“Whilst this planet has been cycling on according to the fixed law of gravity, from so simple a beginning, forms most wonderful . . . have been and are being evolved.” These are the famous closing words of Charles Darwin’s *On the Origin of Species*.

Darwin’s genius was to recognize how “natural selection of favored variations” could have transformed primordial life (formed, he surmised, in a “warm little pond”) into the amazing varieties of creatures that crawl, swim, or fly on Earth. But
this emergence—a higgledy-piggledy process, proceeding without any guiding hand—is inherently very slow. Darwin guessed that it would require hundreds of millions of years. This expanse of time didn’t seem to faze him, because geologists had already invoked such timespans to account for the laying-down of rocks and the shaping of the Earth’s surface features. Indeed, in the first edition of his book, Darwin estimated the rate of erosion in the Weald of Kent, a broad valley near his home, and gauged that this geological feature had to be 300 million years old. He insisted that “there is a grandeur in this view of life.” To his nineteenth century contemporaries, these huge timescales were themselves mind-stretching compared to the constricted timespans of more traditional Western cosmologies.

There were, however, some troubling arguments that seemed to preclude such an ancient Earth. Lord Kelvin, one of the most celebrated physicists of his time, calculated that it would take only a few million years for the heat stored in Earth’s molten core to leak out. And the Sun itself, he claimed, was radiating away its internal heat so fast that it would deflate in 10 million years. Kelvin’s views carried great weight: “We may say, with certainty,” he wrote, “that the inhabitants of the Earth cannot continue to enjoy the light and heat essential to their life, for many millions of years longer, unless sources now unknown to us are prepared in the great storehouse of creation.”¹ The American geologist Thomas Chamberlain retorted that “there is perhaps no beguilement more insidious and dangerous than an elaborate and elegant mathematical process built upon unfortified premises.”

Chamberlain was writing in 1899, but his words strike a
special resonance today. In later chapters, I shall discuss some intellectually alluring theories that may account for the most basic features of our physical universe—why it is expanding the way it is, why it contains atoms (and other particles), the strengths of the forces between them, and the nature of space itself. But the boldest and most ambitious of these are still based on “unfortified premises.”

Even before Kelvin’s death, advances in physics eroded the credibility of his estimates. Henri Becquerel’s discovery in 1896 that uranium emitted mysterious rays gave the first hint of hitherto unenvisioned energy latent within atoms, and of what the Sun’s actual fuel might be. Moreover, it was realized that power generated by radioactive decays in Earth’s core could replenish the heat that was leaking out. But identifying the precise reactions that allowed the Sun to shine for 10 billion years as a gravitationally confined fusion reactor had to await a better understanding of atomic nuclei.

By the 1930s, it was realized that an atom consisted of a positively charged nucleus, surrounded by orbiting electrons with negative electric charge. The atoms of the different elements were distinguished from one another by the charge on their nucleus and the number of electrons needed to neutralize it. This number is 1 for hydrogen and 92 for uranium, the heaviest naturally occurring atom. The chemical properties of the various elements were already known in the nineteenth century, when the great Russian chemist Dmitri Mendeleyev put forward the periodic table of the elements, which showed certain patterns and family resemblances among different groups of elements. During the early twentieth century, the new quantum theory offered a natural explanation for these
patterns, showing how they depended on the detailed orbits of the electrons surrounding the nuclei of atoms.

The atomic nuclei themselves are made up of protons and neutrons: the number of protons determines the electric charge and the place in the periodic table. A nucleus of helium, the second simplest element, consists of two protons and two neutrons. It weighs 0.7 percent less than the four particles from which it is made, however, and this difference in mass corresponds (via Einstein’s famous equation $E = mc^2$) to the energy released when hydrogen is transmuted into helium. Nuclear fusion releases about a million times more energy per kilogram than any chemical process—combustion or explosion. The violence of an H-bomb testifies to fusion’s power. Chemical processes merely alter or reshuffle the orbiting electrons; on the other hand, nuclear fusion taps a far larger reservoir of energy in the nuclei themselves.

The first calculation of how this fusion occurs in a star like the Sun depended on the insights of three great physicists. The first was a Russian, George Gamow, who will reappear in chapter 5 as one of the pioneers of the Big Bang cosmological theory. Gamow used the newly formulated quantum theory to estimate how hot the Sun’s core would need to be for hydrogen fusion to be self-sustaining. The detailed chain reactions were worked out by Carl-Friedrich von Weizsaecker and Hans Bethe. These two eminent (and then young) physicists were both Germans, though Bethe went to the United States in the 1930s, having already achieved distinction as a pioneer of nuclear physics; amazingly, he still remained active in the subject at the beginning of the twenty-first century.
Our conception of the Sun’s life cycle is depicted in figure 1.1 in a time-lapse format, with 150 million years between successive frames. The proto-Sun condensed from a cloud of diffuse interstellar gas. Gravity pulled this entity together until its center was squeezed hot enough to trigger nuclear fusion of hydrogen into helium at a sufficient rate to balance the heat shining from its surface. Less than half of the Sun’s central hydrogen has so far been used up, even though it is already 4.5 billion years old. It will keep shining for a further 5 billion years. It will then swell up to become a “red giant,” large and bright enough to engulf the inner planets and to vaporize all life on Earth. During this red giant phase, lasting some 500 million years, hydrogen will continue to burn in a
shell around the helium core. Next, the Sun will undergo a more rapid convulsion, triggered by the onset of helium fusion in its core. This action will blow off some outer layers—about a quarter of the Sun’s mass altogether. The residue will become a white dwarf—a dense stellar cinder no larger than Earth—which will shine with a bluish glow, no brighter than today’s full Moon, on whatever remains of the solar system.

To convey these vast timescales, an analogy can help. Imagine a walk across America, starting in New York when the Sun was born and ending in California when it is about to die. To pace yourself on this walk, you would have to take one step every 2000 years. All recorded history would be just a few steps. Moreover, those steps would be just before the halfway stage: somewhere in Kansas, perhaps—not the culmination of the journey. Likewise, even our Sun has more time ahead of it than has so far elapsed; and our entire universe could have an infinite future awaiting it.

We may still be near Darwin’s “simple beginning”: if life isn’t prematurely snuffed out, our remote progeny will surely—in the eons that lie ahead—spread far beyond this planet. Even if life is now unique to Earth, there is time enough for it to “green” the entire Galaxy, and even to spread beyond.

Other Solar Systems

In the early twentieth century, our solar system was suspected of being the outcome of a close encounter between our Sun and a passing star, which tore from it a stream of gas. This stream condensed into droplets, each a protoplanet. Stars are
very thinly spread through space, however: in a scale model where the Sun was the size of a tennis ball, the nearest stars would be thousands of kilometers from the Sun and from one another. Close encounters would be freakishly rare; so also, if this theory were right, would be planetary systems. But such “catastrophist” views fell from favor in the second half of the century. Astronomers came to prefer an alternative theory that rendered planets a natural concomitant of star formation.

When interstellar gas contracts to form a star, its density rises a billion billion times. Any slight spin in the original gas would have been so much amplified during the collapse (a cosmic version of what happens when pirouetting ice skaters pull in their arms) that centrifugal forces would prevent all of it from contracting to the size of a star. Instead, any protostar, as it contracted, would naturally spin off a disk. In these disks, made from the gas and dust that pervades interstellar space, dust particles would stick together to make rocky “planetesmals,” which in turn merge to make planets.

Disks have now been detected around newly forming protostars in the Orion Nebula and elsewhere. (Their discoverers coined the name “proplyds,” short for protoplanetary disks). Proplyds are a natural concomitant of any star’s birth, so there is every reason to expect other stars to be orbited by retinues of planets.

Even if they were orbiting one of the nearest stars, planets would be too faint to be seen directly with present-day telescopes. A planet would appear fainter than its parent star by a huge factor—roughly the same factor, in fact, by which Venus and Jupiter appear to us fainter than our Sun. But in
the last few years, planets have been revealed indirectly through their effect on their parent star. Some stars have been found to be wobbling slightly in their positions, just as would be expected if planets were orbiting around them. A planet tugs the star around in a small counterorbit, rather like a small dog pulling its owner on a leash.

The first success was achieved by two Swiss astronomers, Michel Mayor and Didier Queloz. They analyzed the light from a nearby star, 51 Persei, which closely resembles the Sun, to see if there were slight changes in its color (or, in physical terms, in the wavelength of its light). If an object moves toward us, the waves get bunched up and appear to have shorter wavelengths—in other words, they shift toward the blue end of the spectrum. Conversely, light seems reddened if its source recedes from us. This is analogous to the so-called Doppler effect for sound waves, whereby the pitch is higher if the source approaches us. (Doppler famously used trumpeters on a railway truck to demonstrate this now-familiar effect.) Mayor and Queloz found that the light from 51 Persei shifted slightly toward the blue, and then toward the red, then toward the blue again; the pattern repeated regularly. This regularity implied that the star had a near-circular motion, induced by an unseen planet: star and planet pivot around the system’s center of mass, so that the presence of the planet makes the star itself move. The inferred planet is about the size of Jupiter and orbits at around 50 kilometers per second. The star itself moves at only 50 meters per second—one thousand times slower than the planet because it is one thousand times heavier.

It was a real technical triumph to detect these slight mo-
tions. Geoffrey Marcy and Paul Butler in California are now the champion planet hunters, having used the same technique as Mayor and Queloz to find planets around dozens of other stars. This technique measures just the part of the star’s motion that is directed along our line of sight—a small transverse motion does not show up as a Doppler shift. But specially instrumented telescopes should soon be able to reveal the tiny side-to-side motion induced by an orbiting planet by detecting slight changes in a star’s position on the sky. In 1999, Marcy and Butler discovered that the nearby star Upsilon Andromedae had at least three Jupiter-sized planets: one in a very close circular orbit, with period 4.6 days; the other two in larger, slower orbits. More and more stars are now being revealed to have orbiting planets. The systems display surprising variety: there are planets that are up to about twenty times as heavy as Jupiter; the orbital periods can be as short as a few days; and the orbits are sometime near-circular, but surprisingly often they are highly elliptical.

The eventual goal, of course, is to have a sharp and sensitive image that reveals orbiting planets directly. This kind of image is still something for the future. But some claimed planets have already revealed themselves in other ways—for instance, by causing slight changes in the apparent brightness of their parent star. If a planet moved across the face of the star (as, in our own solar system, the planet Venus occasionally transits the face of the Sun), then the star would dim slightly each time, once per orbit, that the planet passed in front of it. This technique only works, of course, if our line of sight is close to the plane of the orbit.
Other Earths?

The planets found so far, orbiting solar-type stars, are all roughly the size of Jupiter or Saturn. But there is every reason to suspect that these are the largest planets in other “solar systems” whose smaller planets are not yet detectable. Planets the size of Earth would induce motions of merely centimeters per second, not meters, in the central star—too small to be discerned by current techniques. They are also hard to find by any other method. If such a planet were to move in front of a star, it would reduce its brightness by less than one part in 10,000. The best hope of detecting this minuscule dimming would be to use a telescope in space, where the starlight is unaffected by Earth’s atmosphere and therefore is steadier. A planned European space mission called “Eddington” (named after the famous English astronomer) should be able to detect transits of Earthlike planets across bright stars within the next decade. The longer-term goal is to observe Earth-sized planets directly, rather than just inferring them. This capability will require very large telescope arrays in space—and it is far from a crazy idea.

From my home base in England, I watch the U.S. space program with interest and admiration. It is far larger than Europe’s, its scale being a legacy of superpower rivalry. I am underwhelmed by the International Space Station. But it is better news that NASA’s somewhat messianic chief executive, Dan Goldin, has focused the less costly unmanned program on the scientific theme of “Origins,” and has included the so-called Terrestrial Planet Finder, capable of detecting planets as small as our Earth, as a main thrust of that program. In Europe, a similar project, called “Darwin,” is also being planned.2
We were all, when young, taught the layout of our own solar system—the sizes of the nine major planets and how they move in orbit around the Sun. But, twenty years from now we shall be able to tell our grandchildren far more interesting things on a starry night. Nearby stars will no longer just be points of light—we will think of them as the Suns of other solar systems. We will know the orbits of each star’s retinue of planets, and the sizes (and even some topographic details) of the bigger ones.

We will be especially interested in possible twins of our Earth—planets the same size as ours, orbiting other Sunlike stars, having temperate climates. We still don’t know how many of these objects there are.

Most of the systems so far discovered, incidentally, are surprisingly different from our own solar system and offer rather poor prospects for habitable planets. Many contain Jupiterlike planets on eccentric orbits much closer to their parent star than our own Jupiter is. These massive bodies would destabilize any Earthlike planet in a near-circular orbit at the “right” distance for its parent star. This discouraging finding may be partly the outcome of observational selection: fast-moving heavy planets, orbiting close to their parent star, have been the easiest for Marcy and Butler to detect. We cannot yet be sure what fraction of planetary systems would permit an Earthlike planet to survive undisturbed for billions of years in a near-circular orbit; but among the many millions of planetary systems (formed with one, two, or three high-mass planets), there would surely be some planets on Earthlike orbits, with temperatures such that water neither boils nor stays frozen.