Chapter 1
Light

> Always the laws of light are the same, but the modes and degrees of seeing vary.
>
> – Henry David Thoreau, *A Week on the Concord and Merrimack Rivers*, 1849

Astronomy is not for the faint of heart. Almost everything it cares for is indescribably remote, tantalizingly untouchable, and invisible in the daytime, when most sensible people do their work. Nevertheless, many – including you, brave reader – have enough curiosity and courage to go about collecting the flimsy evidence that reaches us from the universe outside our atmosphere, and to hope it may hold a message.

This chapter introduces you to astronomical evidence. Some evidence is in the form of material (like meteorites), but most is in the form of light from faraway objects. Accordingly, after a brief consideration of the material evidence, we will examine three theories for describing the behavior of light: light as a wave, light as a quantum entity called a photon, and light as a geometrical ray. The ray picture is simplest, and we use it to introduce some basic ideas like the apparent brightness of a source and how that varies with distance. Most information in astronomy, however, comes from the analysis of how brightness changes with wavelength, so we will next introduce the important idea of spectroscopy. We end with a discussion of the astronomical magnitude system. We begin, however, with a few thoughts on the nature of astronomy as an intellectual enterprise.

1.1 The story

> . . . as I say, the world itself has changed . . . . For this is the great secret, which was known by all educated men in our day: that by what men think, we create the world around us, daily new.
>
> – Marion Zimmer Bradley, *The Mists of Avalon*, 1982

Astronomers are storytellers. They spin tales of the universe and of its important parts. Sometimes they envision landscapes of another place, like the roiling liquid-metal core of the planet Jupiter. Sometimes they describe another time, like the era before Earth when dense buds of gas first flowered into stars, and a darkening Universe filled with the sudden blooms of galaxies. Often the stories solve mysteries or illuminate something commonplace or account for something monstrous: How is it that stars shine, age, or explode? Some of the best stories tread the same ground as myth: What threw up the mountains of the Moon? How did the skin of our Earth come to teem with life? Sometimes there are fantasies: What would happen if a comet hit the Earth? Sometimes there are prophecies: How will the Universe end?

Like all stories, creation of astronomical tales demands imagination. Like all storytellers, astronomers are restricted in their creations by many conventions of language as well as by the characters and plots already in the literature. Astronomers are no less a product of their upbringing, heritage, and society than any other crafts people. Astronomers, however, think their stories are special, that they hold a larger dose of "truth" about the universe than any others. Clearly, the subject matter of astronomy – the Universe and its important parts – does not belong only to astronomers. Many others speak with authority about just these things: theologians, philosophers, and poets, for example. Is there some characteristic of astronomers, besides arrogance, that sets them apart from these others? Which story about the origin of the Moon, for example, is the truer: the astronomical story about a collision 4500 million years ago between the proto-Earth and a somewhat smaller proto-planet, or the mythological story about the birth of the Sumerian/Babylonian deity Nanna-Sin (a rather formidable fellow who had a beard of lapis-lazuli and rode a winged bull)?

This question of which is the "truer" story is not an idle one. Over the centuries, people have discovered (by being proved wrong) that it is very difficult to have a commonsense understanding of what the whole Universe and its most important parts are like. Common sense just isn't up to the task. For that reason, as Morgan le Fey tells us in *The Mists of Avalon,* created stories about the Universe themselves actually *create* the Universe the listener lives in. The real Universe (like most scientists, you and I behave as if there is one) is not silent, but whispers very softly to the storytellers. Many whispers go unheard, so that real Universe is probably very different from the one you read about today in any book that claims to tell its story. People, nevertheless, must act. Most recognize that the bases for their actions are fallible stories, and they must therefore select the most trustworthy stories that they can find.

Most of you won't have to be convinced that it is better to talk about colliding planets than about Nanna-Sin if your aim is to understand the Moon or perhaps plan a visit. Still, it is useful to ask the question: What is it, if anything, that makes astronomical stories a more reliable basis for action, and in that sense more truthful or factual than any others? Only one thing, I think: *discipline*. Astronomers feel an obligation to tell their story with great care, following a rather strict, scientific, discipline.

Scientists, philosophers, and sociologists have written about what it is that makes science different from other human endeavors. There is much discussion and disagreement about the necessity of making scientific stories "broad and deep and simple", about the centrality of paradigms, the importance of predictions, the strength or relevance of motivations, and the inevitability of conformity to social norms and professional hierarchies.

But most of this literature agrees on the perhaps obvious point that a scientist, in creating a story (scientists usually call them "theories") about, say, the Moon, must pay a great deal of attention to all the relevant evidence. A scientist, unlike a science-fiction writer, may only fashion a theory that cannot be shown to violate that evidence.

This is a book about how to identify and collect relevant evidence in astronomy.

1.2 The evidence: astronomical data

> [Holmes said] I have no data yet. It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts
>
> – Arthur Conan Doyle, *The Adventures of Sherlock Holmes*, 1892

> Facts are not pure and unsullied bits of information; culture also influences what we see and how we see it. Theories moreover are not inexorable inductions from facts. The most creative theories are often imaginative visions imposed upon facts; . . .
>
> – Stephen Jay Gould, *The Mismeasure of Man*, 1981

A few fortunate astronomers investigate cosmic rays or the Solar System. All other astronomers must construct stories about objects with which they can have no direct contact, things like stars and galaxies that can't be manipulated, isolated, or made the subject of experiment. This sets astronomers apart from most other scientists, who can thump on, cut up, and pour chemicals over their objects of study. In this sense, astronomy is a lot more like paleontology than it is like physics. Trying to tell the story of a galaxy is like trying to reconstruct a dinosaur from bits of fossilized bone. We will never have the galaxy or dinosaur in our laboratory, and must do guesswork based on flimsy, secondhand evidence. To study any astronomical object we depend on intermediaries, entities that travel from the objects to us. There are two categories of intermediaries – particles with mass, and those without. First briefly consider the massive particles, since detailed discussion of them is beyond the scope of this book.

1.2.1 Particles with mass

Cosmic rays are microscopic particles that arrive at Earth with extraordinarily high energies. ***Primary cosmic rays*** are mostly high-speed atomic nuclei, mainly hydrogen (84%) and helium (14%). The remainder consists of heavier nuclei, electrons, and positrons. Some primary cosmic rays are produced in solar flares, but many, including those of highest energies, come from outside the Solar System. About 6000 cosmic rays strike each square meter of the Earth's upper atmosphere every second. Since all these particles move at a large fraction of the speed of light, they carry a great deal of kinetic energy. A convenient unit for measuring particle energies is the electron volt (eV):

$$1\ \text{eV} = 1.602 \times 10^{-19}\ \text{joules}$$

Primary cosmic rays have energies ranging from 10^6 to 10^{20} eV, with relative abundance declining with increasing energy. The mean energy is around 10 GeV = 10^{10} eV. At relativistic velocities, the relation between speed, v, and total energy, E, is

$$E = \frac{mc^2}{\sqrt{1 - v^2/c^2}}$$

Here m is the rest mass of the particle and c is the speed of light. For reference, the rest mass of the proton (actually, the product mc^2) is 0.93 GeV. The highest-energy cosmic rays have energies far greater than any attainable in laboratory particle accelerators. Although supernova explosions are suspected to be the source of some or all of the higher-energy primary cosmic rays, the exact mechanism for their production remains mysterious.

Secondary cosmic rays are particles produced by collisions between the primaries and particles in the upper atmosphere – generally more that 50 km above the surface. Total energy is conserved in the collision, so the kinetic energy of the primary can be converted into the rest-mass of new particles, and studies of the secondaries gives some information about the primaries. Typically, a cosmic-ray collision produces many fragments, including pieces of the target nucleus, individual nucleons, and electrons, as well as particles not present before the collision: positrons, gamma rays, and a variety of more unusual short-lived particles like kaons. In fact, cosmic-ray experiments were the first to detect pions, muons, and positrons.

Detection of both primary and secondary cosmic rays relies on methods developed for laboratory particle physics. Detectors include cloud and spark chambers, Geiger and scintillation counters, flash tubes, and various solid-state devices. Detection of primaries requires placement of a detector above the bulk of the Earth's atmosphere, and only secondary cosmic rays can be studied directly from the Earth's surface. Since a shower of secondary particles generally spreads over an area of many square kilometers by the time it reaches sea level, cosmic-ray studies often utilize arrays of detectors. Typical arrays consist of many tens or hundreds of individual detectors linked to a central coordinating computer. Even very dense arrays, however, can only sample a small fraction of the total number of secondaries in a shower.

Neutrinos are particles produced in nuclear reactions involving the weak nuclear force. They are believed to have tiny rest masses (the best measurements to date are uncertain but suggest something like 0.05 eV). They may very well be the most numerous particles in the Universe. Many theories predict intense production of neutrinos in the early stages of the Universe, and the nuclear reactions believed to power all stars produce a significant amount of energy in the form of neutrinos. In addition, on Earth, a flux of high-energy "atmospheric" neutrinos is generated in cosmic-ray secondary showers.

Since neutrinos interact with ordinary matter only through the weak force, they can penetrate great distances through dense material. The Earth and the Sun, for example, are essentially transparent to them. Neutrinos can nonetheless be detected: the trick is to build a detector so massive that a significant number of neutrino reactions will occur within it. Further, the detector must also be shielded from secondary cosmic rays, which can masquerade as neutrinos. About a half-dozen such "neutrino telescopes" have been built underground.

For example, the Super-Kamiokande instrument is a 50 000-ton tank of water located 1 km underground in a zinc mine 125 miles west of Tokyo. The water acts as both the target for neutrinos and as the detecting medium for the products of the neutrino reactions. Reaction products emit light observed and analyzed by photodetectors on the walls of the tank.

Neutrinos have been detected unambiguously from only two astronomical objects: the Sun and a nearby supernova in the Large Magellanic Cloud, SN 1987A. These are promising results. Observations of solar neutrinos, for example, provide an opportunity to test the details of theories of stellar structure and energy production.

Meteorites are macroscopic samples of solid material derived primarily from our Solar System's asteroid belt, although there are a few objects that originate from the surfaces of the Moon and Mars. Since they survive passage through the Earth's atmosphere and collision with its surface, meteorites can be subjected to physical and chemical laboratory analysis. Some meteorites have remained virtually unchanged since the time of the formation of the Solar System, while others have endured various degrees of processing. All, however, provide precious clues about the origin, age, and history of the Solar System. For example, the age of the Solar System (4.56 Gyr) is computed from radioisotopic abundances in meteorites, and the inferred original high abundance of radioactive aluminum-26 in the oldest mineral inclusions in some meteorites suggests an association between a supernova, which would produce the isotope, and the events immediately preceding the formation of our planetary system.

Exploration of the Solar System by human spacecraft began with landings on the Moon in the 1960s and 1970s. Probes have returned samples – Apollo and Luna spacecraft brought back several hundred kilograms of rock from the Moon. Humans and their mechanical surrogates have examined remote surfaces *in situ.* The many landers on Mars, the Venera craft on Venus, and the Huygens lander on Titan, for example, made intrusive measurements and conducted controlled experiments.

1.2.2 Massless particles

Gravitons, theorized particles corresponding to gravity waves, have only been detected indirectly through the behavior of binary neutron stars. Graviton detectors designed to sense the local distortion of space-time caused by a passing gravity wave have been constructed, but have not yet detected waves from an astronomical source.

Photons are particles of light that can interact with all astronomical objects.[1] Light, in the form of visible rays as well as invisible rays like radio and X-rays, has historically constituted the most important channel of astronomical information. This book is about using that channel to investigate the Universe.

1.3 Models for the behavior of light

Some (not astronomers!) regard astronomy as applied physics. There is some justification for this, since astronomers, to help tell some astronomical story, persistently drag out theories proposed by physicists. Physics and astronomy differ partly because astronomers are interested in telling the story of an object, whereas physicists are interested in uncovering the most fundamental rules of the material world. Astronomers tend to find physics useful but sterile; physicists tend to find astronomy messy and mired in detail. We now invoke physics, to ponder the question: how does light behave? More specifically, what properties of light are important in making meaningful astronomical observations and predictions?

1.3.1 Electromagnetic waves

> . . . we may be allowed to infer, that homogeneous light, at certain equal distances in the direction of its motion, is possessed of opposite qualities, capable of neutralizing or destroying each other, and extinguishing the light, where they happen to be united; . . .
>
> – Thomas Young, *Philosophical Transactions, The Bakerian Lecture*, 1804

[1] Maybe not. There is strong evidence for the existence of very large quantities of “dark” matter in the Universe. This matter seems to exert gravitational force, but is the source of no detectable light. It is unclear whether the dark matter is normal stuff that is well hidden, or unusual stuff that can’t give off or absorb light. Even more striking is the evidence for the presence of “dark energy” – a pressure-like effect in space itself which contains energy whose mass equivalent is even greater than that of visible and dark matter combined.

Electromagnetic waves are a model for the behavior of light which we know to be incorrect (*incomplete* is perhaps a better term). Nevertheless, the wave theory of light describes much of its behavior with precision, and introduces a lot of vocabulary that you should master. Christian Huygens,[2] in his 1678 book, *Traité de la Lumière*, summarized his earlier findings that visible light is best regarded as a *wave* phenomena, and made the first serious arguments for this point of view. Isaac Newton, his younger contemporary, opposed Huygens' wave hypothesis and argued that light was composed of tiny solid particles.

A ***wave*** is a disturbance that propagates through space. If some property of the environment (say, the level of the water in your bathtub) is disturbed at one place (perhaps by a splash), a wave is present if that disturbance moves continuously from place to place in the environment (ripples from one end of your bathtub to the other, for example). Material particles, like bullets or ping-pong balls, also propagate from place to place. Waves and particles share many characteristic behaviors – both can ***reflect***, (change directions at an interface) ***refract*** (change speed in response to a change in the transmitting medium), and can carry energy from place to place. However, waves exhibit two characteristic behaviors not shared by particles:

Diffraction – the ability to bend around obstacles. A water wave entering a narrow opening, for example, will travel not only in the "shadow" of the opening but will spread in all directions on the far side.

Interference – an ability to combine with other waves in predictable ways. Two water waves can, for example, destructively interfere if they combine so that the troughs of one always coincide with the peaks of the other.

Although Huygens knew that light exhibited the properties of diffraction and interference, he unfortunately did not discuss them in his book. Newton's reputation was such that his view prevailed until the early part of the nineteenth century, when Thomas Young and Augustin Fresnel were able to show how Huygen's wave idea could explain diffraction and interference. Soon the evidence for waves proved irresistible.

Well-behaved waves exhibit certain measurable qualities – amplitude, wavelength, frequency, and wave speed – and physicists in the generation following Fresnel were able to measure these quantities for visible light waves. Since light was a wave, and since waves are disturbances that propagate, it was

[2] Huygens (1629–1695), a Dutch natural philosopher and major figure in seventeenth-century science, had an early interest in lens grinding. He discovered the rings of Saturn and its large satellite, Titan, in 1655–1656, with a refracting telescope of his manufacture. At about the same time, he invented the pendulum clock, and formulated a theory of elastic bodies. He developed his wave theory of light later in his career, after he moved from The Hague to the more cosmopolitan environment of Paris. Near the end of his life, he wrote a treatise on the possibility of extraterrestrial life.

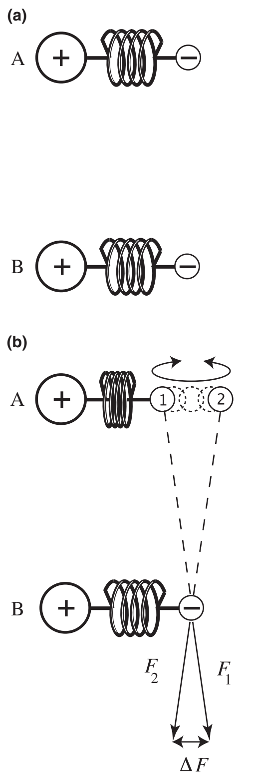

Fig. 1.1 Acceleration of an electron produces a wave. (a) Undisturbed atoms in a source (A) and a receiver (B). Each atom consists of an electron attached to a nucleus by some force, which we represent as a spring. In (b) of the figure, the source electron has been disturbed, and oscillates between positions (1) and (2). The electron at B experiences a force that changes from F_1 to F_2 in the course of A's oscillation. The difference, ΔF, is the amplitude of the changing part of the electric force seen by B.

natural to ask: "What 'stuff' does a light wave disturb?" In one of the major triumphs of nineteenth century physics, James Clerk Maxwell proposed an answer in 1873.

Maxwell (1831–1879), a Scot, is a major figure in the history of physics, comparable to Newton and Einstein. His doctoral thesis demonstrated that the rings of Saturn (discovered by Huygens) must be made of many small solid particles in order to be gravitationally stable. He conceived the kinetic theory of gases in 1866 (Ludwig Boltzmann did similar work independently), and transformed thermodynamics into a science based on statistics rather than determinism. His most important achievement was the mathematical formulation of the laws of electricity and magnetism in the form of four partial differential equations. Published in 1873, ***Maxwell's equations*** completely accounted for separate electric and magnetic phenomena and also demonstrated the connection between the two forces. Maxwell's work is the culmination of classical physics, and its limits led both to the theory of relativity and the theory of quantum mechanics.

Maxwell proposed that light disturbs ***electric and magnetic fields***. The following example illustrates his idea.

Consider a single electron, electron A. It is attached to the rest of the atom by means of a spring, and is sitting still. (The spring is just a mechanical model for the electrostatic attraction that holds the electron to the nucleus.) This pair of charges, the negative electron and the positive ion, is a dipole. A second electron, electron B, is also attached to the rest of the atom by a spring, but this second dipole is at some distance from A. Electron A repels B, and B's position in its atom is in part determined by the location of A. The two atoms are sketched in Figure 1.1a. Now to make a wave: set electron A vibrating on its spring. Electron B must respond to this vibration, since the force it feels is changing direction. It moves in a way that will echo the motion of A. Figure 1.1b shows the changing electric force on B as A moves through a cycle of its vibration.

The disturbance of dipole A has propagated to B in a way that suggests a wave is operating. Electron B behaves like an object floating in your bathtub that moves in response to the rising and falling level of a water wave.

In trying to imagine the actual thing that a vibrating dipole disturbs, you might envision the water in a bathtub, and imagine an entity that fills space continuously around the electrons, the way a fluid would, so a disturbance caused by moving one electron can propagate from place to place. The physicist Michael Faraday[3]

[3] Michael Faraday (1791–1867), considered by many the greatest experimentalist in history, began his career as a bookbinder with minimal formal education. His amateur interest in chemistry led to a position in the laboratory of the renowned chemist, Sir Humphrey Davy, at the Royal Institution in London. Faraday continued work as a chemist for most of his productive life, but conducted an impressive series of experiments in electromagnetism in the period 1834–1855. His ideas, although largely rejected by physicists on the continent, eventually formed the empirical basis for Maxwell's theory of electromagnetism.

supplied the very useful idea of a ***field*** – an abstract *entity* (not a material fluid at all) created by charged particles that permeates space and gives other charged particles instructions about what force they should experience. In this conception, electron B consults the local field in order to decide how to move. Shaking (accelerating) the electron at A distorts the field in its vicinity, and this distortion propagates to vast distances, just like the ripples from a rock dropped into a calm and infinite ocean.

The details of propagating a field disturbance turned out to be a little complicated. Hans Christian Oerstead and Andre Marie Ampère in 1820 had shown experimentally that a changing electric field, such as the one generated by an accelerated electron, produces a magnetic field. Acting on his intuition of an underlying unity in physical forces, Faraday performed experiments that confirmed his guess that a changing magnetic field must in turn generate an electric field. Maxwell had the genius to realize that his equations implied that the electric and magnetic field changes in a vibrating dipole would support one another, and produce a wave-like self-propagating disturbance. Change the electric field, and you thereby create a magnetic field, which then creates a different electric field, which creates a magnetic field, and so on, forever. Thus, it is proper to speak of the waves produced by an accelerated charged particle as ***electromagnetic***. Figure 1.2 shows a schematic version of an electromagnetic wave. The changes in the two fields, electric and magnetic, vary at right angles to one another and the direction of propagation is at right angles to both.

Thus, a disturbance in the electric field does indeed seem to produce a wave. Is this *electromagnetic* wave the same thing as the *light* wave we see with our eyes?

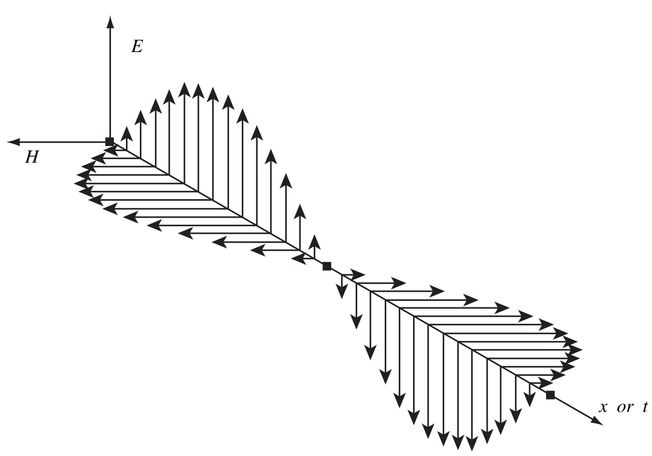

Fig. 1.2 A plane polarized electromagnetic wave. The electric and magnetic field strengths are drawn as vectors that vary in both space and time. The illustrated waves are said to be plane-polarized because all electric vectors are confined to the same plane.

From his four equations – the laws of electric and magnetic force – Maxwell derived the speed of any electromagnetic wave, which, in a vacuum, turned out to depend only on constants:

$$c = (\varepsilon\mu)^{-\frac{1}{2}}$$

Here ε and μ are well-known constants that describe the strengths of the electric and magnetic forces. (They are, respectively, the electric permittivity and magnetic permeability of the vacuum.) When he entered the experimental values for ε and μ in the above equation, Maxwell computed the electromagnetic wave speed, which turned out to be numerically identical to the speed of light, a quantity that had been experimentally measured with improving precision over the preceding century. This equality of the predicted speed of electromagnetic waves and the known speed of light really was quite a convincing argument that light waves and electromagnetic waves were the same thing. Maxwell had shown that three different entities, electricity, magnetism, and light, were really one.

Other predictions based on Maxwell's theory further strengthened this view of the nature of light. For one thing, one can note that for any well-behaved wave the speed of the wave is the product of its frequency and wavelength:

$$c = \lambda\nu$$

There is only one speed that electromagnetic waves can have in a vacuum; therefore there should be a one-dimensional classification of electromagnetic waves (the ***electromagnetic spectrum***). In this spectrum, each wave is characterized only by its particular wavelength (or frequency, which is just c/λ). Table 1.1 gives the names for various portions or ***bands*** of the electromagnetic spectrum.

Maxwell's wave theory of light very accurately describes the way light behaves in many situations. In summary, the theory says:

Table 1.1. *The electromagnetic spectrum. Region boundaries are not well-defined, so there is some overlap. Subdivisions are based in part on distinct detection methods*

Band	Wavelength range	Frequency range	Subdivisions (long λ–short λ)
Radio	>1 mm	<300 GHz	VLF-AM-VHF-UHF
Microwave	0.1 mm–3 cm	100 MHz–3000 GHz	Millimeter–submillimeter
Infrared	700 nm–1 mm	3×10^{11}–4×10^{14} Hz	Far–Middle–Near
Visible	300 nm–800 nm	4×10^{14}–1×10^{15} Hz	Red–Blue
Ultraviolet	10 nm–400 nm	7×10^{14}–3×10^{16} Hz	Near–Extreme
X-ray	0.001 nm–10 nm	3×10^{16}–3×10^{20} Hz	Soft–Hard
Gamma ray	<0.1 nm	$>3 \times 10^{18}$ Hz	Soft–Hard