Marcelo Gleiser

THE ISLAND OF KNOWLEDGE

The Limits of Science and the Search for Meaning

Published by Basic Books,

A Member of the Perseus Books Group

Books published by Basic Books are available at special discounts for bulk purchases in the United States by corporations, institutions, and other organizations. For more information, please contact the Special Markets Department at the Perseus Books Group, 2300 Chestnut Street, Suite 200, Philadelphia, PA 19103, or call (800) 810-4145, ext. 5000, or e-mail special.markets@perseusbooks.com.

Book design by Cynthia Young

Library of Congress Cataloging-in-Publication Data
Gleiser, Marcelo, author.
The island of knowledge : the limits of science and the search for meaning / Marcelo Gleiser.
pages cm
Includes bibliographical references and index.
ISBN 978-0-465-08073-1 (e-book) 1. Science—Philosophy. 2. Meaning (Philosophy)
I. Title.

Q175.32.M43G57 2014

501—dc23

2013046890

10 9 8 7 6 5 4 3 2 1

Contents

Prologue

THE ISLAND OF KNOWLEDGE

> *What I see in Nature is a magnificent structure that we can comprehend only very imperfectly, and that must fill a thinking person with a feeling of humility.*
>
> —ALBERT EINSTEIN

> *What we observe is not Nature itself but Nature exposed to our method of questioning.*
>
> —WERNER HEISENBERG

How much can we know of the world? Can we know *everything?* Or are there fundamental limits to how much science can explain? If there are limits, to what extent can we understand the nature of physical reality? These questions and their surprising consequences are the focus of this book, an exploration of how we make sense of the Universe and of ourselves.

What we see of the world is only a sliver of what's "out there." There is much that is invisible to the eye, even when we augment our sensorial perception with telescopes, microscopes, and other tools of exploration. Like our senses, every instrument has a range. Because much of Nature remains hidden from us, our view of the world is based only on the fraction of reality that we can measure and analyze. Science, as our narrative describing what we see and what we conjecture exists in the natural world, is thus necessarily limited, telling only part of the story. And what about the other part, the one beyond reach? From our past successes we are confident that, in time, part of what is currently hidden will be incorporated into the scientific narrative, unknowns that will become knowns. But as I will argue in this book, other parts will remain hidden, unknowables that are unavoidable, even if what is unknowable in one age may not be in the next one. We strive toward knowledge, always more knowledge, but must understand that we are, and will remain, surrounded by mystery.

This view is neither antiscientific nor defeatist. It is also not a proposal to succumb to religious obscurantism. Quite the contrary, it is the flirting with this mystery, the urge to go beyond the boundaries of the known, that feeds our creative impulse, that makes us want to know more.

The map of what we call reality is an ever-shifting mosaic of ideas. We will follow this mosaic through the history of Western thought, retracing the steps of our shifting scientific worldview from past to present in three separate but complementary parts. In each part, I strive to illuminate a variety of scientific and philosophical viewpoints, always with the intention of exploring how conceptual shifts inform our search for knowledge and meaning. In Part 1 we will focus on the Universe, its origin and physical nature, and the ways in which our evolving cosmic narrative has framed our understanding of ourselves and of the nature of space, time, and energy. In Part 2 we will focus on the nature of matter and the material composition of the world, from ancient alchemical musings to the most modern ideas about the quantum world and what they tell us about the essence of physical reality and our role in defining it. In Part 3 we explore the world of mind, computers, and mathematics, paying particular attention to how they inform our discussion on the limits of knowledge and the nature of reality. As we will see, the incompleteness of knowledge and the limits of our scientific worldview only add to the richness of our search for meaning, as they align science with our human fallibility and aspirations.

As I write these lines, an unknown choreography organizes the firing of millions of neurons in my brain; thoughts emerge and are expressed as words, typed on my laptop by a detailed coordination of eye and hand muscles. Something is in charge, an entity we loosely call “mind.” I’m flying at about thirty thousand feet, returning from a documentary shooting in Los Angeles. The theme was the known universe, a retelling of the wonderful conquests of modern science, in particular those of astronomy and cosmology. I see the white clouds hovering below, the blue sky above; I hear the jet engines humming and feel the annoying tapping of my neighbor as he listens to music on his iPod.

My perception of the world around me, as cognitive neuroscience teaches us, is synthesized within different regions of my brain. What I call reality results from the integrated sum of countless stimuli collected through my five senses, brought from the outside into my head via my nervous system. Cognition, the awareness of being here now, is a fabrication of a vast set of chemicals flowing through myriad synaptic connections between my neurons. I am, and you are, a self-sustaining electrochemical network enacted across a web of biological cells. And yet we are much more. I am me and you are you and we are different, even if made of the same stuff. Modern science has removed the age-old Cartesian dualism of matter and soul in favor of a strict materialism: the theater of the self happens in the brain, and the brain is an assembly of interacting neurons firing nonstop like lights on a Christmas tree.

We have little understanding as to how exactly this neuronal choreography engenders us with a sense of being. We go on with our everyday activities convinced that we can separate ourselves from our surroundings and construct an objective view of reality. I know that I am not you, and I know that I am not the chair that I am sitting on. I can walk away from you and the chair, but I can’t walk away from my own body. (Unless I’m in some trancelike state.) We also know that our perception of reality, on which we base our sense of self, is severely incomplete. Our senses capture only a sliver of what goes on around us. The brain is unaware of much that is happening, deaf and blind to a huge amount of information that wasn’t particularly useful to increase our ancestors’ survival chances in their hostile environments. For example, trillions of neutrinos racing all the way from the heart of the Sun zip through our bodies each second; electromagnetic waves of all sorts—microwaves, radio waves, ultraviolet, infrared—carry information we can’t capture with our eyes; sounds beyond the range of our hearing go unheard; dust particles and bacteria go unseen. As the Fox said to the Little Prince in Antoine de Saint-Exupéry’s fable, “What is essential is invisible to the eye.”

Measuring instruments and tools greatly extend our view, whether of the very small or of the very far. They allow us to “see” invisible bacteria, electromagnetic radiation, subatomic particles, and exploding stars billions of light-years away. High-tech devices allow doctors to visualize tumors inside our lungs and brains, and geologists to locate underground oil reservoirs. Still, any detection or measuring technology has a limited precision or reach. A scale measures weight only with accuracy up to half its smallest graduation: if the ticks are spaced by one ounce, weights can be stated only to within a half-ounce precision. *There is no such thing as an exact measurement.* Every measurement must be stated within its precision and quoted together with “error bars” estimating the magnitude of errors. High-precision measurements are simply measurements with small error bars or high confidence levels; there are no perfect, zero-error measurements.

Consider now a less prosaic example than weighing scales: particle accelerators. These machines are designed to study the fundamental composition of matter, as they search for the smallest bits of stuff that make up all that exists in the world.[1] Particle accelerators make full use of Einstein's famous $E = mc^2$ formula, converting the energy of the motion of high-speed particles into new chunks of matter. As such, they work in a brutish way, colliding particles that move close to the speed of light head-on. How else could scientists study what exists, for instance, inside a proton? Unlike oranges, protons can't be cut. The solution is to throw protons against one another at high speeds and study the bits that come flying off after a collision. In the absence of a sharp knife, the internal composition of oranges could also be studied this way, tossing one fruit against another at high speeds and watching for the seeds, juice, and guts that come spewing out. Pushing the analogy a bit further: the speedier the oranges, the more telling the experiment. For example, only high-speed collisions would reveal the existence of seeds. A few collisions at higher speeds may even crack the seeds open, exposing their innards. This is an essential point: the higher the energy of the collision, the deeper we see into matter.[2]

During the past half century, particle accelerators underwent a tremendous escalation in power. The radioactive particles that Ernest Rutherford used in 1911 to probe the structure of the atomic nucleus had energies about a million times smaller than those currently achieved at the Large Hadron Collider, the behemoth particle accelerator located in Geneva, Switzerland. Consequently, modern-day particle physicists can probe deeply into the nature of matter, "seeing" things Rutherford wouldn't have dreamed of, such as "elementary" particles a hundred times heavier than a proton, like the famous Higgs boson, discovered in July 2012.[3] If funding is secured for future accelerators—a big if, given the enormous price tag for these machines—there is good reason to expect that new technologies will allow for the study of ever-higher energy processes and will produce exciting, perhaps even revolutionary, discoveries.

However, and this is a key point, technology limits how deeply experiments can probe into physical reality. That is to say, machines determine what we can measure and thus what scientists can learn about the Universe and ourselves. Being human inventions, machines depend on our creativity and available resources. When successful, they measure with ever-higher accuracy and on occasion may also reveal the unexpected. A case in point is Rutherford's surprise when his experiments revealed that even though the atomic nucleus occupies a tiny fraction of an atom's volume, it carries most of the atom's mass. To Rutherford and fellow scientists working during the early twentieth century, the world of atoms and subatomic particles looked very different from how it looks today. We can be equally sure that one hundred years from now our knowledge of subatomic physics will be quite different again. With my argument for now restricted to a purely empirical standpoint, scientists can grasp only what happens at energies within their experimental reach.

That being the case, what could we say *with certainty* about the properties of matter at energies thousands or millions of times higher than current limits? Theories may speculate about the properties of matter at such energies, and they may provide compelling arguments based on simplicity and elegance as to why things should be this way and not another. But the essence of empirical science is that Nature always has the last word: data cares little for our yearnings for aesthetic beauty, a point I explored in detail in my book *A Tear at the Edge of Creation*. It then follows that if we only have limited access to Nature through our tools and, more subtly, through our restricted methods of investigation, our knowledge of the natural world is necessarily limited.

Coupled to this technological limitation of how we probe the natural world, advances in physics, mathematics, and computation during the past two hundred years have taught us a lesson or two about the elusiveness of Nature. As we will see in detail, there are fundamental limits to how much we can know of the world not only because of our tools of exploration but also because Nature itself—at least as we humans perceive it—operates within certain limits. The Greek philosopher Heraclitus already sensed this twenty-five centuries ago, when he pronounced that "Nature loves to hide." Through countless tales of tribulation and success, we have found out that Nature cannot be beaten in this game of hide-and-seek. To paraphrase Samuel Johnson's frustration with trying to define certain English verbs, it is as if we were trying to paint a forest's reflection on the surface of a stormy lake.

As a consequence, and despite our ever-increasing efficiency, at any given time large portions of the natural world remain unseen or, more precisely, undetected. This shortsightedness, however, is a powerful tease to our imagination: limits should not be seen as insurmountable obstacles but as challenges. As the French prescient author Bernard le Bovier de Fontenelle wrote in 1686, “We want to know more than we can see.”[4] Galileo’s 1609 telescope could barely discern Saturn’s rings, something that household telescopes can do today. What we know of the world is only what we can detect and measure. We see much more than Galileo, but we can’t see it all. And this restriction is not limited to measurements: speculative theories and models that extrapolate into unknown realms of physical reality must also rely on current knowledge. When there is no data to guide intuition, scientists impose a “compatibility” criterion: any new theory attempting to extrapolate beyond tested ground should, in the proper limit, reproduce current knowledge. For instance, Einstein’s theory of general relativity, which describes gravity as the curvature of spacetime as a result of the presence of matter (and energy), reduces to the old Newtonian theory of universal gravitation within the limit of weak gravitational fields: to launch spaceships to Jupiter we don’t need Einstein’s theory, but to describe black holes we do.

If large portions of the world remain unseen or inaccessible to us, we must consider the meaning of the word “reality” with great care. We must consider whether there is such a thing as an “ultimate reality” out there—the final substrate of all there is—and, if so, whether we can ever hope to grasp it in its totality. Note that I am refraining from calling this ultimate reality God, since God’s nature, as most religions propose, is ungraspable. It is also not the subject of a scientific inquiry. And I am refraining from equating ultimate reality with any of the several Eastern philosophical notions of transcendent reality, as in a nirvana-like state achievable through meditation, the Brahman from Hindu Vedanta philosophy, or an all-encompassing Tao. For now, I am only considering the more concrete nature of *physical* reality, which we can infer through the diligent application of science. We thus must ask whether grasping reality’s most fundamental nature is just a matter of pushing the limits of science or whether we are being quite naïve about what science can and can’t do.

Here is another way of thinking about this: if someone perceives the world through her senses only (as most people do), and another amplifies her perception through the use of instrumentation, who can legitimately claim to have a truer sense of reality? One “sees” microscopic bacteria, faraway galaxies, and subatomic particles, while the other is completely blind to such entities. Clearly, they “see” different things and—if they take what they see literally—will conclude that the world, or at least the nature of physical reality, is very different. Who is right?

Asking who is right misses the point, although surely the person using tools can see further into the nature of things. Indeed, to see more clearly what makes up the world and, in the process, to make more sense of it and of ourselves is the main motivation to push the boundaries of knowledge, as de Fontenelle knew when he wrote, “All philosophy is based on two things only: curiosity and poor eyesight.”[5] Much of what we do can be summed up as different attempts to alleviate our myopic gaze.

What we call “real” is contingent on how deeply we are able to probe reality. Even if there is such a thing as the true or ultimate nature of reality, all we have is what we can know of it. For the sake of argument, let us concede that sometime in the future a brilliant theory, supported by breakthrough experiments, will make remarkable inferences about the ultimate nature of reality. Even if we were to capture a glimpse of this reality through our detectors, all we would be able to conclude is that the theory makes partial sense: the tool-driven methodology we *must* use to learn about the world cannot prove or disprove theoretical statements about the ultimate nature of reality. To stress my point, our perception of what is real evolves with the instruments we use to probe Nature. Gradually, some of what was unknown becomes known. For this reason, what we call “reality” is always changing. The Earth-centered cosmos of Columbus was radically different from the Sun-centered Newtonian cosmos. The expanding cosmos of today, with billions of swirling galaxies, each with billions of stars, would have mystified Newton. It had even Einstein mystified. The version of reality we might call “true” at one time will not remain true at another.

Of course, Newton's laws of motion will always apply within their limit of validity, and water will always be composed of oxygen and hydrogen, at least within the particularly human way of describing physical and chemical processes with atoms. But these belong to our enduring explanations of the natural world, which are valid within their range of applicability and conceptual structure. Given that our instruments will always evolve, tomorrow's reality will necessarily include entities not known to exist today, whether astrophysical objects, elementary particles, or viruses. More to the point, as long as technology advances—and there is no reason to suppose that it will ever stop advancing for as long as we are around—we cannot foresee an end to this quest. The ultimate truth is elusive, a phantom.

Consider, then, the sum total of our accumulated knowledge as constituting an island, which I call the "Island of Knowledge." By "knowledge" I mean mostly scientific and technological knowledge, although the Island could also include all the cultural and artistic creations of humankind. A vast ocean surrounds the Island of Knowledge, the unexplored ocean of the unknown, hiding countless tantalizing mysteries. Whether this ocean extends toward infinity is something we will return to later on. For now, it is enough to imagine the Island of Knowledge growing as we discover more about the world and ourselves. This growth often takes an uncertain path, the coastline delineating the jagged boundary between the known and the unknown. Indeed, the growth may, on occasion, retrocede as ideas once accepted are jettisoned in light of new discoveries.

The Island's growth has a surprising but essential consequence. Naïvely, we would expect that the more we know of the world, the closer we would be to some sort of final destination, which some call a Theory of Everything and others the ultimate nature of reality. However, holding on to our metaphor, we see that as the Island of Knowledge grows, so do the shores of our ignorance—the boundary between the known and the unknown. Learning more about the world doesn't lead to a point closer to a final destination—whose existence is nothing but a hopeful assumption anyway—but to more questions and mysteries. The more we know, the more exposed we are to our ignorance, and the more we know to ask.[6]

Some people, including many of my scientist friends, consider this view of knowledge to be a downer. I've even been called a defeatist before, which is horrifyingly wrong, since I am proposing the precise opposite: a celebration of humanity's achievements resulting from our endless striving for new knowledge. "If we are never going to get to a final answer, what's the point of trying?" they ask. "And how do you know you are right, anyway?" This book will answer these questions. But just for starters, once we explore the nature of human knowledge—that is, how we try to make sense of the world and of our place in it—it should be obvious that our approach is fundamentally limited in scope. This realization should *open* doors, not close them, since it makes the search for knowledge an open-ended pursuit, an endless romance with the unknown. And what could be more inspiring than knowing that there will always be something new to discover in the natural world, that no matter how much we know there will always be plenty of room for the unexpected? To me, the real downer is to presume that there is an end to the search and that we will eventually get there. Paraphrasing Tom Stoppard in *Arcadia*, "It is the need to know that makes us matter."

New discoveries throw light here and there, but farther out their glow fades into darkness. How we choose to relate to this fact traditionally defines how most people approach life's mysteries: either reason will slowly but surely conquer the unknown, or it won't. If the latter, then something beyond reason is needed to cope with our perennial ignorance, such as belief in alternative explanations, which may include the supernatural. If these are the only two options we have, we are left with the unfortunate polarity between scientism versus supernaturalism that so much defines our age. I propose a third path, based on how an understanding of the way we probe reality can be a source of endless inspiration without the need for setting final goals or promises of eternal truths.

As science advances, we will know more. But we will also have more to know. New tools of exploration present us with new questions. Often, these are questions that couldn't even have been imagined before the tools were available. To take two obvious examples, consider astronomy prior to and after the telescope (1609) or biology prior to and after the microscope (1674): no one could have anticipated the revolutions these two instruments and their descendants would bring about. This unsettled existence is the very blood of science. Science needs to fail to move forward. Theories need to break down; their limits need to be exposed. As tools probe deeper into Nature, they expose the cracks of old theories and allow new ones to emerge. However, we should not be fooled into believing that this process has an end. The scientific approach to knowledge has essential limitations; some questions are beyond its reach. In fact, some key aspects of Nature will necessarily remain unknown to us. Some, I will argue, are unknowable.

To expose the limits of science is far from being an obscurantist; quite the contrary, it is a much needed self-analysis in a time when scientific speculation and arrogance are rampant. By describing the limits of explanations of physical reality based on scientific methodology I am attempting to protect science from attacks on its intellectual integrity. I am also trying to explain how science advances because of our ignorance and not because of our knowledge. As Columbia University neuroscientist Stuart Firestein remarked in his recent book *Ignorance: How It Drives Science*, grant proposals are first and foremost statements of the current state of ignorance. To claim to know the "truth" is too heavy a burden for scientists to carry. We learn from what we can measure and should be humbled by how much we can't. It's what we don't know that matters.

Our perception of reality relies on an artificial separation between subject and object, a question that has inspired (and confused) thinkers of all ages. You may think you know where you end and the "outside" world begins, but the issue, as we shall see, is far from simple. No two people have the exact same perspective on the world. On the other hand, science is the best toolkit we have to create a universal language that transcends individual differences. As we explore our longing to conquer the unknown, we will also experience science's power to transform and to inspire.

PART I

THE ORIGIN OF THE WORLD AND THE NATURE OF THE HEAVENS

In the beginning, God created the earth,
and He looked upon it in His cosmic loneliness.
And God said, "Let Us make living creatures out of mud,
so the mud can see what We have done."
And God created every living creature that now moveth,
and one was man. Mud as man alone could speak.
God leaned close to mud as man sat, looked around,
and spoke. "What is the purpose of all this?"
he asked politely.
"Everything must have a purpose?" asked God.
"Certainly," said man.
"Then I leave it to you to think of one for all this,"
said God.
And He went away.

—KURT VONNEGUT, *CAT'S CRADLE*

It is questions with no answers that set the limits of human possibilities, describe the boundaries of human existence.

—MILAN KUNDERA, *THE UNBEARABLE LIGHTNESS OF BEING*

Man has always been his own most vexing problem.

—REINHOLD NIEBUHR, *THE NATURE AND DESTINY OF MAN*

CHAPTER 1

THE WILL TO BELIEVE

(Wherein we explore the role of belief and extrapolation in religion and in scientific creativity)

Can we make sense of the world without belief? This is a central question behind the science and faith dichotomy, one that informs how an individual chooses to relate to the world. Contrasting mythic and scientific explanations of reality, we could say that religious myths attempt to explain the unknown with the unknowable while science attempts to explain the unknown with the knowable. Much of the tension stems from assuming that there are two mutually inconsistent realities, one within this world (and thus "knowable" through the diligent application of the scientific method) and one without (and thus "unknowable" or intangible, traditionally related to religious belief).[1]

In myths, the unknowable reflects the sacred nature of the gods, whose existence transcends the boundaries of space and time. In the words of historian of religion Mircea Eliade:

> For the Australian as well as for the Chinese, the Hindu and the European peasant, the myths are *true* because they are *sacred*, because they tell him about sacred beings and events. Consequently, in reciting or listening to a myth, one resumes contact with the sacred and with reality, and in so doing one transcends the profane condition, the "historical situation."[2]

Throughout the ages, religious myths have allowed the faithful to transcend the "profane condition," the perplexing awareness we humans have of being creatures bound by time, of having a history and an end. At a more pragmatic level, mythic explanations of natural phenomena were prescientific attempts to make sense of things that were beyond human control, answering questions that seemed unanswerable. Why should the Sun go across the sky every day? To the Greeks, because Apollo transported it in his fiery chariot. To the Navajos of the American Southwest, it was Jóhonaa'éí who hauled the Sun daily on his back across the sky. To the Egyptians, this task belonged to Ra, who transported the Sun in his boat. In a strictly naturalistic sense, the motivation behind such myths is not so different from that of science, as both attempt to uncover the hidden mechanisms behind natural phenomena: after all, gods and physical forces make things happen, albeit in very distinct ways.

More to the point, both the scientist and the faithful *believe* in unexplained causation, that is, in things happening for unknown reasons, even if the nature of the cause is completely different for each. In the sciences, this belief is most obvious when there is an attempt to extrapolate a theory or model beyond its tested limits, as in "gravity works the same way across the entire Universe," or "the theory of evolution by natural selection applies to all forms of life, including extraterrestrial ones." These extrapolations are crucial to advance knowledge into unexplored territory. The scientist feels justified in doing so, given the accumulated power of her theories to explain so much of the world.

We can even say, with slight impropriety, that her faith is empirically validated.[3]

Here is an example. Newton's theory of universal gravitation, as explained in Book III of his revolutionary *Mathematical Principles of Natural Philosophy*, the *Principia*, should really have been called a theory of solar system gravitation, since by the late seventeenth century no tests were conceivable beyond its confines. Yet Newton called Book III *The System of the World*, assuming that his description of gravitational attraction as a force proportional to the quantity of mass in two bodies and decreasing with the square of the distance between them would extend to the whole "world," that is, the cosmos. In his own words, from Book III,

> Finally, if it is universally established by experiments and astronomical observations that all bodies on or near the earth gravitate toward the earth, and do so in the proportion of matter in each body, and that the moon gravitates toward the earth in proportion to the quantity of its matter, and that our sea in turn gravitates toward the moon, and that all planets gravitate toward one another, and that there is a similar gravity of comets toward the sun, it will have to be concluded by this third rule that all bodies gravitate toward one another.[4]

Newton cleverly avoided speculating on the cause of gravity itself—"I feign no hypothesis"—attaching it universally to all bodies with mass: "And to us it is enough that gravity does really exist, and acts according to the laws which we have explained, and abundantly serves to account for all the motions of the celestial bodies, and of our sea," he wrote in the General Scholium of the *Principia*, a sort of concluding explanatory text. He didn't know why masses attract one another, but he knew how they did so. The *Principia* was a book concerned with the hows and not with the whys.

Later, in a letter to the Cambridge theologian Richard Bentley dated December 10, 1692, Newton used his extrapolation on the nature of the gravitational force to justify why the universe should be infinite, a major turning point in the history of cosmological thought. If gravity acted across a spatially finite universe according to the same law of attraction, Bentley wondered, why wouldn't all matter be concentrated in a huge ball at the center? Newton agreed that this would indeed be the case if the universe were finite in extent. However, he went on, "if the matter was evenly diffused through an infinite space, it would never convene into one mass but some of it convene into one mass and some into another so as to make an infinite number of great masses scattered at great distances from one to another throughout all that infinite space." Newton's belief in the universal nature of gravity was strong enough to let him speculate confidently about the spatial extent of the cosmos as a whole.

Centuries later, Einstein did something similar. He formulated his general theory of relativity in final form in 1915, wherein he went a step beyond Newton and attributed gravity to the curvature of space about a massive body (and time, but let's leave this aside for now): the larger the mass, the more space is bent around it, like the elastic surface of a trampoline around people of different weights. No more was a mysterious action-at-a-distance called forth to explain how massive bodies tend toward one another: in a curved space trajectories are no longer straight. Of course, Einstein didn't explain why mass should have this effect on the geometry of space. I suspect that, like Newton, he would have answered that he "feigned no hypothesis." His theory worked beautifully, explaining things Newton's couldn't, as observational tests concerned with solar system dynamics attested. And that was enough.

In 1917, less than two years after the publication of his general theory, Einstein wrote a remarkable paper, "Cosmological Considerations on the General Theory of Relativity." Like Newton, Einstein extrapolated the validity of his theory beyond the solar system, where it was tested at the time, to the universe as whole, and he proceeded to consider the shape of the entire cosmos. In true Platonist fashion, he wanted the cosmos to have the most perfect of shapes, that of a sphere. For convenience, and because of the lack of any opposing observation at the time, he also wanted the universe to be static. His equations produced the desired answer—a static and spherically symmetric Universe—but with a hidden surprise: in order to avoid the total collapse of matter to a central point (which would occur, just as Bentley had worried in Newton's case), Einstein refrained from making space infinite in extent. Instead, he introduced what he called a "universal constant" and added a new term to the equations describing the curvature of space, noting that, if sufficiently small, this constant was “also compatible with the facts of experience derived from the solar system.” This constant, “not justified by our actual knowledge of gravitation,” he conceded, is now called the “cosmological constant” and may indeed play a key role in the dynamics of the cosmos, although one quite different from that which Einstein had prescribed. Einstein needed it to ensure that his static spherical universe would not collapse onto itself. Displaying complete faith in his theory, he not only extrapolated his equations from the solar system to the entire universe but also imposed on his theory of the cosmos the effects of a strange repulsion whose job was to balance the cosmic dome.

To go beyond the known, both Newton and Einstein had to take intellectual risks, making assumptions based on intuition and personal prejudice. That they did so, knowing that their speculative theories were necessarily faulty and limited, illustrates the power of belief in the creative process of two of the greatest scientists of all time. To a greater or lesser extent, every person engaged in the advancement of knowledge does the same.

CHAPTER 2

BEYOND SPACE AND TIME

(Wherein we explore how different religions have faced the question of the origin of all things)

Backtrack ten thousand years, to just before the dawn of the first great civilizations along the Tigris and the Euphrates Rivers, where Iraq is now. To divinize Nature was an attempt to have a certain measure of control over what was uncontrollable. Floods, droughts, earthquakes, volcanoes, tidal waves—what even today insurance companies (shamelessly) call "acts of God"—were attributed to angry gods who needed to be placated. A language had to be developed, a common dialect between human and deity, enacted through ritualistic practices and mythic narratives, to bridge the enormous power imbalance between humans and the forces of Nature. As threats to survival came from everywhere—within the Earth, from its surface, and from the skies—gods had to be everywhere as well. Religion was born of necessity and reverence. Quite possibly, any thinking being with widespread but limited powers must assume the existence of other beings, be they gods or, more recently, aliens, with powers beyond his. The alternative, to leave natural disasters to chance, was just too scary to contemplate, as it would imply in accepting humankind's helplessness and utter loneliness in confronting the unknown. To have a fighting chance to control their destiny, humans had to believe.

Fear was not the only driving force toward belief, although it was possibly the main one. But not everything was bad. Good things also happened: a good crop, a productive hunt, quiet weather, bountiful oceans. Nature didn't only take away; it also gave plenty. In its dual role as giver and taker, it kept people alive, and it could kill them. Reflecting this polar tension, natural phenomena could be regular and safe—the day-night cycle, the seasons, the phases of the Moon, the tides—or irregular and fearsome, as in solar eclipses, comets, avalanches, and forest fires. It is then not surprising that regularity was (and is) associated with good and irregularity with bad: natural phenomena gained a moral dimension that, through the divinization of Nature, reflected directly the whims of intangible gods.

Across the world, ancient cultures erected monuments and temples to celebrate and to clock the regularity of the heavens. In England, Stonehenge's function as a burial place is probably related to the yearly alignment of the "Heel Stone" with the rising Sun during summer solstice, establishing a link between the periodic return of the Sun and the human's cycle of life and death. If the mechanisms behind the cyclic motions in the heavens were unknown—and there was no desire to "know" them, at least in the way we understand knowing now—they were still noted and in some cases measured with great care. Some three thousand years ago, the Babylonians, for example, had a well-established astronomical tradition, reflected in their creation myth, the Enuma Elish ("When Above"). They made detailed tables mapping the motions of the planets and the Moon across the sky, and registered any observed periodicities, as in the Ammisaduqa tablet, which recorded the risings and settings of Venus for twenty-one years.

There is comfort in repetition. If Nature beats to a drum, perhaps we do too. A cyclic time brings with it the promise of rebirth, establishing a deep connection between human and cosmos: our existence reflects that of the whole world. No wonder the myth of the eternal return resonates with so many cultures. What could be better than to believe that we return over and over again, that death is not the end but a transition to a new beginning?

As a father of five, I see this struggle with endings in all my children. My son Lucian, who was six when I wrote these lines, has been obsessed with death since he was four. Death sounds like an absurdity when time seems endless. “What happens after we die?” is one of those questions every parent hears, and most struggle to answer. Lucian is convinced that we return. He is just not quite sure if we return the same or as a different person. His choice, of course, is to come back the same, with the same parents and siblings, essentially reliving life twice or, better still, endlessly. What could be safer than not to have to face loss? It breaks my heart to have to tell him that what happens to us is the same thing that happens to the ant he crushes under his feet. He, of course, is not convinced. “How do you know, Dad?” “I don’t know for sure, son. Some people believe we do come back; others that we go to a place called Paradise, where we meet everybody else who has died. The problem is that I haven’t heard back from any of them to be sure that that’s where we are headed.” The conversation usually ends with a very tight hug and many utterings of “I love you.” What could be harder than to know that I cannot love him forever? And that one day, in the normal course of things, he will have to cope with my death?

With the advent of the Abrahamic faiths, a radically different way to think about the nature of time made a triumphal entrance: instead of ongoing cycles of creation and destruction, of life and death, time becomes linear, with a single beginning and an end. “Profane history,” as Eliade called it, is what happens between birth and death. The stakes suddenly became much higher, since with a single lifetime there is only one chance to be happy. For Christians and Muslims, the notion of an after-death Paradise comes to the rescue, and time begets a dual role, linear in life and inexistent in Paradise.

Linear or cyclic, time has always been a measure of transformation. Follow it to the future; it leads to endings. Follow it to the past; it leads to beginnings. In mythic narratives, humans are always subjected to the changes that time brings, while gods live outside time, never aging or getting sick. As life begets life, and generations succeed one another, following time backwards will necessarily lead to first life, the first living thing, be it bacteria, human, or beast. It is here the key question arises: How did the first living creature emerge, if there was nothing already living to give it birth? The same reasoning can be extrapolated to the world: How did the world come to be, if it had a beginning? The mythic answer, in the vast majority of cases, is clear: gods created the world first and then life. Only that which exists without time can *first* create that which exists within time. Although some creation myths, most notably among the Maori of New Zealand, suggest that the first creation could have happened without the interference of gods, in most myths time itself becomes a creation, starting once the world comes into being, as Saint Augustine cleverly proposed in *The Confessions* (Book 11, Chapter 13):

> Seeing then Thou art the Creator of all times, if any time was before Thou madest heaven and earth, why say they that Thou didst forego working? For that very time didst Thou make, nor could times pass by, before Thou madest those times. But if before heaven and earth there was no time, why is it demanded, what Thou then didst? For there was no “then,” when there was no time.

The origin of the world and the beginning of time are thus deeply enmeshed with the nature of the heavens, a connection that remains true in our time as modern cosmological models attempt to describe the origin of the Universe and astrophysicists study the origin of stars and planets. Not surprisingly, as I have examined in my book *The Dancing Universe*, both cyclic and linear notions of time reappeared in modern cosmology. More surprisingly, an essential characteristic of ancient creation myths—the deep relation between human and cosmos—also returns with current astronomical thought, after a long post-Copernican hiatus when our existence played second fiddle to the material splendor of the Universe. When Copernicus and, more pointedly, Johannes Kepler and Galileo

Galilei displaced Earth from the epicenter of Creation during the first decades of the seventeenth century, we lost our special status to become mere inhabitants of one among countless many worlds. Four hundred years later, as the ongoing search for life in the Universe reveals the fragility and relative scarcity of Earthlike planets, life and, more crucially, the uniqueness of human life is regaining its cosmic relevance: we matter because we are rare. The many steps from nonlife to life and then to complex multicellular life are hard to duplicate. Furthermore, the many particulars depend on our planet's detailed history. However, even with the current lack of evidence we cannot establish conclusively that other kinds of intelligent life don't exist in the Universe. They may or may not be out there. But what we *can* do is to state with confidence that *if* intelligent aliens exist, they are distant and rare. (Or, if ubiquitous, they certainly know how to hide extremely well, something we will get to at the end of this book.) In effect we are alone and must learn to live with our cosmic loneliness.

The urge to know our origins and our place in the cosmos is a defining part of our humanity. Creation myths of all ages ask questions not so different from those scientists ask today, when they ponder the quantum creation of the Universe "out of nothing," or whether our Universe is but one among countless others, all of them exhalations of a timeless multiverse. The specifics of the questions and of the answers are, of course, entirely different, but not the motivation: to understand where we came from and what our cosmic role is, if any. To the authors of those myths, ultimate questions of origins were solely answerable through invocations of the sacred, as only the timeless could create that which exists within time. To those who do not believe that answers to such questions remain exclusively within the realm of the sacred, the challenge is to scrutinize the reach of our rational explanations of the world and examine how far they can go in making sense of reality and, by extension, of ultimate questions of origins.

CHAPTER 3

TO BE, OR TO BECOME? THAT IS THE QUESTION

(Wherein we encounter the first philosophers of ancient Greece and delve into their remarkable notions about the meaning of reality)

A major shift in perspective happened sometime during the sixth and fifth centuries BCE in ancient Greece. Although influential new ideas concerning humankind's social and spiritual dimension were appearing elsewhere, such as with Confucius and Lao Tzu in China, and with Siddhartha Gautama, the Buddha, in India, it is to Greece that we turn to witness the birth of Western philosophy, a new mode of understanding based on questioning and argumentation, devised to examine the fundamental nature of knowledge and existence. In opposition to creation myths and sacred knowledge based on intangible revelation, the first Greek philosophers, known collectively as the Presocratics (since most lived before Socrates), sought to understand reality through logic and conjecture. This transition, in which tremendous faith is placed in the powers of reasoning to address the key questions of existence, redefined humanity's relation with the unknown from a passive reliance on fate and the supernatural to an active approach to knowledge and personal freedom.

Central to the Ionians, the first group of Presocratic thinkers, was a preoccupation with the material composition of the world. "What is the stuff that makes everything that is?" they asked. That this remains the defining question of modern particle physics serves to show that the value of a great question is that it keeps generating answers that, in turn, keep changing as our methods of inquiry change. Different members of the Ionian school suggested different answers, although all shared a fundamental characteristic: they believed that "All is One," that is, the material essence of reality is captured in a single substance or entity. This centralizing Oneness stands in sharp contrast with previous pantheistic mythologies, in which different gods are responsible for different parts of Nature. To the Ionians, everything that we witness is a manifestation of a single material essence undergoing various types of physical transformation.

And so Thales, whom none other than Aristotle considered to be the first philosopher, is said to have declared that "the principle of all things [is] water. For [Thales] says from water come all things and into water do all things decompose."[1] This quote, from the Byzantine physician Aëtius of Amida, is typical of thoughts attributed to Thales. Unfortunately, none of Thales's works survived, and we must rely on indirect sources to make sense of his ideas. When we read the literature, it becomes clear that Thales did propose water to be the source of everything, recognizing its role as the giver of life. To him, water symbolized the ongoing transformations of Nature, never at rest, even when apparently so. To explain the power source for these transformations, Thales invoked a soul-like force: "Some say that [soul] is mixed in the totality; this is perhaps the reason Thales thought all things are full of gods," wrote Aristotle in *On the Soul*.[2] These, however, are not the anthropomorphic gods of past mythologies, but the unexplained harmonious is a gateway to the inner workings of the psyche, Pythagoras and his followers found a bridge between the outside world and its perception through the senses. That this bridge was built with mathematical relations established the foundation to what would come next: to understand the world we must describe it mathematically. More than that, since what is harmonious is beautiful, the beauty of the world is expressible through mathematics. A new aesthetics emerges, equating mathematical law with beauty and beauty with truth.

Apart from pioneering the role of mathematics in the description of the world and of our interactions with it, the Pythagoreans had plenty to contribute to cosmology. The Pythagorean cosmos did more than displace the Earth from the center of the cosmos in favor of the "central fire." Extrapolating from harmony in music to the celestial spheres, the Pythagoreans believed that the distances between the planets stood in the same numerical ratios as those of the musical scales. As the planets revolved in the heavens they made music, "the harmony of the spheres," although according to legend only Pythagoras himself was able to hear it. The cosmic design, from the sensorial pleasure of music to the distances between planets, embodied harmoniously strict proportions: the beauty of creation was in its essence mathematical. Nothing could be more ennobling to the human soul than to decipher it.

Before we transition into Plato and his pupil Aristotle, we should briefly review where we are. On the one hand, we have the Ionians proposing that the essence of Nature is transformation and that all that exists is a manifestation of a single material essence. On the other, we have the Pythagoreans proposing that mathematics offers the key to Nature's mysteries and to our perception of reality. There were other voices. Also from Italy, Parmenides and the Eleatics countered the Ionians, proposing that what is essential is not that which can change but that which is unchangeable, or "what is." They proposed that change is an illusion, a distortion our imperfect senses cause in how we perceive reality. These are some of the earliest considerations in the West about the nature of reality and our perception of it: Is the essence of reality "in your face," the transformations we witness with our senses? Or is its essence hidden, locked in an abstract realm perceived only through thought?

To perceive change we need to sense it. But if our senses feed us only imperfect reconstructions of what exists, how can we grasp what is truly real? On the other hand, if we follow Parmenides, how can we possibly have any idea of this "thing" that doesn't change? After all, if something doesn't change, it becomes imperceptible to us, like a humming to which we grow deaf. Worse, if this unchangeable reality exists somehow in a rarefied realm, how are we to make sense of it? How can we probe it? And so the Ionians would accuse the Eleatics of empty abstractions, while the Eleatics would think the Ionians fools, as they trusted what could not be trusted. Meanwhile, the Pythagoreans would ignore both, following their belief in the power of mathematics to describe the harmony and beauty of the world.

Taken together, the plurality of Presocratic worldviews is staggering. The first Western philosophers were pushing the boundaries of knowledge in all directions, increasing the territory on which rational debate could ensue. Already twenty-five centuries ago conflicting ideas about the nature of reality abounded and clashed with one another. In their richness and complexity, they carried the kernel of the question that is still with us today and is the central theme of this book: To what extent can we make sense of reality? The Island was growing fast, and the shores facing the unknown offered endless possibilities.

CHAPTER 4

LESSONS FROM PLATO'S DREAM

(Wherein we explore how Plato and Aristotle dealt with the question of the First Cause and with the limits of knowledge)

Both Parmenides and the Pythagoreans deeply influenced Plato, who lived between circa 428 and 348 BCE. In a sense, Plato unified their modes of thinking, since, like Parmenides, he despised sensorial experience as a reliable source to attain the truth while, like Pythagoras, he embraced geometrical notions as the bridge between the human mind and the world of pure thought, where this elusive truth was to be found. Living in a time of political and social unrest, when Athens finally fell to Sparta at the end of the Peloponnesian War in 404 BCE, Plato searched for unchangeable truths as a path to stability and wisdom.

Nowhere is Plato's mode of thinking as clear as in his famous Allegory of the Cave, which appears in *The Republic*. The Allegory is also one of the first explicit meditations on the nature of reality. Imagine a group of people inside a cave, chained and unable to move since birth. Let's call them the Chained Ones. All the Chained Ones could do was to stare at the cave wall in front of them. They had no knowledge of the world outside or around them; their "reality" consisted of what they saw projected on the cave wall. They were unaware of the large fire burning behind them and of the small path and low-lying wall standing between them and the fire. Other persons could walk along the path and hold statues and various other objects near the fire. The Chained Ones saw the shadows of the statues and objects projected on the cave wall and took them for real. Their inability to turn back and understand their condition precluded them from seeing the truth. Theirs was a world of false illusions.

Plato proposed that even if a Chained One were freed and able to move toward the burning fire and the moving statues, the pain and temporary blindness from the bright flames would be such that he would quickly return to his previous spot in the cave. He would choose to believe the shadows to be truer than the blinding truth that was revealed to him: knowledge comes with a price that not all are willing to pay. To learn takes courage and tolerance, as it may cause an uncomfortable change of perspective. Plato insisted that if the freed Chained One had been dragged out of the cave and into the sunlight and hence even closer to the truth, he would beg to return to the shadows on the cave wall.

Plato compared the ascension of the Chained One toward the sunlight with the "mounting of the soul to the intellectual region," that is, with a literal enlightenment. He further suggested that the truth—a derivative of what he called "the essential Form of Good"—is hard to grasp, given that we are chained to our sensorial perception of reality. However, once we are ready to see (what we can of) it, the drive toward more knowledge is inevitable:

> My opinion is that in the world of knowledge the idea of good appears last of all, and is seen only with an effort; and, when seen, is also inferred to be the universal author of all things beautiful and right, parent of

light and of the lord of light in this visible world, and the immediate source of reason and truth in the intellectual; and that is the power upon which he who acts rationally either in public or private life must have his eyes fixed.[1]

Plato's main task in *The Republic* was to propose how a just and equitable society should be ruled and what kind of man should rule it. His answer, the philosopher-king, would ideally be someone who could probe into the abstract realm of pure forms, someone who would "set this Form of the Good before his eyes" and feed from the undying light of wisdom that forever shines there.

Plato's Forms are a source of much debate and confusion. Fortunately, we need not bother with the details. Suffice it to think of Forms as some kind of blueprint of perfection, as the core ideas behind things or feelings. For example, the Form of chairness contains in it the possibility of all chairs, which, once built, become mere shadows of the true Form. Forms are the universal essence of what is potentially existing, as they themselves are nonexistent in space and time. As limited beings, we can only dimly perceive what they comprise of, as we clumsily attempt to represent them within our perceived reality. Thus the idea of a circle, the one we sustain in our minds when prompted to think of a circle, is the only real circle: drawings or other forms of concrete representation are never the real thing, as they are never perfect.

In *Timaeus*, Plato extended these notions to cosmology. The Universe is the handiwork of a godlike entity called the demiurge, who uses Forms as a blueprint for his creation: the cosmos is spherical, and motions are circular and uniform, as those are the "most appropriate to mind and intelligence." Plato was proposing a cosmic aesthetics, the most symmetric and thus perfect shape being the only possible choice for celestial objects and their motions. Mind dictates the paths that matter follows: the world comes from an idea, and its physical structure must reflect this idea. This is a teleological cosmic view, a cosmoteleology, the notion that either the Universe has a purpose of its own or that it reflects the purpose of its Creator. It clashes frontally with the Atomistic notion of cosmic purposelessness, which, instead, proposes that nothing happens as the result of a preprescribed reason or intention: everything comes from atoms randomly moving and combining in the Void. As Lucretius wrote in *The Nature of Things*:

Especially since this world is the product of Nature,
  the happenstance
Of the seeds of things colliding into each other by pure chance
In every possible way, no aim in view, at random, blind,
Till sooner or later certain atoms suddenly combined
So that they lay the warp to weave the cloth of mighty things:
Of earth, of sea, of sky, of all species of living beings.[2]

Most of the philosophical discussions on the nature of the Universe after Plato, all the way to present-day ones involving the possibility of a multiverse that encompasses countless many universes including our own, or the possibility that our existence has some kind of cosmic purpose, reflect this charged ancient dichotomy.

From a scientific perspective, the main obstacle to any teleological explanation is our inability to determine whether it is right or wrong. The scientific method relies on empirical validation: a scientific hypothesis must be falsifiable; that is, scientists must be able to prove it wrong. If we can't, or better, while we can't—since every hypothesis must fail sooner or later—we accept it as true.[3] So if someone states that "the Universe has a purpose," we must first identify this purpose and then verify if indeed it is there. A popular contender is conscious life: "the Universe has the explicit intent of creating conscious life." A Creator Universe is not so different from a Creator God. It transforms a supernatural teleology into a teleology of the supranatural. A Universe with a scientifically justified intent is the modern answer to the pressure that the countless successes of science have imposed on revelation-based explanations: it gives purpose a scientific credibility. A Universe that intentionally engenders conscious beings reflects our ever-present need to be not just special creatures but special creations.

Attractive as it may be, the question of purpose has a serious challenge. How are we to test purpose in Nature? Unless we receive a very clear message from the perpetrator(s) stating their purpose and thus resolving the issue, this kind of naturalistic teleology becomes a categorical unknowable: if there is cosmic purpose and we cannot be aware of it, we cannot know whether it is not there either. We are ignorant either way, and all we can do is choose to believe in it or not, just as Plato believed in his demiurge but could not prove his existence.

When Plato's pupil Aristotle came into the game, his goal was diametrically opposite to that of his master. Aristotle was in many ways a pragmatist; he tried to build a towerlike structure of interlinked rational arguments to explain the way Nature works. This verticalism would, of course, be extremely attractive to the church later on, as it adopted Aristotle's cosmic arrangement as its own. Aristotle would posit that the natural bottom-up vertical arrangement of the four basic substances—earth, water, air, and fire—explains why an object made of one of them, when displaced from its medium, would naturally move back to it: an air bubble will float upward in water until it mixes with air, while a rock will sink to the bottom of a lake. He dismissed Plato's Forms and artisan demiurge as mere abstractions, suggesting instead that teleological drive was to be found within the objects themselves, in their "natures," which he took to be principles of change inherent in living things.

However, Aristotle wouldn't dismiss godlike presences in the world. Even though he would posit that the Universe was eternal and uncreated, he would invoke a detached kind of deity to be responsible for the motions of the heavens, the "unmoved movers." These immaterial godlike entities had as their mission to impart the motions observed in the skies without themselves moving or either suffering or imposing any material cause. Mysteriously, they would impart motion through some sort of "aspiration or desire." Since Aristotle's cosmos was a complex onion-like contraption of spheres within spheres, with Earth static at the center and the sphere carrying the stars at the periphery, there was a vertical hierarchy within the unmoved movers, with the outermost one being the First Mover. Its role was to impart motion from the outside in, a kind of Cosmic Winder responsible for initiating the causal chain that animated the Universe.[4]

With his First Mover and subordinate collective of unmoved movers, Aristotle addressed two fundamental issues one faces when trying to explain Nature: how the change from rest into motion occurs, and how motion is maintained. How else could he explain what kept his huge cosmic machine in motion for all eternity? Aristotle missed the notion of inertia, the natural condition of a body to remain in its state of motion unless compelled to change it by a force. For that, some eighteen centuries had to pass.

Being eternal in time, Aristotle's cosmos had an added simplicity over a cosmos that appeared at some moment in the past, as in the biblical narrative or, more to the point, the modern Big Bang cosmology. As we remarked before, a Universe with a beginning needs a causal explanation to its appearance. Why should there be a Universe in the first place? What caused it to become? Religions resolve the issue by imposing the existence of a godlike First Cause that exists beyond the constraints of physical laws. To explain the emergence of the physical Universe through science is a tremendous conceptual challenge, one that still haunts modern cosmologists, even if many insist that quantum mechanics can take care of it all, which, as we shall argue later, is bad philosophy and scientifically fallacious. To claim that we know how the Universe emerged is both untrue and a great disservice to the public understanding of science. Like it or not, there is a horizon around every island. The Island of Knowledge is no exception.

Back to Aristotle. We see that he drove philosophy out of Plato's cave, dissolving the distinction between a world of abstract Forms and a world of sensorial perception. Change on Earth and its immediate surroundings is to be understood through the interplay of the four basic substances. As we ascend to the skies, we transition into the realm of the heavenly spheres, responsible for carrying the Moon and five planets in their circular orbits. (Mercury, Venus, Mars, Jupiter, and Saturn were the only planets known until the discovery of Uranus in 1781.) Heavenly objects, however, were made of a fifth essence, the perfect and eternal aether, immune to any kind of change. Aristotle's cosmos still retained a dualistic nature, a division between the terrestrial world of ordinary matter and the ethereal worlds above, inaccessible to matter and perfect. It also retained a version of an unknowable teleology, incorporated into the immaterial but active unmoved movers, a shoo-in for medieval Christian theology.