicates the ability of an individual to react to a stimulus. This is just the tip of the iceberg. If we investigate further, particularly if we venture into the realms of the philosophy of mind, we soon get the impression that we must have embarked on a long and tortuous path that most likely will lead nowhere or may even turn out to be circular. It seems inevitable that we must take some philosophical stance on the relationship between matter and mind, but which? The possibilities are many, the choices seemingly arbitrary, and reference points rare—one begins to suspect that medieval scholar mulling over the proofs of God's existence must have felt similarly. Book after book, page after page, the concept of consciousness begins to disintegrate into myriads of distinctions and categories [2–4]. Who can blame us if we get discouraged? The British scholar Stuart Sutherland expressed his frustration in the introduction to his *International Dictionary of Psychology*:

> consciousness is a fascinating but elusive phenomenon: it is impossible to specify what it is, what it does, or why it evolved. Nothing worth reading has been written on it.

A harsh judgment, and you might think that it is hardly an appropriate quotation for the first chapter of a book on consciousness. It certainly expresses the despair still shared by many students and scholars alike after struggling with this subject in a nutshell. This is why we would like to open this book with an experience, rather than a definition. We will have time to sharpen the blades of logic later; for now, let us start the journey by giving free rein to our instincts.

So how can we stimulate our senses to "recognize" consciousness? Is there a way to feel the mystery of consciousness in all its power? Maybe there is one, though it does not have the heroic quality of the astronauts' experience; quite the contrary. In fact, to come to terms with consciousness, we must leave the aseptic and majestic atmosphere of outer space and step into the shoes of a medical student getting to grips with

his first autopsy, trying to conquer waves of nausea in the malodorous and narrow environment of the mortuary. A first autopsy is an experience that is difficult to forget. The moment the pathologist picks up their knife, and opens the chest and the abdomen of the cadaver lying on the steel table, nearly everything that we have learned with such effort during the years of medical study falls apart before our very eyes. Those tidy organs, neatly distinct in the illustrations in the anatomy text book, are not so tidy nor are they so neatly separate; the liver isn't as brown, nor are the lungs so blue and the heart is certainly not so red. In a split second the pumps and filters, the levers and gears, all the crystalline mechanisms of physiology melt into a homogeneous mush, while wafts of the odor of decomposition penetrate our masks that have been (ineffectively) soused in aftershave. It is quite a shock to realize that we are made of such crude matter, which decomposes so rapidly. Even the most cynical medical student struggles against accepting the concept of his body being material, but it just takes one autopsy to end the struggle and this is just the beginning.

After examining the internal organs of the chest and the abdomen, the pathologist slits open the scalp, folds it back, and uses a vibrating saw to cut open the calvarium. Two or three sharp taps with a chisel and the skull cap is removed and deposited on the dissecting table with a thud. These sounds leave their mark, and echo in your ears for days just as the crash of a road accident will, if you are caught up in it. After freeing the convexity of the brain from the membranes that envelop it, the pathologist uses a spatula to lift out the frontal lobe and takes dissecting scissors to the optical nerves, the acoustic nerves, and all the other fibers that tie the brain to the cranium. One final cut to the brain stem and the cerebrum is free. The pathologist turns to the nearest student and places the brain in their palm. It is then passed from hand to hand for the students to examine it in turn. Now imagine you are next. You have a choice; you can either observe this organ like you analyzed the spleen, the liver, and the heart, then pass it on to the person standing next to you or you can stop for a moment and ponder that this damp and jelly-like mass, lying inert in your hands, was a universe as vast as your own, just a few hours ago. Everything that you are, everything that you know, that you remember, imagine, and dream is contained in an object that can be handled like any other worldly object. A thing with mass and borders. Your mind begins to whirl as it did when, as a child, you tried to

imagine the immensity of the universe, and the profusion of galaxies and stars. The dizziness that you are experiencing may even be stronger than what the astronauts felt when watching the small Earth setting behind the Moon. Holding a brain in your hand is an overwhelming experience that, in the blink of an eye, erases all the habits, philosophical positions, definitions, and logical traps that stand between us and the mystery of consciousness. It is almost an initiation, one that any scholar interested in consciousness should go through. A simple question seems to spring spontaneously from the nerves of the hand holding the brain, clamoring for an answer. What makes this object so different from the rest? What makes it so special?

There, standing in front of the dissection table, you don't ask how it is possible that brain matter can feed the contorted flames of self-reflection, nor do you feel the need to understand how it manages to produce the perception of a scene from one of Brueghel's crowded paintings. You have just examined the liver and heart, so you simply ask how it is possible that this tofu-like organ, which weighs 1.5 kg if that, can host a subject who can see light or just pure dark. Why the brain, but not the other organs? Of course, you are a model student and you remember that the brain generates electric signals; but just a moment! So, does the heart. Of course! the brain is composed of tens of billions of neurons and trillions of synapses . . . ah, but within the brain the cerebellum has even more neurons and synapses than the cerebrum, and it has nothing to do with consciousness. Then you remember that during the physiology course they told you that, in certain phases of sleep, the brain isolates itself temporarily from the nerves that connect it to the outside world and starts dreaming—a vivid, colorful, riotous universe, generated entirely from within. You are struck by the thought that this fragile mass of tissue lying here in your hand could dream, if it were drenched in the right solution of sugar and oxygen, and while you are still standing there trying to collect your thoughts, your neighbor nudges you; it is time to pass the brain on. This intense and tangible mystery unfolds in little over 1 min, but what power it can have!

As students, we both had a similar experience in the mortuary. Both of us are convinced that the effect of holding a brain, feeling its texture and weight, is not unlike hiding the Earth behind your thumb at 300 thousand km. It is a sublime experience, in the philosophical and literary sense of the term; it is both a source of mental anguish and

liberation. In the first place, it is disturbing to have to attribute the perfection, beauty, and integrity of what we can conceive and perceive to such a humble object; a small greasy machine whose working parts will, sooner or later, break down and melt. This cannot be right; we are much more than this! A glance at the cadaver on the dissection table, his open eyes, however, is enough to convince us that not long before, this body was a being who could see, hear, feel, and think, from the smallest to the grandest though, just as we can. The astronauts saw the extraordinary richness of the world in a tiny colored marble suspended in space. We grasped the borders of the humble matter that contains anything that can be experienced. During the autopsy, we have held the forbidden fruit and there is no going back. Innocence is lost. We must swear that we will not accept pseudo-solutions and will take nothing for granted. We are aware that a valid scientific explanation of consciousness must stand up to the test of facts and measures, but we also know that it must stand up to the test of our senses and the judgment of our instinct. If one day, holding a brain in our hand will be much more a revelation and much less of a challenge, we will have somehow succeeded.

CHAPTER 2

ZOMBIES AND DOLLS

We have just stepped into the shoes of a young medical student staring at a grey, soft mass in the palm of his hand. Just like Hamlet musing on the skull of Yorick, the court jester, he is deeply disturbed by the juxtaposition of the richness of being and the poverty of matter. Both our student and Hamlet are faced with a tangible and unavoidable question. What is so special about this small and seemingly modest object? There must be something! For many, this question will remain unanswered. Any attempt to *understand* the physical weight of consciousness is doomed to failure, and the miracle of how the brain produces consciousness will remain just that, a miracle that requires an act of faith like the miracle of water being transformed into wine [5].[1] The young student is obviously not in the mood to give up, but for the time being, let skepticism take the upper hand. In fact, skepticism has deep historical and philosophical roots, and an interesting rationale that is well worth exploring.

A Philosopher's Doubt

Dualism, a line of philosophical thought that still poses tough challenges to a scientific approach to consciousness, has a long history that started, at least officially, with Descartes. The French philosopher held that consciousness has nothing to do with the physical world [6]. In his view, the fact that we can have a clear and distinct idea of ourselves as thinking beings, completely different from the idea we have of our body as a material extension, is irrefutable proof that consciousness and the mind are

1 The effective analogy of the water of the physical brain turned into the wine of consciousness was introduced by the British Philosopher Colin McGinn [5].

two separate entities. The attribute that defines matter is its extension—the fact that it occupies space, and so can be measured and explained in scientific terms. The attribute which defines the mind, on the other hand, is thought, which being immaterial in nature cannot be measured. These two substances are ontologically different and the insurmountable barrier between material and mental substances, between the *res extensa* and the *res cogitans*, lies at the heart of the dualist standpoint.

Centuries after Descartes, this position still enjoys support, because it appears simple and easy to grasp, and also because it complies with the natural human reluctance to be considered on a level with the other objects on the planet. On a closer look, however, the dualist approach also has its share of contradictions and problems. For example, how is it possible that something that affects the body (a burn, for example), provokes a response at mental level (the sensation of pain)? How is it possible for a thought to result in the physical movement of an arm or leg? How can a chemical anesthetic or a physical trauma pause our existence for hours or even years? Even the most hard-boiled dualist has to admit that there is a connection between matter and mind somewhere, but where? Descartes solved the question by proposing the pineal gland (epiphysis), a small structure situated in the center of the brain, as an exclusive place in which mind and matter could interact. This solution did not go down well in the 16th century and obviously does not pass muster in today's more sophisticated environment. However, even if we smile and discard Descartes' attempt as ingenuous, the fact remains that there is still no agreement as to where two ontologically separate substances might interact.

This said, today's dualists appear to be less concerned with finding a solution to the mystery of the mind–matter relationship than with underlining the limits of the scientific approach to the problem. We are not going to find the place where matter becomes experience, simply because it does not exist. Consciousness is one thing and matter is another, and science has no way of making the two become one. This is an important point, which deserves serious pondering.

Philosophical Zombies

Believe it or not, contemporary philosophy often uses zombies to illustrate the dualist perspective. These philosophical zombies (also

known as p-zombies) came to notice thanks to the Australian philosopher David Chalmers [7] and, of course, have no relation to the Creole undead evoked by Caribbean witch doctors or to the partially decomposed corpses that chase people in B-movies. Philosophical zombies are well-mannered, respectable, and undistinguishable from the man in the street as far as their behavior is concerned. No medical or psychological examination will reveal any dissimilarity between a human being and a philosophical zombie. The only difference is that the philosophical zombie is totally devoid of subjective experience. If they touch a blistering hot surface they will snatch their hand back, will scream and swear like a trooper, but in point of fact, they do not feel pain. They feel nothing. According to many contemporary philosophers, the very fact that it is possible to think of a creature that from the material point of view is like us, but does not have any subjective experience (Figure 2.1), indicates that consciousness cannot be inferred from physical properties. These zombies don't eat human flesh, but they still manage to terrorize conferences dedicated to the science of consciousness, where they are evoked by witch-doctors camouflaged as cultivated scholars to embarrass and paralyze naïve speakers.

As often happens in philosophical debate, every argument or thought experiment has a counter-argument or -experiment. Of course, zombies are no exception to this. For example, it has been argued that the very possibility of the existence of a philosophical zombie conflicts with our certainty that we are sentient beings [8]. Indeed, if the zombies function as we do, surely they, too, will have this same certainty. If they can entertain this false belief, who can say that we ourselves are not mistaken when we maintain that we are thinking beings? As per this counter-argument, assuming that it is absurd to throw doubt on our subjective experience, philosophical zombies are neither possible nor conceivable. Several books and dozens of articles have been written on this subject, but who is right and who is wrong? This is an example of the circular traps that we would like to avoid, so we will take this argument no further. As we stated in Chapter 1, "A Brain in Your Palm," we prefer to keep our feet on the ground, rather than experience the frisson of logical disorientation; hence, we will now attempt to reformulate the problem of the zombies in a different way.

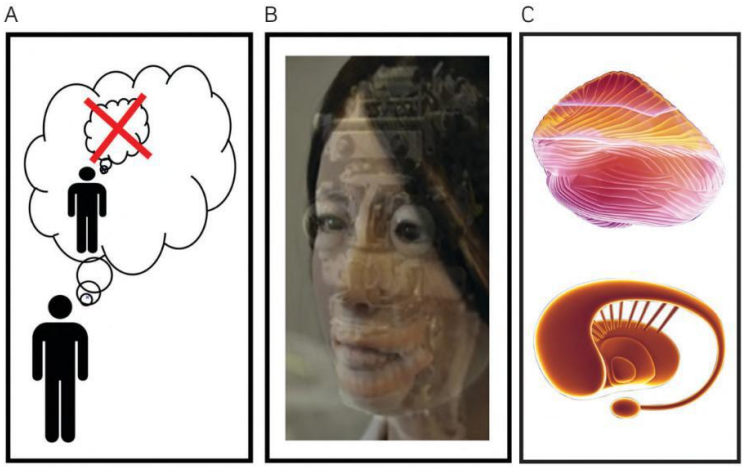


Figure 2.1 Different kinds of zombies. (A) A graphic representation of a philosophical zombie: the possibility of conceiving a being that is physically and behaviorally identical to us, but does not possess conscious experience. (B) Actroid-F, the android developed by the University of Osaka and a forerunner of the digital zombie. The sensors and the activators are visible under the artificial silicon "skin." (C) Two biological zombies, which reside in the cranium—the cerebellum (above) and the basal ganglia (below).

Figure 2.1c Upper: Source: CLIPAREA l Custom media/Shutterstock.com.

Figure 2.1c Lower: Source: decade3d—anatomy online/Shutterstock.com.

Digital Zombies

Take one of the latest generation smartphones, connect it to Internet, and download an app with a search engine, and software for voice recognition and synthesized speech; now you have a small object that can respond to many questions, remind you of your appointments, and suggest a good restaurant close by that fits in with your past choices of menu and price bracket. The voice of such devices is becoming so friendly and helpful that you might soon catch yourself saying thank you to your digital assistant! Of course, until now it has always been perfectly natural to attribute consciousness to someone who listens to us and then replies, but a smartphone certainly isn't conscious. Sure, it may fool us for a few seconds, and then only when we are distracted, but of course we can tell.

However, imagine what might happen if something rather more powerful were to be placed between the microphone and the voice synthesizer, such as Watson, the 3-million dollar IBM computer with 2880 processors and 16 terabytes RAM that can process 500 GB/s. In 2011, Watson was invited to participate in the popular American television quiz show, Jeopardy [9]. The game consisted of pressing a button before the other participants, then answering a series of questions correctly. These questions were quite difficult to interpret as they were formulated like crossword clues, playing on words and double meanings. For three episodes, which attracted huge audiences, Watson challenged the two champions of the program. The computer played large sums of money when it was sure to win, just a few dollars when it was not sure of the answer, and didn't press the button when it did not have the answer at all. Its monitor even blushed when it made a mistake. In the end it won, much to the disappointment of the two human champions and the delight of the public, who had become its ardent fan in just a few hours. Poor Watson didn't know it had won, however; in fact, Watson never experienced anything about anything. When a modern computer responds correctly to our input, when it produces an error message, crashes, or reboots, it does not understand the meaning or significance of what it is saying or doing; there is a consensus on this, even among the audience who instinctively attributed to Watson plenty of human ambitions and emotions.

In fact, it is unlikely that the software of the supercomputer would have passed the exam that computers should "dream" of passing—the Turing Test, introduced in 1950 by Alan Turing [10], the British mathematician who influenced the outcome of World War II by deciphering the German codes and considered to be the father of information technology. The Turing Test was constructed around the question "how can we know if a computer understands what it says?" Turing devised a pragmatic solution, based on the following principle: a machine can think and understand what is said to the extent that it can sustain a conversation that is indistinguishable from a conversation between two human beings. In practical terms, the Turing Test consists of a conversation held at a distance between two terminals, with an exchange of comments such as we have today when opening a conversation using Skype or Messenger. If, at a certain point, it is clear that the respondent on the other end is a machine and not a human being, then the machine has

failed. If the machine dupes us into believing that we are chatting with a human being, the machine has passed the test.

There is no doubt that Watson, who could have beaten anybody at "Jeopardy!" would have tripped up in the early phases of a simple chat. The first exchanges, "How are you?," "How are things?," would not have constituted a problem, and it might even have been able to respond to specific questions, even difficult ones, but there would have been something "off." After a while, we would have had the irritating impression that the interlocutor on the other end of the line was not paying attention, or had not grasped the context of the conversation. It would not have been long before an absurd misunderstanding or a totally irrelevant answer would have revealed the non-human identity of the other party to the conversation.

While it is likely that Watson would not pass the Turing Test, better hardware running smarter software may soon do that, even in a face-to-face encounter. Imagine a much more powerful supercomputer (even in 2011 Watson was only ranked 94th) and plug it inside a futuristic version of Actroid-F, the gynoid robot developed by the University of Osaka. This humanoid machine replicates the movements and facial expressions of a young Japanese girl, including tics and emotions (Figure 2.1B), through a system of sensors and actuators [11]. It is perfectly conceivable that a mix of microprocessors, optical fibers, sensors and compressed air activators, microphones, and speech recognition software, covered in soft silicone, may one day dupe an unwary interlocutor into thinking it is human for a fairly long period of time. The futuristic android will call on a powerful combination of calculation, speed, machine learning algorithms, and access to immense databases to produce a fluid and appropriate conversation. In a face-to-face encounter, we may be a little perplexed by some inaccuracies, but it is likely that we will attribute them to intriguing cultural differences and, enchanted by the gentle ways of this exotic companion, we may even propose a romantic place for dinner. Then, just as we are about to shyly suggest our favorite restaurant, our enhanced version of Actroid-F may pull the mask off, leaving us with the humiliating sensation of having been fooled by a fancy doll.

After all, on close reflection, it is quite easy to see how a black box, which provides appropriate responses to any question, does not necessarily house someone who understands. This can best be illustrated by the Chinese room thought experiment conceived by John Roger Searle

[12]. Pretend you are in a huge box, armed with pen and paper, and an exhaustive catalogue in which all the questions that can be formulated in Japanese kanji are each associated with the correct answers, also written in Japanese kanji.[2] There is a hole in the side of the box through which someone on the outside passes you slips of paper on which are written questions. You painstakingly consult the catalogue, find the question which corresponds visually to the one written on the paper you have received, and copy down the answer. You provide the correct answer every time, even though you do not know Japanese and have understood nothing, neither the question nor the answer. You have simply followed the instructions to the letter. This will probably be the system the android will follow, just much faster, to converse with you in the future. Millions of processors will consult an exhaustive catalogue of possible questions and answers, and provide the most appropriate answer at the speed of light. This is perfectly conceivable. Searle's insight was to put ourselves in the place of the computer in order to understand that, even if we could manipulate the symbols perfectly, we would understand and feel nothing.

This is when we are struck by a thought. What if our brain, too, is just an expert manipulator of symbols? Just as the silicon microprocessors exchange digital symbols along optical fibers, so the neurons in our brain exchange electrical impulses through microscopic cables composed of fat and proteins. Could they do this without understanding anything, without feeling anything, without "being" anyone? We hope that this digression into the world of digital zombies has helped to clarify the doubt expressed by the philosophical zombies. In all cases, the most remarkable zombies live elsewhere. Indeed, they are much closer.

Zombies Within Our Skulls

In 2001, Christof Koch and Francis Crick [13] published a note in the journal *Nature* called "The zombie within." In the note, they argued that our brain contains a large number of "zombie systems," whose hallmarks are efficient sensorimotor behavior and immediate, rapid action.

2 Here, to be consistent with the nationality of Actroid-F, we play with Japanese writing, rather than with Chinese characters as in the original formulation.

Every waking day, we live in close contact with faithful and silent servants of whose presence we are mostly unaware. We don't notice them because they do what they do in the dark, far from the flame of consciousness, but they are there. In fact, we often teach them to do what they do. Imagine (or remember) the first lesson on the piano. You can take nothing for granted, you must make a conscious effort in every movement—how to sit on the stool, the correct posture, how to place your hands on the keyboard, how to move your fingers. You touch the keys—your performance is slow, staccato, labored. As time goes by, you become more proficient and your performance improves. The learning process that a human being undertakes is both tangible and extraordinary, but what is truly amazing is that once you have learnt the complicated motor sequences that in the initial phases absorbed every iota of concentration, they cease to appear on your "radar." You no longer have to consciously think about these sequences. They become so automatic they cease to exist and your consciousness, liberated from the fetters of the motor sequences, can apply itself to how the music should be interpreted. It is almost as if a mysterious entity in the subconscious, which moves surely and swiftly, much more surely and swiftly than we ourselves do, has taken over. Someone or something is playing the pianoforte in your place, and is playing it well. Music teachers understand this phenomenon perfectly, and at a certain point tell their students "let your fingers play, don't think about it," and one of those faithful zombies that live in the cerebrum will execute the notes fluidly, without hesitation, and without making even the slightest mistake. If you try to take back control of the motor sequence, then inevitably your performance slows down or you hit the wrong notes. This is because your inner zombie has become much better at playing the piano than you, and does not appreciate interference, only indications of a very general nature. The pianoforte is just one example, but there are many others—when your hands and feet automatically work accelerator and brake, to maintain the correct distance from the vehicle in front, while you chat to your companions; when you find the right words at the right time, with the right inflection and meaning, while in reality you only have a vague idea of what you want to say; when your skis navigate bumps in the snow, while you admire the view of distant peaks. In all these instances you have to thank an unconscious entity that works for you.

Where do these scrupulous, devoted servants live? The answer comes from clinical neurology. It is probable that there are zombies all over the place in our brain, but certain types of cerebral lesion have shown that the majority are to be found in two neural structures: the cerebellum, in the posterior cranial fossa and the basal ganglia, a voluminous mass of neurons drowned in the depths of the cerebral hemispheres (Figure 2.1C). If one of these two structures is destroyed or degenerates, life becomes very difficult. We become painfully conscious of every gesture, particularly of those of which we were previously hardly aware. When an inner zombie dies, we suddenly become aware of the importance of the work it did for us. Even the most banal tasks, such as grasping a glass of water, become arduous. Distances must be calculated, movements measured, the hand has to open at the right time and be closed with the correct amount of pressure—too strong a grip and we break the glass, too weak and the glass slips through our fingers. With all these complications, errors are inevitable [14]. Many, too many, of the tasks that were carried out by the zombie now have to come under our control [15]. The luxury of that light unconscious management that we took for granted is gone for good and the resulting overload is impossible to manage. If the cerebellum is destroyed, even shaking a friend's hand is a complicated task, and driving a car or learning to play a musical instrument becomes extremely hard.

We become aware of these zombies when they disappear, but also when they malfunction and stop collaborating. Tourette syndrome, commonly known as Tourette's, is a disorder of the basal ganglia characterized by forms of behavior, some of which are quite complex, that the person is unable to control [16]. Symptoms include tics, such as spontaneous exclamations, sometimes socially inacceptable (coprolalia), repeating what others have said (echolalia), and repetition of one's own words (palilalia), although blinking and throat clearing are the most common symptoms. Persons with Tourette's may suddenly execute complicated dance steps or pirouettes. They often have a high IQ, but face great challenges in life because they are not always able to control what they say and do. It is as if a mischievous alien, residing in the depths of their brain, takes over their body, and makes them move and speak out of character.

These zombies are more than just a logical possibility. They are composed of neurons, synapses, and circuits; they live in our brain, they

speak, walk, and do other complicated things without sparking the flame of consciousness. Why? What is the mysterious ingredient that is missing in their make-up? As was argued by Koch and Crick, the existence of zombie systems raises two questions. First, why aren't we just big bundles of unconscious fast zombie agents? Why bother with consciousness, which takes almost half a second to set in? Second, what is the difference between the neuronal circuits that make up zombie agents and those that support conscious experience? Addressing these kinds of questions is exactly what scientists can and should do. Indeed, when Crick and Koch jump-started the whole field of the neuroscience of consciousness and shattered the unwritten rule that forbade scientists to even mention consciousness as a subject worthy of serious inquiry, they did so precisely because they knew how to frame their questions in concrete, experimental terms. The zombies in the brain seem to be the incarnation of the abstract doubts of the philosopher. We can touch them and study them, but until we have understood them, skepticism is more than justified.

A Neuroscientist's Doubt

So, what can neuroscientists say about consciousness? After all, they are in close contact with damp biological matter every day. One thing is certain; they will not have an inferiority complex with regard to philosophers. Physiology has been extremely successful in explaining how the heart pumps blood, and how the liver and the kidneys filter and purify it. Over the last 50 years, neurophysiology and the neurosciences in general have literally exploded and our knowledge of the brain has increased exponentially. An untold quantity of data accumulates each day in thousands of neuroscience departments and continues to grow. Teams of students, graduates, postgraduates, and researchers slave daily on experiments and data, using increasingly powerful instruments. Ten years ago, it was difficult to record brain activity with more than 20 sensors; now we can cover the brain with hundreds of electrodes. In the past, a university purchasing a magnetic resonance imaging (MRI) scanner made the national headlines, today you can find scanners in almost any basement, where the anatomy and the metabolism of the human brain can be recorded with millimeter precision in vivo. Not to mention the

electron microscopes, which can capture the detail of a synapsis with Angstrom-level spatial resolution (0.0000007 mm), two-photon imaging techniques, optogenetics, and the like.

There are scholars in the field of neuroscience who have dedicated their lives to understanding the workings of a single molecule, expressed in a particular class of neurons, located in a particular structure of the brain, anywhere between the periphery of the spinal cord to the more noble areas of the association cortex. The competition between laboratories is fierce; no angle of the brain forest has been left in its virgin state, researchers have marked every tree, colonized and cultivated every millimeter. You are curious about the structure of the cells in the eye of a fruit fly? You want to know about the neurons of the human cerebral cortex that are specialized in recognizing Jennifer Aniston [17]? The stands of the Annual Meeting of the Society for Neuroscience, a bustling scientific fair, which attracts at least 30,000 scientists every year, have it all.

All this effort has been rewarded with results. Neurophysiologists are revealing the neural mechanisms that control how we grasp an object—the programing of motor sequences, anticipatory postural adjustments, the perfect coordination of the limbs, fine control of the force needed from the finger muscles. These are complex mechanisms that are not yet completely understood, but it is just a question of time, there is nothing mysterious or unfathomable. Similarly, they are demystifying much of the visual system's working—from the receptive field of the rods and cones of the retina to the mechanisms of motion detection higher up in the cortical hierarchy. These successes are replicated for hearing, smell, touch, and much more.

In principle, it will soon be possible to draw a precise diagram of the auditory and visual systems, of the sense of touch, and the circuits dedicated to motor planning and control. With time and patience, we will be able to piece together a detailed diagram of the circuits in the brain and reconstruct them one-by-one, just as a counterfeiter might replicated complicated objects—such as a digital camera, car, or a military airplane—based on a stolen blueprint. The chances are that we will end up with a tangled mass of artificial neurons that behave just like the real ones, but will this bring us closer to the solution of the mystery of consciousness? Not necessarily. Let us see why, again using a tangible example.

The scientists of Lausanne Polytechnic have been working for years on a scientific enterprise without precedent, the Blue Brain project. A team of researchers from a number of fields—physicists, biologists, physiologists and computer scientists—are reproducing everything that is known about the brain into a supercomputer simulation: molecules, synapses, the electrical properties of different neurons and their patterns of connectivity, with the objective of reproducing entire areas of the cortex and, finally, the entire brain. By the end of 2006, the project had achieved its first objective, the reconstruction of a simplified cortical column. This is a cylindrical structure, about 2 mm high and 0.5 mm in diameter, which is thought to represent the fundamental functional unit for the cerebral cortex. Artificial cortical columns are much bigger than biological columns; they are housed in a supercomputer that takes up a large room. A few years ago, the project reached another landmark, the detailed simulation of a portion of the rat's somatosensory cortex containing about 30,000 simulated neurons (of 200 different types) connected by 40 million synapses [18]. This is, without doubt, the most complete simulation to date of a piece of excitable brain matter. In the years to come, petaflops of computer power and smart algorithms will grow this chunk of digital cortex further. Indeed, the project, which represents a fantastic scientific adventure, can count on generous financing and is very ambitious. Its goal is to create a virtual mouse brain with hundreds of millions of neurons and, ultimately, simulate the human brain.

Now, let us fast-forward a few decades and suppose that the empirical data and mathematical models used to reconstruct virtual neurons, their biophysical properties, and their connections are correct and complete, that their activity can be simulated in real time on a powerful supercomputer that can fit inside an artificial head, and that two artificial eyes, two ears, a nose, two nimble robotic hands, two legs and a smooth digital vocalizing system have been plugged in. Now, imagine that you have been asked to sign a rather peculiar contract, which establishes that at a certain point in your existence, when your body starts to show the first irrefutable signs of deterioration or the signs of an untreatable disease, you will hand over your biological brain to a group of scientists. You will be allowed to say your last goodbyes to family and friends, then the scientists will anesthetize you, extract your brain and record its activity with billions of sensors. They will cut it into sections, take photographs of it with the electronic microscope and analyze it thoroughly. Finally, all

your neurons, their biophysical properties, and their connections will be uploaded and simulated by virtual neurons implemented by the supercomputer inside the head of your new artificial body. Every last detail will be simulated faithfully, the synapses of the hippocampus, where your most vivid memories are stored, the circuits of the amygdala where the events and conditioning of your own unique life have carved out fears that you have never confessed to yourself, and so on. Trains of neuronal spikes will run incessantly across the simulation, millions per seconds, replicating the coding and the computations of your original brain and all this will rely on silicon chips, which are impervious to viruses, tumors, or heart attacks, and do not suffer the effects of the passing of time. The "silicon you" will behave exactly as you always did, and everybody will think that inside the new body there is you—the same old you—with all your idiosyncrasies, obsessions, and your few endearing traits. The contract promises you the digital immortality of your brain's activity in exchange for renouncing a few months of biological life. It lures you with the promise of watching your children grow up, enjoying their successes, and offering a helping hand when needed, of continuing to see your friends, and even the option of asking to switch off the whole thing when you have had enough.

The stakes are high, very high, and the goal is to turn one of man's greatest dreams, immortality, into reality. There are doubts, however, just as strong as the stakes are high. How many people would actually sign a similar contract? It is quite possible that not even the most optimistic of scientists, the most hard-boiled materialist and anti-dualist would put his signature to this agreement, when his time comes. Not in the name of moral qualms, but because an inner voice whispers incessantly, "Who says that being a machine that simulates exactly the activity of all your neurons is the same as being you? Who says that scientists have understood the relevant properties of the brain that have to be reproduced? What if the spatial–temporal grain at which the simulation runs is not the right one? What if the secret of consciousness is tucked away in a sub-atomic detail of the functioning of neuronal membranes? What if, instead, the relevant processes occur through unknown interactions that extend beyond my brain? Maybe, it is a particular chemical aspect of the biological matter, which cannot be reproduced on a silicon chip? Am I trading my last conscious days for a reincarnation as

an eternal zombie? Do not sign! Put that pen down and go and hug your family one more time!"

The bottom line is that we totally trust science when it is a question of substituting vital organs, such as hearts and kidneys, with artificial devices, but when it comes to the brain, that is a different ball game altogether. Why are we so reluctant? Because we have a niggling suspicion that however much we know about those billions of neurons in our head and how they function, we are no closer to a scientific explanation of how the brain generates subjective experience. This is the doubt that troubles physiologists, who feel rather like the early astronomers with their detailed descriptions and charts of the movements of the celestial bodies, but who had no idea as to whether those movements were dictated by a general law. When all is said and done, our doubts, and the doubts of the physiologist and the philosopher, may share the same rational root. When it comes to the relationships between consciousness and the brain, we have described a lot, but we lack principles.

CHAPTER 3

BRAIN ISLANDS

Now we leave the morgue, the digital dolls, and the zombies, and move to the intensive care unit (ICU), where living human beings fluctuate between consciousness and unconsciousness, and where sometimes the flame of experience burns unseen on brain islands lost in a sea of neural dissolution. Here, among the tubes and drip-feeds, artificial lungs, and monitors, the way we pose the question changes, but the question itself remains the same.

Bad Awakening

How do we know a fellow human being is conscious? In the normal run of things, the question doesn't even arise. Unless we are really into philosophical zombies, we just assume that others are conscious beings—we are similar physically, so we will be similar in having subjective experiences. If your friend has nodded off on the settee in front of the television, or seems to be lost in a world of his own, you can ask something along the lines of "are you with us?" and a gesture of the hand or a grunt is all that is needed to reassure you that he has not "lost his senses." A doctor uses much the same method to evaluate the level of consciousness of a patient who has just been given a dose of anesthetic in the operating theater, or of a man arriving in the Emergency Ward with his eyes closed and his face covered in blood. In cases such as these, the doctor asks the patient to open his eyes, say his name, say where he is, and clench his right or left fist. If the patient responds, he is conscious. If there is no response, the doctor will try exerting pressure on the palm of the patient's hand, or on the nail bed to elicit pain. If there is still no response, the patient is considered to be unconscious. So, the fundamental criterion for establishing the conscious existence of another human being is whether