The TROUBLE with PHYSICS

The Rise of String Theory, the Fall of a Science, and What Comes Next

Lee Smolin

HOUGHTON MIFFLIN COMPANY
BOSTON • NEW YORK 2006

Introduction

There may or may not be a God. Or gods. Yet there is something ennobling about our search for the divine. And also something humanizing, which is reflected in each of the paths people have discovered to take us to deeper levels of truth. Some seek transcendence in meditation or prayer; others seek it in service to their fellow human beings; still others, the ones lucky enough to have the talent, seek transcendence in the practice of an art.

Another way of engaging life's deepest questions is science. Not that every scientist is a seeker; most are not. But within every scientific discipline, there are those driven by a passion to know what is most essentially true about their subject. If they are mathematicians, they want to know what numbers are, or what kind of truth mathematics describes. If they are biologists, they want to know what life is, and how it started. If they are physicists, they want to know about space and time, and what brought the world into existence. These fundamental questions are the hardest to answer and progress is seldom direct. Only a handful of scientists have the patience for this work. It is the riskiest kind of work, but the most rewarding: When someone answers a question about the foundations of a subject, it can change everything we know.

Because it is their job to add to our growing store of knowledge, scientists spend their days confronting what they don't understand.

And those scientists who work on the foundations of any given field are fully aware that the building blocks are never as solid as their colleagues tend to believe.

This is the story of a quest to understand nature at its deepest level. Its protagonists are the scientists who are laboring to extend our knowledge of the basic laws of physics. The period of time I will address — roughly since 1975 — is the span of my own professional career as a theoretical physicist. It may also be the strangest and most frustrating period in the history of physics since Kepler and Galileo began the practice of our craft four hundred years ago.

The story I will tell could be read by some as a tragedy. To put it bluntly — and to give away the punch line — we have failed. We inherited a science, physics, that had been progressing so fast for so long that it was often taken as the model for how other kinds of science should be done. For more than two centuries, until the present period, our understanding of the laws of nature expanded rapidly. But today, despite our best efforts, what we know for certain about these laws is no more than what we knew back in the 1970s.

How unusual is it for three decades to pass without major progress in fundamental physics? Even if we look back more than two hundred years, to a time when science was the concern mostly of wealthy amateurs, it is unprecedented. Since at least the late eighteenth century, significant progress has been made on crucial questions every quarter century.

By 1780, when Antoine Lavoisier's quantitative chemistry experiments were showing that matter is conserved, Isaac Newton's laws of motion and gravity had been in place for almost a hundred years. But while Newton gave us a framework for understanding all of nature, the frontier was wide open. People were just beginning to learn the basic facts about matter, light, and heat, and mysterious phenomena like electricity and magnetism were being elucidated.

Over the next twenty-five years, major discoveries were made in each of these areas. We began to understand that light is a wave. We discovered the law that governs the force between electrically charged particles. And we made huge leaps in our understanding of matter with John Dalton's atomic theory. The notion of energy was introduced; interference and diffraction were explained in terms of

the wave theory of light; electrical resistance and the relationship between electricity and magnetism were explored.

Several basic concepts underlying modern physics emerged in the next quarter century, from 1830 to 1855. Michael Faraday introduced the notion that forces are conveyed by fields, an idea he used to greatly advance our understanding of electricity and magnetism. During the same period, the conservation of energy was proposed, as was the second law of thermodynamics.

In the quarter century following that, Faraday's pioneering ideas about fields were developed by James Clerk Maxwell into our modern theory of electromagnetism. Maxwell not only unified electricity and magnetism, he explained light as an electromagnetic wave. In 1867, he explained the behavior of gases in terms of the atomic theory. During the same period, Rudolf Clausius introduced the notion of entropy.

The period from 1880 to 1905 saw the discoveries of electrons and X rays. The study of heat radiation was developed in several steps, leading to Max Planck's discovery, in 1900, of the right formula to describe the thermal properties of radiation — a formula that would spark the quantum revolution.

In 1905, Albert Einstein was twenty-six. He had failed to find an academic job in spite of the fact that his early work on the physics of heat radiation alone would come to be seen as a major contribution to science. But that was just a warm-up. He soon zeroed in on the fundamental questions of physics: First, how could the relativity of motion be reconciled with Maxwell's laws of electricity and magnetism? He told us in his special theory of relativity. Should we think of the chemical elements as Newtonian atoms? Einstein proved we must. How can we reconcile the theories of light with the existence of atoms? Einstein told us how, and in the process showed that light is both a wave and a particle. All in the year 1905, in time stolen from his work as a patent examiner.

The working out of Einstein's insights took the next quarter century. By 1930, we had his general theory of relativity, which makes the revolutionary claim that the geometry of space is not fixed but evolves in time. The wave-particle duality uncovered by Einstein in 1905 had become a fully realized quantum theory, which gave us a detailed understanding of atoms, chemistry, matter, and radiation. By 1930 we also knew that the universe contained huge numbers of galaxies like our own, and we knew they were moving away from one another. The implications were not yet clear, but we knew we lived in an expanding universe.

With the establishment of quantum theory and general relativity as part of our understanding of the world, the first stage in the twentieth-century revolution in physics was over. Many physics professors, uncomfortable with revolutions in their areas of expertise, were relieved that we could go back to doing science the normal way, without having to question our basic assumptions at every turn. But their relief was premature.

Einstein died at the end of the next quarter century, in 1955. By then we had learned how to consistently combine quantum theory with the special theory of relativity; this was the great accomplishment of the generation of Freeman Dyson and Richard Feynman. We had discovered the neutron and the neutrino and hundreds of other apparently elementary particles. We had also understood that the myriad phenomena in nature are governed by just four forces: electromagnetism, gravity, the strong nuclear force (which holds atomic nuclei together), and the weak nuclear force (responsible for radioactive decay).

Another quarter century brings us to 1980. By then we had constructed a theory explaining the results of all our experiments on the elementary particles and forces to date — a theory called the standard model of elementary-particle physics. For example, the standard model told us precisely how protons and neutrons are made up of quarks, which are held together by gluons, the carriers of the strong nuclear force. For the first time in the history of fundamental physics, theory had caught up with experiment. No one has since done an experiment that was not consistent with this model or with general relativity.

Going from the very small to the very large, our knowledge of physics now extended to the new science of cosmology, where the Big Bang theory had become the consensus view. We realized that our universe contains not only stars and galaxies but exotic objects such as neutron stars, quasars, supernovas, and black holes. By 1980, Stephen Hawking had already made the fantastic prediction

that black holes radiate. Astronomers also had evidence that the universe contains a lot of dark matter — that is, matter in a form that neither emits nor reflects light.

In 1981, the cosmologist Alan Guth proposed a scenario for the very early history of the universe called inflation. Roughly speaking, his theory asserts that the universe went through a spurt of enormous growth extremely early in its life, and it explains why the universe looks pretty much the same in every direction. The theory of inflation made predictions that seemed dubious, until the evidence began to swing toward them a decade ago. As of this writing, a few puzzles remain, but the bulk of the evidence supports the predictions of inflation.

Thus, by 1981, physics had enjoyed two hundred years of explosive growth. Discovery after discovery deepened our understanding of nature, because in each case theory and experiment had marched hand in hand. New ideas were tested and confirmed and new experimental discoveries were explained in terms of theory. Then, in the early 1980s, things ground to a halt.

I am a member of the first generation of physicists educated since the standard model of particle physics was established. When I meet old friends from college and graduate school, we sometimes ask each other, "What have we discovered that our generation can be proud of?" If we mean new fundamental discoveries, established by experiment and explained by theory — discoveries on the scale of those just mentioned — the answer, we have to admit, is *"Nothing!"* Mark Wise is a leading theorist working on particle physics beyond the standard model. At a recent seminar at the Perimeter Institute of Theoretical Physics, in Waterloo, Ontario, where I work, he talked about the problem of where the masses of the elementary particles come from. "We've been remarkably unsuccessful at solving that problem," he said. "If I had to give a talk on the fermion-mass problem now, I'd probably end up talking about things I could have in the 1980s."[1] He went on to tell a story about when he and John Preskill, another leading theorist, arrived at Caltech in 1983, to join its faculty. "John Preskill and I were sitting together in his office, talking. . . . You know, the gods of physics were at Caltech, and now we were there! John said, 'I'm not going to forget what is important to work on.' So he took what was known about the quark and lepton masses, and he wrote it on a yellow sheet of paper and stuck it on his bulletin board . . . so as not to forget to work on them. Fifteen years later, I come into his office . . . and we're talking about something, and I look up at his bulletin board and [notice that] that sheet of paper is still there but the sun has faded everything that was written on it. So the problems went away!"

To be fair, we've made two experimental discoveries in the past few decades: that neutrinos have mass and that the universe is dominated by a mysterious dark energy that seems to be accelerating its expansion. But we have no idea *why* neutrinos (or any of the other particles) have mass or what explains their mass value. As for the dark energy, it's not explained in terms of any existing theory. Its discovery cannot then be counted as a success, for it suggests that there is some major fact we are all missing. And except for the dark energy, no new particle has been discovered, no new force found, no new phenomenon encountered that was not known and understood twenty-five years ago.

Don't get me wrong. For the past twenty-five years we have certainly been very busy. There has been enormous progress in applying established theories to diverse subjects: the properties of materials, the molecular physics underlying biology, the dynamics of vast clusters of stars. But when it comes to extending our knowledge of the laws of nature, we have made no real headway. Many beautiful ideas have been explored, and there have been remarkable particle-accelerator experiments and cosmological observations, but these have mainly served to confirm existing theory. There have been few leaps forward, and none as definitive or important as those of the previous two hundred years. When something like this happens in sports or business, it's called hitting the wall.

Why is physics suddenly in trouble? And what can we do about it? These are the central questions of my book.

I'm an optimist by nature, and for a long time I fought the conclusion that this period in physics — the period of my own career — has been an unusually fallow one. For me and many of my friends who entered science with the hope of making important contributions to what then was a rapidly moving field, there is a shocking fact we must come to terms with: Unlike any previous generation,

we have not achieved anything that we can be confident will outlive us. This has given rise to personal crises. But, more important, it has produced a crisis in physics.

The main challenge for theoretical particle physics over the last three decades has been to explain the standard model more deeply. Here there has been a lot of activity. New theories have been posited and explored, some in great detail, but none has been confirmed experimentally. And here's the crux of the problem: In science, for a theory to be believed, it must make a new prediction — different from those made by previous theories — for an experiment not yet done. For the experiment to be meaningful, we must be able to get an answer that disagrees with that prediction. When this is the case, we say that a theory is *falsifiable* — vulnerable to being shown false. The theory also has to be *confirmable;* it must be possible to verify a new prediction that only this theory makes. Only when a theory has been tested and the results agree with the theory do we advance the theory to the ranks of true theories.

The current crisis in particle physics springs from the fact that the theories that have gone beyond the standard model in the last thirty years fall into two categories. Some were falsifiable, and they were falsified. The rest are untested — either because they make no clean predictions or because the predictions they do make are not testable with current technology.

Over the last three decades, theorists have proposed at least a dozen new approaches. Each approach is motivated by a compelling hypothesis, but none has so far succeeded. In the realm of particle physics, these include Technicolor, preon models, and supersymmetry. In the realm of spacetime, they include twistor theory, causal sets, supergravity, dynamical triangulations, and loop quantum gravity. Some of these ideas are as exotic as they sound.

One theory has attracted more attention than all the others combined: string theory. The reasons for its popularity are not hard to understand. It purports to correctly describe the big and the small — both gravity and the elementary particles — and to do so, it makes the boldest hypotheses of all the theories: It posits that the world contains as yet unseen dimensions and many more particles than are presently known. At the same time, it proposes that all the elementary particles arise from the vibrations of a single entity — a string — that obeys simple and beautiful laws. It claims to be the one theory that unifies *all* the particles and *all* the forces in nature. As such, it promises to make clean and unambiguous predictions for any experiment that has ever been done or ever could be done. Much effort has been put into string theory in the last twenty years, but we still do not know whether it is true. Even after all this work, the theory makes no new predictions that are testable by current — or even currently conceivable — experiments. The few clean predictions it does make have already been made by other well-accepted theories.

Part of the reason string theory makes no new predictions is that it appears to come in an infinite number of versions. Even if we restrict ourselves to theories that agree with some basic observed facts about our universe, such as its vast size and the existence of the dark energy, we are left with as many as 10^{500} distinct string theories — that's 1 with 500 zeros after it, more than all the atoms in the known universe. With such a vast number of theories, there is little hope that we can identify an outcome of an experiment that would not be encompassed by one of them. Thus, no matter what the experiments show, string theory cannot be disproved. But the reverse also holds: No experiment will ever be able to prove it true.

At the same time, we understand very little about most of these string theories. And of the small number we do understand in any detail, every single one disagrees with the present experimental data, usually in at least two ways.

So we face a paradox. Those string theories we know how to study are known to be wrong. Those we cannot study are thought to exist in such vast numbers that no conceivable experiment could ever disagree with all of them.

These are not the only problems. String theory rests on several key conjectures, for which there is some evidence but no proof. Even worse, after all the scientific labor expended in its study, we still do not know whether there is a complete and coherent theory that can even go by the name "string theory." What we have, in fact, is not a theory at all but a large collection of approximate calculations, together with a web of conjectures that, if true, point to the existence of a theory. But that theory has never actually been written down. We don't know what its fundamental principles are. We don't know what mathematical language it should be expressed in — perhaps a new one will have to be invented to describe it. Lacking both fundamental principles and the mathematical formulation, we cannot say that we even know what string theory asserts.

Here is how the string theorist Brian Greene puts it in his latest book, *The Fabric of the Cosmos:* "Even today, more than three decades after its initial articulation, most string practitioners believe we still don't have a comprehensive answer to the rudimentary question, What is string theory? . . . [M]ost researchers feel that our current formulation of string theory still lacks the kind of core principle we find at the heart of other major advances."[2]

Gerard 't Hooft, a Nobel Prize winner for his work in elementary-particle physics, has characterized the state of string theory this way: "Actually, I would not even be prepared to call string theory a 'theory,' rather a 'model,' or not even that: just a hunch. After all, a theory should come with instructions on how to deal with it to identify the things one wishes to describe, in our case the elementary particles, and one should, at least in principle, be able to formulate the rules for calculating the properties of these particles, and how to make new predictions for them. Imagine that I give you a chair, while explaining that the legs are still missing, and that the seat, back and armrest will perhaps be delivered soon. Whatever I did give you, can I still call it a chair?"[3]

David Gross, a Nobel laureate for his work on the standard model, has since become one of the most aggressive and formidable champions of string theory. Yet he closed a recent conference intended to celebrate the theory's progress by saying, "We don't know what we are talking about. . . . The state of physics today is like it was when we were mystified by radioactivity. . . . They were missing something absolutely fundamental. We are missing perhaps something as profound as they were back then."[4]

But though string theory is so incomplete that its very existence is an unproved conjecture, that does not keep many who work on it from believing that it is the only way forward for theoretical physics. One prominent string theorist, Joseph Polchinski, of the Kavli Institute for Theoretical Physics at UC Santa Barbara, was asked not long ago to give a talk on "Alternatives to String Theory." His first reaction, he said, "was that this was silly, there are no alternatives. . . . All good ideas are part of string theory."[5] Lubos Motl, an assistant professor at Harvard, recently asserted on his blog that "the most likely reason why no . . . person has convinced others about [an] alternative to string theory is that there probably exists no alternative to string theory."[6]

What is going on here? Usually in science one means something quite definite by the term *theory.* Lisa Randall, an influential particle theorist and Motl's colleague at Harvard, defines a theory as "a definite physical framework embodied in a set of fundamental assumptions about the world — and an economical framework that encompasses a wide variety of phenomena. A theory yields a specific set of equations and predictions — ones that are borne out by successful agreement with experimental results."[7]

String theory does not fit this description — at least not yet. How, then, are some experts sure there is no alternative to string theory, if they don't know precisely what it is? What exactly is it that they are sure has no alternative? These are some of the questions that led me to write this book.

Theoretical physics is hard. Very hard. Not because a certain amount of math is involved but because it involves great risks. As we will see over and over again as we examine the story of contemporary physics, science of this kind cannot be done without risk. If a large number of people have worked on a question for many years and the answer remains unknown, it may mean that the answer is not easy or obvious. Or this may be a question that has no answer.

String theory, to the extent it is understood, posits that the world is fundamentally different from the world we know. If string theory is right, the world has more dimensions and many more particles and forces than we have so far observed. Many string theorists talk and write as if the existence of those extra dimensions and particles were an assured fact, one that no good scientist can doubt. More than once, a string theorist has said to me something like "But do you mean you think it's possible that there are *not* extra dimensions?" In fact, neither theory nor experiment offers any evidence at all that extra dimensions exist. One of the goals of this book is to demystify the claims of string theory. The ideas are beautiful and well motivated. But to understand why they have not led to greater

progress, we have to be clear about exactly what the evidence supports and what is still missing.

Because string theory is such a high-risk venture — unsupported by experiment, though very generously supported by the academic and scientific communities — there are only two ways the story can end. If string theory turns out to be right, string theorists will turn out to be the greatest heroes in the history of science. On the basis of a handful of clues — none of which has an unambiguous reading — they will have discovered that reality is far more vast than previously imagined. Columbus discovered a new continent unknown to the king and queen of Spain (as the Spanish royals were unknown to the residents of the New World). Galileo discovered new stars and moons, and later astronomers discovered new planets. All this would pale in the face of the discovery of new dimensions. Moreover, many string theorists believe that the myriad worlds described by the huge number of string theories really do exist — as other universes impossible for us to see directly. If they are right, we see far less of reality than any group of cave dwellers saw of the earth. No one in human history has ever guessed correctly about such a large expansion of the known world.

On the other hand, if string theorists are wrong, they can't be just a little wrong. If the new dimensions and symmetries do not exist, then we will count string theorists among science's greatest failures, like those who continued to work on Ptolemaic epicycles while Kepler and Galileo forged ahead. Theirs will be a cautionary tale of how not to do science, how not to let theoretical conjecture get so far beyond the limits of what can rationally be argued that one starts engaging in fantasy.

One result of the rise of string theory is that the community of people who work on fundamental physics is split. Many scientists continue to work on string theory, and perhaps as many as fifty new PhDs are awarded each year for work in this field. But there are some physicists who are deeply skeptical — who either never saw the point or have by now given up waiting for a sign that the theory has a consistent formulation or makes a real experimental prediction. The split is not always friendly. Doubts are expressed on each side about the professional competence and ethical standards of the other, and it is real work maintaining friendships across the divide.

According to the picture of science we all learned in school, situations like this are not supposed to develop. The whole point of modern science, we are taught, is that there is a method that leads to progress in our understanding of nature. Disagreement and controversy are of course necessary for science to progress, but there is always supposed to be a way to resolve a dispute by means of experiment or mathematics. In the case of string theory, however, this mechanism seems to have broken down. Many adherents and critics of string theory are so confirmed in their views that it is difficult to have a cordial discussion on the issue, even among friends. "How can you not see the beauty of the theory? How could a theory do all this and not be true?" say the string theorists. This provokes an equally heated response from skeptics: "Have you lost your mind? How can you believe so strongly in *any* theory in the complete absence of experimental test? Have you forgotten how science is supposed to work? How can you be so sure you are right when you do not even know what the theory is?"

I have written this book in the hope that it will contribute to an honest and useful discussion among experts and lay readers alike. In spite of what I have seen in the last few years, I believe in science. I believe in the ability of the scientific community to rise above acrimony and resolve controversy through rational argument based on the evidence in front of us. I am aware that just by raising these issues, I will anger some of my friends and colleagues who work on string theory. I can only insist that I am writing this book not to attack string theory or those who believe in it but out of admiration for them and, above all, as an expression of faith in the physics scientific community.

So this is not a book about "us" versus "them." During my career, I have worked on both string theory and on other approaches to quantum gravity (the reconciliation of Einstein's general theory of relativity with quantum theory). Even if most of my efforts have gone into these other approaches, there have been periods when I avidly believed in string theory and devoted myself to solving its key problems. While I didn't solve them, I wrote eighteen papers in the subject; thus, the mistakes I will discuss are my mistakes as much as anyone else's. I will speak of conjectures that were widely

believed to be true, in spite of never having been proved. But I was among the believers, and I made choices about my research based on those beliefs. I will speak of the pressures that young scientists feel to pursue topics sanctioned by the mainstream in order to have a decent career. I have felt those pressures myself, and there were times when I let my career be guided by them. The conflict between the need to make scientific judgments independently and make them in a way that doesn't alienate you from the mainstream is one that I, too, have experienced. I write this book not to criticize scientists who have made choices different from mine but to examine why scientists need to be confronted with such choices at all.

In fact, it took me a long time to decide to write this book. I personally dislike conflict and confrontation. After all, in the kind of science we do, anything worth doing is a risk and all that really matters is what our students' students will think worthy of teaching their own students fifty years down the road. I kept hoping someone in the center of string-theory research would write an objective and detailed critique of exactly what has and has not been achieved by the theory. That hasn't happened.

One reason to take these issues public goes back to the debate that took place a few years ago between scientists and "social constructivists," a group of humanities and social science professors, over how science works. The social constructivists claimed that the scientific community is no more rational or objective than any other community of human beings. This is not how most scientists view science. We tell our students that belief in a scientific theory must always be based on an objective evaluation of the evidence. Our opponents in the debate argued that our claims about how science works were mainly propaganda designed to intimidate people into giving us power, and that the whole scientific enterprise was driven by the same political and sociological forces that drove people in other fields.

One of the main arguments we scientists used in that debate was that our community was different because we governed ourselves according to high standards — standards that prevented us from embracing any theory until it had been proved, by means of published calculations and experimental data, beyond the doubt of a competent professional. As I will relate in some detail, this is not always the case in string theory. Despite the absence of experimental support and precise formulation, the theory is believed by some of its adherents with a certainty that seems emotional rather than rational.

The aggressive promotion of string theory has led to its becoming the primary avenue for exploring the big questions in physics. Nearly every particle theorist with a permanent position at the prestigious Institute for Advanced Study, including the director, is a string theorist; the exception is a person hired decades ago. The same is true of the Kavli Institute for Theoretical Physics. Eight of the nine MacArthur Fellowships awarded to particle physicists since the beginning of the program in 1981 have also gone to string theorists. And in the country's top physics departments (Berkeley, Caltech, Harvard, MIT, Princeton, and Stanford), twenty out of the twenty-two tenured professors in particle physics who received PhDs after 1981 made their reputation in string theory or related approaches.

String theory now has such a dominant position in the academy that it is practically career suicide for young theoretical physicists not to join the field. Even in areas where string theory makes no predictions, like cosmology and particle phenomenology, it is common for researchers to begin talks and papers by asserting a belief that their work will be derivable from string theory sometime in the future.

There are good reasons to take string theory seriously as a hypothesis about nature, but this is not the same as declaring its truth. I invested several years of work in string theory because I believed in it enough to want to try my hand at solving its key problems. I also believed that I had no right to an opinion until I knew it in detail, as only a practitioner could. At the same time, I have worked on other approaches that also promise to answer fundamental questions. As a result, I'm regarded with some suspicion by people on both sides of the debate. Some string theorists consider me "anti-string." This couldn't be less true. I would never have put so much time and effort into working on string theory, or written three books largely motivated by its problems, if I wasn't fascinated by it and didn't feel that it might turn out to be part of the truth. Nor am I *for* anything except science, or *against* anything except that which threatens science.

But there's more at stake than amity among colleagues. To do our work, we physicists require significant resources, which are provided largely by our fellow citizens — through taxes as well as foundation money. In exchange, they ask only for the chance to look over our shoulders as we forge ahead and deepen humanity's knowledge of the world we share. Those physicists who communicate with the public, whether through writing, public speaking, television, or the Internet, have a responsibility to tell the story straight. We must be careful to present the failures along with the successes. Indeed, being honest about failures is likely to help rather than hurt our cause. After all, the people who support us live in the real world. They know that progress in any endeavor requires that real risks be taken, that sometimes you will fail.

In recent years, many books and magazine articles for the general public have described the amazing new ideas that theoretical physicists have been working on. Some of these chronicles have been less than careful about explaining just how far the new ideas are from both experimental test and mathematical proof. Having benefited from the public's desire to know how the universe works, I feel a responsibility to make sure that the story told in this book sticks close to the facts. I hope to lay out the various problems we have been unable to solve, explain clearly what experiment supports and doesn't support, and distinguish fact from speculation and intellectual fad.

Above all, we physicists have a responsibility to the future of our craft. Science, as I shall argue later, is based on an ethic, and that ethic requires good faith on the part of its practitioners. It also requires that each scientist be the judge of what he or she believes, so that every unproved idea is met with a healthy dose of skepticism and criticism until it is proved. This, in turn, requires that a diversity of approaches to unsolved problems be supported and welcomed into the community of science. We do research because even the smartest among us doesn't know the answer. Often it lies in a direction other than the one pursued by the mainstream. In those cases, and even when the mainstream guesses right, the progress of science depends on healthy support for scientists who hold divergent views.

Science requires a delicate balance between conformity and variety. Because it is so easy to fool ourselves, because the answers are unknown, experts, no matter how well trained or smart, will disagree about which approach is most likely to yield fruit. Therefore, if science is to move forward, the scientific community must support a variety of approaches to any one problem.

There is ample evidence that these basic principles are no longer being followed in the case of fundamental physics. While few would disagree with the rhetoric of diverse views, it is being practiced less and less. Some young string theorists have told me that they feel constrained to work on string theory whether or not they believe in it, because it is perceived as the ticket to a professorship at a university. And they are right: In the United States, theorists who pursue approaches to fundamental physics other than string theory have almost no career opportunities. In the last fifteen years, there have been a total of three assistant professors appointed to American research universities who work on approaches to quantum gravity other than string theory, and these appointments were all to a single research group. Even as string theory struggles on the scientific side, it has triumphed within the academy.

This hurts science, because it chokes off the investigation of alternative directions, some of them very promising. Despite the inadequate investment in these approaches, a few have moved ahead of string theory to the point of suggesting definite predictions for experiments, which are now in progress.

How is it possible that string theory, which has been pursued by more than a thousand of the brightest and best-educated scientists, working in the best conditions, is in danger of failing? This has puzzled me for a long time, but now I think I know the answer. What I believe is failing is not so much a particular theory but a style of doing science that was well suited to the problems we faced in the middle part of the twentieth century but is ill suited to the kinds of fundamental problems we face now. The standard model of particle physics was the triumph of a particular way of doing science that came to dominate physics in the 1940s. This style is pragmatic and hard-nosed and favors virtuosity in calculating over reflection

on hard conceptual problems. This is profoundly different from the way that Albert Einstein, Niels Bohr, Werner Heisenberg, Erwin Schrödinger, and the other early-twentieth-century revolutionaries did science. Their work arose from deep thought on the most basic questions surrounding space, time, and matter, and they saw what they did as part of a broader philosophical tradition, in which they were at home.

In the approach to particle physics developed and taught by Richard Feynman, Freeman Dyson, and others, reflection on foundational problems had no place in research. This freed them from the debates over the meaning of quantum physics that their elders were embroiled in and led to thirty years of dramatic progress. This is as it should be: Different styles of research are needed to solve different kinds of problems. Working out the applications of established frameworks requires very different kinds of thinking — and thinkers — than inventing those frameworks in the first place.

However, as I will argue in detail in the pages to come, the lesson of the last thirty years is that the problems we're up against today cannot be solved by this pragmatic way of doing science. To continue the progress of science, we have to again confront deep questions about space and time, quantum theory, and cosmology. We again need the kinds of people who can invent new solutions to long-standing foundational problems. As we shall see, the directions in which progress is being made — which are taking theory back into contact with experiment — are led by people who have an easier time inventing new ideas than following popular trends and for the most part do science in the reflective and foundational style of the early-twentieth-century pioneers.

I want to emphasize that my concern is not with string theorists as individuals, some of whom are the most talented and accomplished physicists I know. I would be the first to defend their right to pursue the research they think is most promising. But I am extremely concerned about a trend in which only one direction of research is well supported while other promising approaches are starved.

It is a trend with tragic consequences if, as I will argue, the truth lies in a direction that requires a radical rethinking of our basic ideas about space, time, and the quantum world.

1

The Five Great Problems in Theoretical Physics

FROM THE BEGINNING of physics, there have been those who imagined they would be the last generation to face the unknown. Physics has always seemed to its practitioners to be almost complete. This complacency is shattered only during revolutions, when honest people are forced to admit that they don't know the basics. But even revolutionaries still imagine that the big idea — the one that will tie it all up and end the search for knowledge — lies just around the corner.

We live in one of those revolutionary periods, and have for a century. The last such period was the Copernican revolution, beginning in the early sixteenth century, during which Aristotelian theories of space, time, motion, and cosmology were overthrown. The culmination of that revolution was Isaac Newton's proposal of a new theory of physics, published in his *Philosophiae Naturalis Principia Mathematica* in 1687. The current revolution in physics began in 1900, with Max Planck's discovery of a formula describing the energy distribution in the spectrum of heat radiation, which demonstrated that the energy is not continuous but quantized. This revolution has yet to end. The problems that physicists must solve today are, to a large extent, questions that remain unanswered because of the incompleteness of the twentieth century's scientific revolution.

The core of our failure to complete the present scientific revolution consists of five problems, each famously intractable. These problems confronted us when I began my study of physics in the 1970s, and while we have learned a lot about them in the last three decades, they remain unsolved. One way or another, any proposed theory of fundamental physics must solve these five problems, so it's worth taking a closer look at each.

Albert Einstein was certainly the most important physicist of the twentieth century. Perhaps his greatest work was his discovery of general relativity, which is the best theory we have so far of space, time, motion, and gravitation. His profound insight was that gravity and motion are intimately related to each other and to the geometry of space and time. This idea broke with hundreds of years of thinking about the nature of space and time, which until then had been viewed as fixed and absolute. Being eternal and unchanging, they provided a background, which we used to define notions like position and energy.

In Einstein's general theory of relativity, space and time no longer provide a fixed, absolute background. Space is as dynamic as matter; it moves and morphs. As a result, the whole universe can expand or shrink, and time can even begin (in a Big Bang) and end (in a black hole).

Einstein accomplished something else as well. He was the first person to understand the need for a new theory of matter and radiation. Actually, the need for a break was implicit in Planck's formula, but Planck had not understood its implications deeply enough; he felt that it could be reconciled with Newtonian physics. Einstein thought otherwise, and he gave the first definitive argument for such a theory in 1905. It took twenty more years to invent that theory, known as the quantum theory.

These two discoveries, of relativity and of the quantum, each required us to break definitively with Newtonian physics. However, in spite of great progress over the century, they remain incomplete. Each has defects that point to the existence of a deeper theory. But the main reason each is incomplete is the existence of the other.

The mind calls out for a third theory to unify all of physics, and for a simple reason. Nature is in an obvious sense "unified." The universe we find ourselves in is interconnected, in that everything interacts with everything else. There is no way we can have two

theories of nature covering different phenomena, as if one had nothing to do with the other. Any claim for a final theory must be a complete theory of nature. It must encompass all we know.

Physics has survived a long time without that unified theory. The reason is that, as far as experiment is concerned, we have been able to divide the world into two realms. In the atomic realm, where quantum physics reigns, we can usually ignore gravity. We can treat space and time much as Newton did — as an unchanging background. The other realm is that of gravitation and cosmology. In that world, we can often ignore quantum phenomena.

But this cannot be anything other than a temporary, provisional solution. To go beyond it is the first great unsolved problem in theoretical physics:

Problem 1: Combine general relativity and quantum theory into a single theory that can claim to be the complete theory of nature.

This is called the *problem of quantum gravity.*

Besides the argument based on the unity of nature, there are problems specific to each theory that call for unification with the other. Each has a problem of infinities. In nature, we have yet to encounter anything measurable that has an infinite value. But in both quantum theory and general relativity, we encounter predictions of physically sensible quantities becoming infinite. This is likely the way that nature punishes impudent theorists who dare to break her unity.

General relativity has a problem with infinities because inside a black hole the density of matter and the strength of the gravitational field quickly become infinite. That appears to have also been the case very early in the history of the universe — at least, if we trust general relativity to describe its infancy. At the point at which the density becomes infinite, the equations of general relativity break down. Some people interpret this as time stopping, but a more sober view is that the theory is just inadequate. For a long time, wise people have speculated that it is inadequate because the effects of quantum physics have been neglected.

Quantum theory, in turn, has its own trouble with infinities. They appear whenever you attempt to use quantum mechanics to describe fields, like the electromagnetic field. The problem is that the electric and magnetic fields have values at every point in space.

This means that there are an infinite number of variables (even in a finite volume there are an infinite number of points, hence an infinite number of variables). In quantum theory, there are uncontrollable fluctuations in the values of every quantum variable. An infinite number of variables, fluctuating uncontrollably, can lead to equations that get out of hand and predict infinite numbers when you ask questions about the probability of some event happening, or the strength of some force.

So this is another case where we can't help but feel that an essential part of physics has been left out. There has long been the hope that when gravity is taken into account, the fluctuations will be tamed and all will be finite. If infinities are signs of missing unification, a unified theory will have none. It will be what we call a *finite theory*, a theory that answers every question in terms of sensible, finite numbers.

Quantum mechanics has been extremely successful at explaining a vast realm of phenomena. Its domain extends from radiation to the properties of transistors and from elementary-particle physics to the action of enzymes and other large molecules that are the building blocks of life. Its predictions have been borne out again and again over the course of the last century. But some physicists have always had misgivings about it, because the reality it describes is so bizarre. Quantum theory contains within it some apparent conceptual paradoxes that even after eighty years remain unresolved. An electron appears to be both a wave and a particle. So does light. Moreover, the theory gives only statistical predictions of subatomic behavior. Our ability to do any better than that is limited by the *uncertainty principle*, which tells us that we cannot measure a particle's positon and momentum at the same time. The theory yields only probabilities. A particle — an atomic electron, say — can be anywhere until we measure it; our observation in some sense determines its state. All of this suggests that quantum theory does not tell the whole story. As a result, in spite of its success, there are many experts who are convinced that quantum theory hides something essential about nature that we need to know.

One problem that has bedeviled the theory from the beginning is the question of the relationship between reality and the formalism. Physicists have traditionally expected that science should give an account of reality as it would be in our absence. Physics should be more than a set of formulas that predict what we will observe in an experiment; it should give a picture of what reality *is*. We are accidental descendants of an ancient primate, who appeared only very recently in the history of the world. It cannot be that reality depends on our existence. Nor can the problem of no observers be solved by raising the possibility of alien civilizations, for there was a time when the world existed but was far too hot and dense for organized intelligence to exist.

Philosophers call this view *realism*. It can be summarized by saying that the real world out there (or RWOT, as my first philosophy teacher used to put it) must exist independently of us. It follows that the terms by which science describes reality cannot involve in any essential way what we choose to measure or not measure.

Quantum mechanics, at least in the form it was first proposed, did not fit easily with realism. This is because the theory presupposed a division of nature into two parts. On one side of the division is the system to be observed. We, the observers, are on the other side. With us are the instruments we use to prepare experiments and take measurements, and the clocks we use to record when things happen. Quantum theory can be described as a new kind of language to be used in a dialogue between us and the systems we study with our instruments. This quantum language contains verbs that refer to our preparations and measurements and nouns that refer to what is then seen. It tells us nothing about what the world would be like in our absence.

Since quantum theory was first proposed, a debate has raged between those who accept this way of doing science and those who reject it. Many of the founders of quantum mechanics, including Einstein, Erwin Schrödinger, and Louis de Broglie, found this approach to physics repugnant. They were realists. For them quantum theory, no matter how well it worked, was not a complete theory, because it did not provide a picture of reality absent our interaction with it. On the other side were Niels Bohr, Werner Heisenberg, and many others. Rather than being appalled, they embraced this new way of doing science.

Since then, the realists have scored some successes by pointing to inconsistencies in the present formulation of quantum theory. Some of these apparent inconsistencies arise because, if it is universal, quantum theory should also describe *us*. Problems, then, come from the division of the world required to make sense of quantum theory. One difficulty is where you draw the dividing line, which depends on who is doing the observing. When you measure an atom, you and your instruments are on one side and the atom is on the other side. But suppose I watch you working through a videocam I have set up in your laboratory. I can consider your whole lab — including you and your instruments, as well as the atoms you play with — to constitute one system that I am observing. On the other side would be only me.

You and I hence describe two different "systems." Yours includes just the atom. Mine includes you, the atom, and everything you use to study it. What you see as a measurement, I see as two physical systems interacting with each other. Thus, even if you agree that it's fine to have the observers' actions as part of the theory, the theory as given is not sufficient. Quantum mechanics has to be expanded, to allow for many different descriptions, depending on who the observer is.

This whole issue goes under the name *the foundational problems of quantum mechanics*. It is the second great problem of contemporary physics.

Problem 2: Resolve the problems in the foundations of quantum mechanics, either by making sense of the theory as it stands or by inventing a new theory that does make sense.

There are several different ways one might do this.

1. Provide a sensible language for the theory, one that resolves all puzzles like the ones just mentioned and incorporates the division of the world into system and observer as an essential feature of the theory.
2. Find a new interpretation of the theory — a new way of reading the equations — that is realist, so that measurement and observation play no role in the description of fundamental reality.
3. Invent a new theory, one that gives a deeper understanding of nature than quantum mechanics does.

All three options are currently being pursued by a handful of smart people. There are unfortunately not many physicists who work on this problem. This is sometimes taken as an indication that the problem is either solved or unimportant. Neither is true. This is probably the most serious problem facing modern science. It is just so hard that progress is very slow. I deeply admire the physicists who work on it, both for the purity of their intentions and for their courage to ignore fashion and attack the hardest and most fundamental of problems.

But despite their best efforts, the problem remains unsolved. This suggests to me that it's not just a matter of finding a new way to think about quantum theory. Those who initially formulated the theory were not realists. They did not believe that human beings were capable of forming a true picture of the world as it exists independent of our actions and observations. They argued instead for a very different vision of science: In their view, science can be nothing but an extension of the ordinary language we use to describe our actions and observations to one another.

In more recent times, that view looks self-indulgent — the product of a time we hope we have advanced beyond in many respects. Those who continue to defend quantum mechanics as formulated, and propose it as a theory of the world, do so mostly under the banner of realism. They argue for a reinterpretation of the theory along realist lines. However, while they have made some interesting proposals, none has been totally convincing.

It is possible that realism as a philosophy will simply die off, but this seems unlikely. After all, realism provides the motivation driving most scientists. For most of us, belief in the RWOT and the possibility of truly knowing it motivates us to do the hard work needed to become a scientist and contribute to the understanding of nature. Given the failure of realists to make sense of quantum theory as formulated, it appears more and more likely that the only option is the third one: the discovery of a new theory that will be more amenable to a realist interpretation.

I should admit that I am a realist. I side with Einstein and the others who believe that quantum mechanics is an incomplete description of reality. Where, then, should we look for what is missing in quantum mechanics? It has always seemed to me that the solution will require more than a deeper understanding of quantum physics itself. I believe that if the problem has not been solved after all this time, it is because there is something missing, some link to other problems in physics. The problem of quantum mechanics is unlikely to be solved in isolation; instead, the solution will probably emerge as we make progress on the greater effort to unify physics.

But if this is true, it works both ways: We will not be able to solve the other big problems unless we also find a sensible replacement for quantum mechanics.

The idea that physics should be unified has probably motivated more work in physics than any other problem. But there are different ways that physics can be unified, and we should be careful to distinguish them. So far we have been discussing *unification through a single law.* It is hard to see how anyone could disagree that this is a necessary goal.

But there are other ways to unify the world. Einstein, who certainly thought as much about this as anyone, emphasized that we must distinguish two kinds of theories. There are *theories of principle* and *constructive theories.* A theory of principle is one that sets up the framework that makes a description of nature possible. By definition, a theory of principle must be universal: It must apply to everything because it sets out the basic language we use to talk about nature. There cannot be two different theories of principle, applying to different domains. Because the world is a unity, everything interacts ultimately with everything else, and there can be only one language used to describe those interactions. Quantum theory and general relativity are both theories of principle. As such, logic requires their unification.

The other kind of theories, constructive theories, describe some particular phenomenon in terms of specific models or equations.[1] The theory of the electromagnetic field and the theory of the electron are constructive theories. Such a theory cannot stand alone; it must be set within the context of a theory of principle. But as long as the theory of principle allows, there can be phenomena that obey different laws. For example, the electromagnetic field obeys laws different from those governing the postulated cosmological dark matter (thought to vastly outnumber the amount of ordinary atomic

matter in our universe). One thing we know about the dark matter is that, whatever it is, it is dark. This means it gives off no light, so it likely doesn't interact with the electromagnetic field. Thus two different theories can coexist side by side.

The point is that the laws of electromagnetism do not dictate what else exists in the world. There can be quarks or not, neutrinos or not, dark matter or not. Similarly, the laws that describe the two forces — strong and weak — that act within the atomic nucleus do not necessarily require that there be an electromagnetic force. We can easily imagine a world with electromagnetism but no strong nuclear force, or the reverse. As far as we know, either possibility would be consistent.

But it is still possible to ask whether all the forces we observe in nature might be manifestations of a single, fundamental force. There seems, as far as I can tell, no logical argument that this *should* be true, but it is still something that *might* be true.

The desire to unify the various forces has led to several significant advances in the history of physics. James Clerk Maxwell, in 1867, unified electricity and magnetism into one theory, and a century later, physicists realized that the electromagnetic field and the field that propagates the weak nuclear force (the force responsible for radioactive decay) could be unified. This became the *electroweak* theory, whose predictions have been repeatedly confirmed in experiments over the last thirty years.

There are two fundamental forces in nature (that we know of) that remain outside the unification of the electromagnetic and weak fields. These are gravity and the strong nuclear force, the force responsible for binding the particles called quarks together to form the protons and neutrons making up the atomic nucleus. Can all four fundamental forces be unified?

This is our third great problem.

Problem 3: Determine whether or not the various particles and forces can be unified in a theory that explains them all as manifestations of a single, fundamental entity.

Let us call this problem *the unification of the particles and forces,* to distinguish it from the unification of laws, the unification we discussed earlier.

At first, this problem appears easy. The first proposal for how to unify gravity with electricity and magnetism was made in 1914, and many more have been offered since. They all work, as long as you forget one thing, which is that nature is quantum mechanical. If you leave quantum physics out of the picture, unified theories are easy to invent. But if you include quantum theory, the problem gets much, much harder. Since gravity is one of the four fundamental forces of nature, we must solve the problem of quantum gravity (that is, problem no. 1: how to reconcile general relativity and quantum theory) along with the problem of unification.

Over the last century, our physical description of the world has simplified quite a bit. As far as particles are concerned, there appear to be only two kinds, quarks and leptons. Quarks are the constituents of protons and neutrons and many particles we have discovered similar to them. The class of leptons encompasses all particles not made of quarks, including electrons and neutrinos. Altogether, the known world is explained by six kinds of quarks and six kinds of leptons, which interact with each other through the four forces (or interactions, as they are also known): gravity, electromagnetism, and the strong and weak nuclear forces.

Twelve particles and four forces are all we need to explain everything in the known world. We also understand very well the basic physics of these particles and forces. This understanding is expressed in terms of a theory that accounts for all of these particles and all of the forces except for gravity. It's called *the standard model of elementary-particle physics* — or the standard model, for short. This theory does not have the problem of infinities mentioned earlier. Anything we want to compute in this theory we can, and it results in a finite number. In the more than thirty years since it was formulated, many predictions made by this theory have been checked experimentally. In each and every case, the theory has been confirmed.

The standard model was formulated in the early 1970s. Except for the discovery that neutrinos have mass, it has not required adjustment since. So why wasn't physics over by 1975? What remained to be done?

For all its usefulness, the standard model has a big problem: It has a long list of adjustable constants. When we state the laws of the theory, we must specify the values of these constants. As far as we

know, any values will do, because the theory is mathematically consistent no matter which values we put in. These constants specify the properties of the particles. Some tell us the masses of the quarks and the leptons, while others tell us the strengths of the forces. We have no idea why these numbers have the values they do; we simply determine them by experiments and then plug in the numbers. If you think of the standard model as a calculator, then the constants will be dials that can be set to whatever positions you like each time the program is run.

There are about twenty such constants, and the fact that there are that many freely specifiable constants in what is supposed to be a fundamental theory is a tremendous embarrassment. Each one represents some basic fact of which we are ignorant: namely, the physical reason or mechanism responsible for setting the constant to its observed value.

This is our fourth big problem.

Problem 4: Explain how the values of the free constants in the standard model of particle physics are chosen in nature.

It is devoutly hoped that a true unified theory of the particles and forces will give a unique answer to this question.

In 1900, William Thomson (Lord Kelvin), an influential British physicist, famously proclaimed that physics was over, except for two small clouds on the horizon. These "clouds" turned out to be the clues that led us to quantum theory and relativity theory. Now, even as we celebrate the encompassing of all known phenomena in the standard model plus general relativity, we, too, are aware of two clouds. These are the dark matter and the dark energy.

Apart from the issue of its relationship with the quantum, we think we understand gravity very well. The predictions of general relativity have been found to be in agreement with observation to a very precise degree. The observations in question extend from falling bodies and light on Earth, to the detailed motion of the planets and their moons, to the scales of galaxies and clusters of galaxies. Formerly exotic phenomena — such as gravitational lensing, an effect of the curvature of space by matter — are now so well understood that they are used to measure the distributions of mass in galactic clusters.

In many cases — those in which velocities are small compared with that of light, and masses are not too compact — Newton's laws of gravity and motion provide an excellent approximation to the predictions of general relativity. Certainly they should help us predict how the motion of a particular star is influenced by the masses of stars and other matter in its galaxy. But they don't. Newton's law of gravity says that the acceleration of any object as it orbits another is proportional to the mass of the body it is orbiting. The heavier the star, the faster the orbital motion of the planet. That is, if two stars are each orbited by a planet, and the planets are the same distances from their stars, the planet orbiting the more massive star will move faster. Thus if you know the speed of a body in orbit around a star and its distance from the star, you can measure the mass of that star. The same holds for stars in orbit around the center of their galaxy; by measuring the orbital speeds of the stars, you can measure the distribution of mass in that galaxy.

Over the last decades, astronomers have done a very simple experiment in which they measure the distribution of mass in a galaxy in two different ways and compare the results. First, they measure the mass by observing the orbital speeds of the stars; second, they make a more direct measurement of the mass by counting all the stars, gas, and dust they can see in the galaxy. The idea is to compare the two measurements: Each should tell them both the total mass in the galaxy and how it is distributed. Given that we understand gravity well, and that all known forms of matter give off light, the two methods should agree.

They don't. Astronomers have compared the two methods of measuring mass in more than a hundred galaxies. In almost all cases, the two measurements don't agree, and not by just a small amount but by factors of up to 10. Moreover, the error always goes in one direction: There is always more mass needed to explain the observed motions of the stars than is seen by directly counting up all the stars, gas, and dust.

There are only two explanations for this. Either the second method fails because there is much more mass in a galaxy than is visible, or Newton's laws fail to correctly predict the motions of stars in the gravitational field of their galaxy.

All the forms of matter we know about give off light, either directly as in starlight or reflected from planets or interstellar rocks, gas, and dust. So if there is matter we don't see, it must be in some novel form that neither emits nor reflects light. And because the discrepancy is so large, the majority of the matter in galaxies must be in this new form.

Today most astronomers and physicists believe that this is the right answer to the puzzle. There is missing matter, which is actually there but which we don't see. This mysterious missing matter is referred to as the *dark matter*. The dark-matter hypothesis is preferred mostly because the only other possibility — that we are wrong about Newton's laws, and by extension general relativity — is too scary to contemplate.

Things have become even more mysterious. We have recently discovered that when we make observations at still larger scales, corresponding to billions of light-years, the equations of general relativity are not satisfied even when the dark matter is added in. The expansion of the universe, set in motion by the Big Bang some 13.7 billion years ago, appears to be accelerating, whereas, given the observed matter plus the calculated amount of dark matter, it should be doing the opposite — decelerating.

Again, there are two possible explanations. General relativity could simply be wrong. It has been verified precisely only within our solar system and nearby systems in our own galaxy. Perhaps when one gets to a scale comparable to the size of the whole universe, general relativity is simply no longer applicable.

Or there is a new form of matter — or energy (recall Einstein's famous equation $E = mc^2$, showing the equivalence of energy and mass) — that becomes relevant on these very large scales: That is, this new form of energy affects only the expansion of the universe. To do this, it cannot clump around galaxies or even clusters of galaxies. This strange new energy, which we have postulated to fit the data, is called the *dark energy*.

Most kinds of matter are under pressure, but the dark energy is under tension — that is, it pulls things together rather than pushes them apart. For this reason, tension is sometimes called negative pressure. In spite of the fact that the dark energy is under tension, it causes the universe to expand faster. If you are confused by this, I sympathize. One would think that a gas with negative pressure would act like a rubber band connecting the galaxies and slow the expansion down. *But it turns out that when the negative pressure is negative enough, in general relativity it has the opposite effect.* It causes the expansion of the universe to accelerate.

Recent measurements reveal a universe consisting mostly of the unknown. Fully 70 percent of the matter density appears to be in the form of dark energy. Twenty-six percent is dark matter. Only 4 percent is ordinary matter. So less than 1 part in 20 is made out of matter we have observed experimentally or described in the standard model of particle physics. Of the other 96 percent, apart from the properties just mentioned, we know absolutely nothing.

In the last ten years, cosmological measurements have gotten much more precise. This is partly a side effect of Moore's law, which states that every eighteen months or so, the processing speeds of computer chips will double. All the new experiments use microchips in either satellites or ground-based telescopes, so as the chips have gotten better, so have the observations. Today we know a lot about the basic characteristics of the universe, such as the overall matter density and the rate of expansion. There is now a standard model of cosmology, just as there is a standard model of elementary-particle physics. Just like its counterpart, the standard model of cosmology has a list of freely specifiable constants — in this case, about fifteen. These denote, among other things, the density of different kinds of matter and energy and the expansion rate. No one knows anything about why these constants have the values they do. As in particle physics, the values of the constants are taken from observations but are not yet explained by any theory.

These cosmological mysteries make up the fifth great problem.

Problem 5: Explain dark matter and dark energy. Or, if they don't exist, determine how and why gravity is modified on large scales. More generally, explain why the constants of the standard model of cosmology, including the dark energy, have the values they do.

These five problems represent the boundaries to present knowledge. They are what keep theoretical physicists up at night. Together they drive most current work on the frontiers of theoretical physics.

Any theory that claims to be a fundamental theory of nature

must answer each one of them. One of the aims of this book is to evaluate just how well recent physical theories, such as string theory, have done in achieving this goal. But before we do that, we need to examine some earlier attempts at unification. We have a great deal to learn from the successes — and also from the failures.