In December 2013, the United Nations 68th General Assembly declared 2015 to be “The International Year of Light and Light-Based Technologies.” The following essay explores the importance of light to all branches of physics. Optics is the study of light, but here we imagine physics as the study of optics.

In a concept map of physics the study of light stands at all the major intersections. Insights into light illuminate the whole of physics, just as scattered light rays illuminate a whole house. This article is not a scholarly history but an illustrative overview, written with hindsight, of the central role of light in making connections.

In 1267 Roger Bacon, with whom the post-medieval “awakening began,”[2] published Opus Majus. In Book V, the Optics section of that encyclopedic work, Bacon wrote:[3]

“It is possible that some other science may be more useful, but no other science has so much sweetness and beauty of utility. Therefore it is the flower of the whole of philosophy and through it, and not without it, can other sciences be known.”

Seven hundred years later this motif was made explicit by Jacob Bronowski:[4]

“We see matter by light; we are aware of the presence of light by the interruption of matter. And that thought makes up the world of every great physicist, who finds that he cannot deepen his understanding of one without the other.”

Let us begin at the beginning.

Geometrical Optics

“About 10 months ago a rumor came to our ears that a spyglass had been made . . . This finally caused me to apply myself totally to investigating the principles and figuring out the means by which I might arrive at the invention of a similar instrument, which I achieved shortly afterward on the basis of the science of refraction.” –Galileo Galilei [5]

Navigation and surveying have long depended on the straightness of light rays. Through the practical experience provided by these activities, the optical laws of rectilinear propagation and reflection became known in antiquity. The first unified theory in physics came from Hero of Alexandria (c. 10–70 CE), who set forth the principle that light rays follow the path of minimum distance; rectilinear propagation and the law of reflection follow as consequences.[6]

Refraction has been known qualitatively from time immemorial. A partially immersed stick appearing to be sharply bent at the water’s surface was mentioned in Plato’s Republic (c. 360 BCE). “Burning glasses,” lenses for starting fires by focusing sunlight, were part of ancient technology, as documented by artifacts such as a magnifier found in the ruins of the palace of Assyrian King Sennacherib (708–681 BCE). Refraction was made a quantitative
science in the Middle Ages by Muslim scholars such as Ibn al-Haytham (c. 965–1040), known to us as Alhazen, who introduced the practice of measuring angles from the normal for reflected and refracted rays. Alhazen’s contemporary Abu Sâd al-Alâ ibn Sahl (c. 940–1000) expressed the law of refraction in terms of the hypotenuses of right triangles.[7] Willebrord Snellius (or Snell) rediscovered in 1621 the law of refraction, which René Descartes rediscovered again and published in its well-known sine form in 1637.

Refraction made possible the lens, which made the cell and the stars accessible to human senses. Galileo’s Starry Messenger of 1610 and Robert Hooke’s Micrographia of 1665 opened new worlds to investigation. They deepened the questions, and not only for scholars:

. . . He burned his house down for the fire insurance  
And spent the proceeds on a telescope  
To satisfy a lifelong curiosity  
About our place among the infinities.  

Hero’s principle of minimum distance does not explain refraction. That gap was remedied by Pierre de Fermat in 1657 through a broader unifying principle: Of all possible paths connecting two fixed points, the path followed by a light ray minimizes the time for light to go between the points. Fermat’s principle requires light to travel at finite speed. Astronomy offered the first meaningful estimate of this speed in 1676 when Ole Rømer used as a clock the periodic emergence of Io from behind Jupiter’s shadow. (The moon has an orbital period of 42.5 hours.) During the time of year when Earth recedes from the Jupiter–Io system, after each orbit of Io around Jupiter the clock is seen from Earth to run slow. Rømer interpreted the delay as the time light took to travel the additional distance between Earth and Io. Astronomy, which possesses information carried from the heavens to us by light, now gave back from the heavens information about light itself.

Lenses and Spectra

“I procured me a triangular glass-prisme, to try therewith the celebrated Phenomena of Colours . . .” –Isaac Newton

The edge of every lens forms a prism. The rainbow of colors that emerges from prisms was familiar in Aristotle’s time. Received doctrine held that a prism somehow modifies the color of light. Isaac Newton had to investigate. He made a hole in his window shutter to let in a fine beam of sunlight. The prism produced the expected colors of the rainbow, but Newton noticed the significance of something else: the circular beam that entered the prism emerged as an elongated ellipse. Each color refracted at a different angle.[8]

With a second aperture Newton could select from this rainbow one color to enter a second prism. This prism did not change the color. Allowing all the colors to enter the second prism produced white light on its far side. A prism did not modify light but separated it. Newton wrote, “A naturalist would scarce expect to see ye science of those colours become mathematical, and yet I dare affirm that there is as much certainty in it as in any other part of Opticks.”[9] This image of a prism separating white light into a spectrum and the inverse operation of synthesizing distinct colors into white light, illustrates visually the mathematics of synthesis and analysis, such as the harmonic series of Fourier’s theorem.

William Herschel and his sister Caroline made some of the first catalogs of stars, discovering many binary systems and the planet Uranus. While testing a red filter for observing sunspots, William happened to place his hand at the focal point of his reflecting telescope and noticed the region to be unexpectedly warm. To study the temperature of light, in 1800 William inserted thermometers into the separate colors of the sun’s spectrum. He noticed that in going from violet to red, the temperature increased. Intrigued, he placed a thermometer beyond the red, and there found the highest temperature. Herschel called this warm invisible light beyond the red “caloric rays,” which we know as infrared. Herschel’s results were anticipated by 63 years by Emilie du Châtelet. This remarkable woman essentially discovered the work–energy theorem, translated Newton’s Principia into the French translation used to this day, and collaborated with Voltaire across many years. Her opus was Eléments de la Philosophie de Newton (1738), which went deep into the philosophical foundations of mechanics and was influential in shifting French scientists from the mechanics of Descartes to that of Newton. In 1737 du Châtelet entered an essay competition on the nature of fire. In her essay “Dissertation on the Nature and Propagation of Fire,” she argued that fire is not a material substance, and different colors of light transport different quantities of heat. The way to demonstrate this, she suggested, was to line up an array of thermometers, one inserted into each of the separated colors of the spectrum, which was precisely what William Herschel did in 1800. du Châtelet was not able to perform the experiment herself for lack of thermometers.[10]

Joseph von Fraunhofer supervised glass melting and grinding processes in his Munich optical institute. He needed to measure the refractive indices for different colors in various kinds of glass. In one of his experiments, light from an oil lamp flame passed through a prism to be viewed through a telescope. Fraunhofer noted dark lines in the spectrum. Intrigued, he looked for generalizations. Repeating Newton’s experiment on sunlight with his telescope-equipped prism, in 1814–15 dark lines were revealed in the solar spectrum.

In 1857 the “daring and resourceful experimenter” Robert Bunsen invented a burner that produced a colorless flame.[11] With Bunsen’s burner the spectra of chemicals placed in the flame could be cleanly separated. His collaborator Gustav Kirchhoff added a prism to complete the basic tool of modern spectroscopy, the spectroscope. Payoffs came quickly. In 1860 Bunsen and Kirchhoff discovered rubidium and cesium in a sample of Dükkheim mineral water. In 1868 two astronomers, Pierre Janssen from France and Norman Lockyer from England, independently reported a yellow line in the solar spectrum that fit no known element. Interpreting it as an unknown element, Lockyer named it after helios, Greek for “the Sun.”[12] Terrestrial helium was not confirmed until 1895 when William Ramsey isolated it as a byproduct of uranium ore. In 1907 Ernest Rutherford and Thomas Roysd collected alpha particles emitted by radioactive decay, examined their spectra, and showed that the particles were helium.

Classical Mechanics

“Following in the footsteps of Hero and Fermat, he [Mau- pertuis] then proclaimed that this simplicity causes nature to act in such a way as to render a certain quantity, which he named the ‘action,’ a minimum.” –Wolfgang Yourgrau and Stanley Mandelstam [13]
After Newton revolutionized optics he turned to mechanics. Generalizing inductively from specific problems solved in quantitative detail [14]—Archimedes on the lever, Galileo on projectiles, Huygens on the pendulum, and Newton himself on gravitation—he postulated in 1687 three laws of motion that turned mechanics into an axiomatic system. As the laws of geometrical optics could be derived from Fermat’s least time principle, could the same be done for mechanics? Several proposals were forthcoming. These included Johann Bernoulli’s 1717 principle of virtual work for statics, extended to dynamics by Jean le Rond d’Alembert in 1743.

Around 1740 Pierre Louis Moreau de Maupertuis (who tutored young Emilie du Châtelet in calculus) suggested that a particle acted on by specific forces moves in a way that minimizes the “action.” This approach was successfully demonstrated for central forces by Leonhard Euler in 1744. In his Mécanique Analytique of 1788, Joseph Lagrange generalized Maupertuis’ principle to all conservative forces and clarified “action” as the line integral of momentum. The generalization of this principle to all of mechanics (later extended to most of physics) was published in two papers by William R. Hamilton in 1834–35.[15] Hamilton’s principle postulates that of all the conceivable trajectories whereby a particle might travel between two fixed points, the trajectory actually followed minimizes the time-averaged difference between the particle’s kinetic and potential energies. The principles of Hamilton and Fermat arose from similar motivations, but a logical connection between them would have to await general relativity.

**Ontology**

> “From the multitude of experiences it [science] selects a few simple forms, and constructs from them, by thought, an objective world of things.” –Max Born [16]

> “You know something and then the quality stimulus hits . . . , but to define it all you’ve got to work with what you know. So your definition is made up of what you know. It’s an analogue to what you already know.” –Robert Pirzsig [17]

A debate about the ultimate reality of light began in the time of Plato and the Sophists. By the time of Newton and Huygens, those arguing the question “What is light?” faced a binary choice: What is light—wave or particle? Robert Hooke’s *Micrographia* describes how colors of thin films depended on a film’s thickness, suggesting a standing wave condition. Christaan Huygens argued that the tremendous speed of light would be feasible only if light was a disturbance through a medium, not the bulk motion of a medium. He gave the wave hypothesis predictive power by postulating that each point on a wave front behaves as the source of another wave. If that were so, then light should radiate into regions that would otherwise remain in geometric shadow. Hooke and Francesco Grimaldi had noticed diffraction in the fine structure of shadows cast by a needle.

Initially ambivalent (“I make no hypotheses”), Newton eventually argued that light was a beam of particles. While acknowledging that something periodic occurs with waves (and discovering an interference pattern called “Newton’s rings”), he interpreted the periodicity as something that matter does to light. To Newton, the diffraction reports did not require light to be a wave. Gravity acts between separated massive bodies, so matter could bestow its periodic influence on light from a distance.

Refraction offered one way to decide the question. When light passes from air into water the ray bends toward the normal. If light consists of waves, the speed of light in water would be less than its speed in air. If light consists of particles the reverse would happen.

In 1800 Thomas Young demonstrated that the interference of light passed through a double aperture. Such a pattern could be interpreted only as the superposition of waves. Augustin Fresnel worked out a comprehensive theory of diffraction based on the assumption that light consists of waves, and his predictions were vindicated, famously so with the notorious “Poisson’s spot,” a bright spot, due to wave diffraction, in the shadow behind an illuminated disk. In 1850 Léon Foucault measured the speed of light in water and found it to be less than the speed of light in air. The riddle “What is light?” seemed answered.[18]

Lingering questions remained, as they always do with important questions that have multiple layers. First, supposing light to be a wave, what is waving? Second, acoustical waves require a medium; what serves as the medium for light, the “aether”? Third, light had been found to be polarized by birefringent crystals. Reconciling polarization and the rapid speed of light with our ability to breeze freely through the aether offered a perplexing situation.

**Electromagnetism**

> "Maxwell shewed light to be an electromagnetic phenomenon, so that the whole science of Optics became a branch of Electromagnetism. . . ." –James Jeans [19]

Hints at a connection between electricity and magnetism came when Hans Christian Ørsted showed that moving electric charge makes magnetism and when Michael Faraday showed that changing magnetism makes electricity. A unified theory of electromagnetism was written by James Maxwell in 1862. Action at a distance, which served well for static interactions, was replaced with the dynamic concept of the field, a function of space and time.

The interactions of matter proceed through fields. On one hand, local fields tell a particle of matter how to move. Newton’s second law with the Lorentz force, for instance, predicts the motion of a charged particle in response to electromagnetic fields. On the other hand, matter determines the fields around it. Maxwell’s equations relate the electric and magnetic fields to their charged particle sources and relate the fields to each other. When a charged particle accelerates, Maxwell’s equations say the fields it produces must...
change. A changing electric field produces a magnetic field that also changes, and the changing magnetic field produces a changing electric field. Together the changing fields make a self-propagating wave moving at the speed of light.

In response to the “What is waving?” question, light must thus be a wave in the electromagnetic field! The equations describing this wave have no restriction on the frequency, suggesting the existence of a continuous electromagnetic spectrum of harmonics whose frequencies range from zero to infinity. The equations also say that the propagating fields are transverse to the direction of wave travel, implying polarization and explaining the effects of birefringent crystals.

In 1886–89 Heinrich Hertz affirmed Maxwell by broadcasting and detecting radio waves in the laboratory. While doing so the alert Hertz noticed a spurious glitch in his apparatus. Radiation of low intensity but sufficiently high frequency immediately stimulates an electric current in certain materials; at low frequencies the incoming light produces no current even at high intensity. Dubbed the photoelectric effect, this anomaly in the interaction of light with matter did not fit Maxwell's theory. For two decades it remained a mystery.

Maxwell had answered important questions about light, but others remained. The equations say that electromagnetic waves need no medium, that they travel in empty space at the speed of light, c, but the equations are silent on the frame of reference. In 1895 16-year-old Albert Einstein wondered what he would observe if he rode on a beam of light. Intuition said that Einstein's light-wave surfer should observe a stationary crest of the electromagnetic wave. But Maxwell's equations insist that electromagnetic waves travel at speed c even from the surfer's perspective! This paradox, like all paradoxes, suggested that the question should be restated.

Einstein held the question in his mind for 10 years. Then the 26-year-old Einstein wrote “On the Electrodynamics of Moving Bodies,” noting that “Maxwell's electrodynamics—as usually understood at the present—when applied to moving bodies, leads to asymmetries that do not seem to be inherent in the phenomena.”[20]

The relative motion between a magnet and a coil of conducting wire illustrates the issue. Whatever the reference frame, the relative motion results in a force on the charge carriers, driving an electric current in the coil. An observer aboard the coil sees a changing magnetic flux as the magnet sweeps by. Faraday's law says an electric field $E$ gets induced in the coil, producing the force $qE$ on the charges. An observer aboard the magnet sees a different picture. The coil sweeps by with velocity $v$, carrying the charged particles through the magnetic field $B$. Each charge $q$ feels the force $q\mathbf{v}\times\mathbf{B}$. Thus do distinct mechanisms describe the same result, an asymmetry that do not seem to be inherent in the phenomena. Einstein wondered what principle would unify the two explanations.

The thought experiment about light surfing suggested a clue in light itself. If you ride on the beam of light that bounces off a clock at 10:00 am, then you stay with the information that says the time is 10 o'clock.[21] For the light-wave surfer, time stands still. Newtonian relativity of inertial frames postulates the separate invariance of length and time intervals; as a consequence, the speed of light must be relative. Einstein replaced those assumptions with the postulate of the invariance of the speed of light between inertial frames, which requires space and time intervals to be relative. Mechanics had to adapt to light, instead of the light adapting to mechanics.

Special relativity, which linked light to space and time, also linked light to mass and energy. Energy and momentum became the time and space components of a vector in four-dimensional space-time. Its geometry was not Euclidean but hyperbolic. The square of the energy–momentum four-vector was given by a difference, not a sum, with the particle's mass as the vector's magnitude. For a free particle, $E^2 - (pc)^2 = (mc^2)^2$.

### Thermodynamics and Quantum Physics

“By 1906 or 1908 Planck had come to see that his compromise over cavity radiation had loosed something brand new and menacing into the world of physics.” —J.L. Heilbron [22]

The thermodynamics of light motivated the extension of Newtonian mechanics to quantum mechanics. Macroscopic thermodynamics serves as a boundary condition on microscopic statistical mechanics. After many triumphs with engines and phase changes and the kinetic theory of gases, statistical thermodynamics confronted the question of finding the energy density of light as a function of frequency. Light and matter in thermal equilibrium was produced in the laboratory by a metal box held at temperature $T$. The atoms in the box walls are made of oscillating charged particles and radiate light. According to Newtonian mechanics, the energy of a harmonic oscillator is proportional to the frequency squared. The sum over all microscopic states, a procedure required by statistical mechanics, thus predicts an energy density that diverges as frequency cubed, the “ultraviolet catastrophe.” Although the experimental spectrum of light in thermal equilibrium with matter goes as the frequency cubed at low frequencies, as the frequency increases the distribution mapped by data reaches a peak and then slides toward zero at the highest frequencies.

Max Planck realized that the predicted distribution function could be made to peak and trail off at high frequencies if the energy of an oscillator of frequency $f$ was linear in $f$ and exhibited only a harmonic series of discrete overtones, so that $E_n = nhf$, where $n = 0,1,2,3, \ldots$ with $h$ a constant to be fit to data. The distribution function that resulted had the right shape, whatever the value of Planck's constant $h$. It fit the data precisely if $h$ had the astonishingly small value $6.6\times10^{-34} \text{ J s}$. Planck had solved this important problem, but at the price of making an ad hoc hypothesis about energy quantization, a drastic move which at the time pointed to nothing else.

Five years after Planck's hypothesis Einstein revisited the thermodynamics of light. He calculated the entropy of radiation and compared the result to the entropy of a box filled with ideal gas molecules. Then came the heretical punch line. The entropy of the
radiation matches the entropy of the molecules, said Einstein, if a light wave of frequency \( f \) corresponds to a swarm of particles, each carrying energy \( E = hf \). According to Einstein, light itself is quantized. He showed how this corpuscle interpretation of light solved outstanding mysteries in the interaction of matter and radiation. Most famously, the photoelectric effect made sense as a collision between a light corpuscle and an electron if Einstein’s \( h \) has the same value as Planck’s \( h \). Planck’s constant \( h \) pointed to something deep.[23] The name of the light corpuscle, the photon, came years later, in 1926.[24]

With the concept of the photon in mind, one can look again to special relativity, which requires any particle moving at the speed of light to carry zero mass. With zero mass, the energy–momentum relation for a photon simplifies to \( E = pc \). Together with \( E = hf \) and \( c = \lambda f \), it follows that a light wave of wavelength \( \lambda \) corresponds to a swarm of photons, each carrying momentum \( p = h/\lambda \). This idea, rigorous for massless particles, was boldly postulated by Louis de Broglie in 1924 to hold for massive particles too. Thus did the thermodynamics of light—along with spectroscopy’s stained-glass window into the atom—lead the way into quantum mechanics.

General Relativity

“Another important consequence of the theory, which can be tested experimentally, has to do with the path of rays of light. . . . We can therefore draw the conclusion from this, that a ray of light passing near a large mass is deflected . . . The existence of this deflection, which amounts to 1.7 . . . was confirmed, with remarkable accuracy, by the English Solar Eclipse Expedition in 1919. . . .” –Albert Einstein [25]

Between 1905 and 1915 Einstein extended special relativity to arbitrarily accelerated frames. Thanks to the principle of the equivalence of gravitational and inertial mass, general relativity serves as a theory of gravitation. Early tests of general relativity checked its predictions for the behavior of light, including the deflection of a light ray grazing the sun, gravitational redshift, and radar echo delay.

David Hilbert realized that Einstein’s gravitational field equations could be derived in analogy to Fermat’s principle: Of all the possible trajectories that a particle might follow between two events in space-time, the trajectory actually followed maximizes the particle’s proper time for the trip. In the limiting case of a particle moving slowly in a weak gravitational field, this “Fermat’s principle for gravity” reduces to Hamilton’s principle of classical mechanics.

Newtonian cosmology had originally envisioned a static, everlasting, infinite universe. However, the Newtonian universe was unstable and paradoxical—how could a universe filled to infinity with stars show a dark sky at night (Olbers’ paradox)? In 1917, with his new tool expressing gravitation as the curvature of space-time, Einstein solved the cosmological problem at infinity by abolishing infinity. He postulated the three-dimensional universe to be the surface of a static sphere embedded in four-dimensional Euclidean space. Alexander Friedmann and Georges Lemaître asked why the universe must be static. Their equations predicted a universe in which space could contract or stretch to show a velocity–distance relation. At the cosmic scale the relative speed of two points would be proportional to their separation.

Measuring astronomical distances requires the light of standard candles. Henrietta Swan Leavitt provided crucial candles in 1912 when she discovered a relationship between the periods and luminosities of Cepheid variable stars. Edwin Hubble used Cepheids in 1924 to probe distances to spiral nebulae, which turned out to be millions of light-years away. The universe suddenly became very big. By applying the Cepheid distance indicators and Doppler shifts to the spectra of galaxies, in 1929 he offered the first evidence for the cosmic velocity–distance relation. The journey toward big-bang cosmology was underway.

In a universe that begins in the big-bang scenario, after the primordial gas of relativistic particles cools sufficiently for atoms to form, an afterglow of photons must remain. The wavelengths of those photons are continuously stretched by the cosmic expansion. In 1948 the existence in our universe of this background radiation was predicted by Ralph Alpher and Robert Hermann. Their first estimate placed its temperature today near 5 K. Alpher and Hermann tried throughout the 1950s to convince radio astronomers to look for the afterglow.[26] In 1964 it was accidentally found by Arno Penzias and Robert Wilson. Their measurements gave a temperature of 2.7 K.[27] Ever since, it has offered a window into the genesis of the universe.

Today light has become the most incisive of tools in cosmology. Precision measurements of the cosmic afterglow of the big bang heralded the era of precision cosmology; the harmonics in the afterglow’s power spectrum offer a kind of electrocardiogram for the early universe. The irony of our present state of fertile ignorance is that the greatest mysteries at present are not about the existence of light, but its absence: dark matter and dark energy. Could dark energy be our aether?

Quantum Electrodynamics and Beyond

“The diagrams we make of quarks exchanging gluons are very similar to the pictures we draw for electrons exchanging photons. So similar, in fact, that you might say that the physicists have no imagination—that they just copied the theory of quantum electrodynamics for the strong interactions! And you’re right: that’s what we did, but with a little twist.” –Richard Feynman [28]

In the mid-1920s, quantum mechanics developed into the form now taught to physics majors. But it took two more decades to reconcile quantum mechanics with electrodynamics. An electron is not an ideal point charge. The “total” electron includes its ideal “bare” charge plus the interactions of the electron with its own electromagnetic field. An electron emits and reabsorbs photons, and some of those photons briefly turn into electron–positron pairs that combine back into a photon before returning to the original electron. The energy budget for producing these virtual particles comes from the energy fuzziness inherent in the Heisenberg uncertainty principle. Thus we see as “the electron” in the laboratory includes a cloud of virtual photons and electron–positron pairs. This is a serious problem because these intermediate processes contribute infinity to the quantum state!

The remedy is “renormalization.” A theory is said to be renormalizable when all divergent pieces cancel out each other in perturbation theory, leaving as a residue the observed charge and mass. According to our present understanding, renormalizability presents a necessary condition for any sensible theory of fundamental interactions.

Quantum electrodynamics—the interaction of light with electrically charged matter—was the first renormalizable theory of elementary particle interactions. It serves as the template for the
other theories of elementary particle physics.[29] At its foundation stands a principle of least action, adapted to quantum field theory, that traces its inspiration back through the analogous principles of Hamilton and Fermat.[30]

***

From quarks to cosmology, light has been a tool, a model, and an inspiration to all of physics. Light has also been a metaphorical symbol of hope and wisdom in all cultures. The Hindu four-day festival of lights, Diwali, celebrates the triumph of knowledge over ignorance, hope over darkness. In the Book of Genesis, the “poem of the dawn” in the Judeo-Christian mythos, God speaks the universe into existence by uttering “Let there be light.” In Buddhism one seeks enlightenment, the lights of wisdom and compassion. Let 2015, the Year of Light, be a celebration of knowledge and wisdom overcoming poverty and ignorance.[31] Physics and its technological applications have essential roles in achieving these ends. May we use them wisely and in the service of all that lives. May the secular world of physics help us find “our place among the infinities” in a festival of light.

Acknowledgments

I thank Devin Powell, Kendra Redmond, and Daniel Golombek for suggestions that resulted in an improved manuscript.

References

[2] “The awakening began with Roger Bacon” comes from Will Durant, The Story of Philosophy (Garden City Publishing Co., New York, NY, 1938), 117. In the present article, events, persons, and principles well known from introductory and intermediate physics textbooks are not referenced.
[31] This may also be an opportune moment to lament the thoughtless excesses of light pollution and light trespass.